

Bioorthogonal Noncanonical Amino Acid Tagging for  
Selective Analysis of the *Pseudomonas aeruginosa*  
Proteome

Thesis by  
Brett M. Babin

In Partial Fulfillment of the Requirements for the  
degree of  
Doctor of Philosophy in Chemical Engineering

The logo for the California Institute of Technology (Caltech), featuring the word "Caltech" in a bold, orange, sans-serif font.

CALIFORNIA INSTITUTE OF TECHNOLOGY  
Pasadena, California

2016  
Defended May 10, 2016

© 2016

Brett M. Babin

ORCID: 0000-0002-4133-6665

All rights reserved

## ACKNOWLEDGEMENTS

First, thanks to my advisor, Prof. Dave Tirrell. Dave's attention to detail and thoughtful curiosity foster a unique environment for scientific exploration and most of what I know about being a scientist, scholar, and communicator I learned by following his example. Thanks also to Prof. Dianne Newman. Dianne's excitement about the beauty and importance of microbiology is contagious and her enthusiasm and encouragement drove me to delve further into the study of bacterial physiology than I ever expected. Thanks to the rest of my thesis committee, Prof. Rustem Ismagilov and Prof. Frances Arnold for their help and advice.

Many thanks to Dr. Megan Bergkessel. Megan and I worked closely for many years, and our successes are very much a testament to her technical prowess, her creativity, and her impressive ability to cut through biological complexity to reach new and interesting conclusions.

The projects described here are the result of many other collaborations with talented researchers. First, nearly all of my work relied on high quality mass spectrometry data and I'm indebted to the talent and expertise of the Proteome Exploration Laboratory: Dr. Sonja Hess, Dr. Annie Moradian, Dr. Mike Sweredoski, Roxana Eggleston-Rangel, and Dr. Bobby Graham. I benefited greatly from my summer at the University of Copenhagen, where I learned methods for the study of biofilms. Thanks to Prof. Tim Tolker-Nielsen and Dr. Mustafa Fazli for their generous training. Thanks to Lydia Atangcho and Dr. Mark van Eldijk for their contributions to the biofilm project. Thanks to Dr. Ben Ramirez, director of the Center for Structural Biology at the University of Illinois at Chicago for his efforts to determine the solution structure of SutA via NMR. Thanks to Prof. Victoria Orphan and Dr. Roland Hatzenpichler for initiating an interesting and productive collaboration to introduce the BONCAT method to the field of environmental microbiology.

I thank the members of the Tirrell and Newman labs, both present and past, for making my time at Caltech an enjoyable one. I especially appreciate all members of the BONCAT team for fruitful discussions: Prof. John Ngo, Dr. Beverly Lu, Dr. Alborz Mahdavi, Dr. JD Bagert, Kai Yuet, Shannon Stone, Dr. Graham Hamblin, Dr. Weslee Glenn, and Dr. Mark van Eldijk.

Finally, thank you to my friends and family, near and far, and especially Vanessa, Oreo, and Kitten.

## ABSTRACT

In natural environments, bacterial physiology is frequently characterized by slow metabolic rates and complex cellular heterogeneities. The opportunistic pathogen *Pseudomonas aeruginosa* provides one such example; *P. aeruginosa* forms untreatable chronic biofilm infections of the cystic fibrosis lung, where oxygen limitation can lead to states of metabolic dormancy. To better understand the biology of these states, *in vitro* experiments must be adapted to better recapitulate natural settings. However, low rates of protein turnover and cellular or phenotypic complexity make these systems difficult to study using established methods. Here we adapt the bioorthogonal noncanonical amino acid tagging (BONCAT) method for time- and cell-selective proteomic analysis to the study of *P. aeruginosa*. Analysis of proteins synthesized in an anoxic dormancy state led to the discovery of a new type of transcriptional regulator which we designated SutA. We performed detailed analyses of SutA's role in transcription under slow growth states and we elucidated the structural basis for its regulatory behavior. Additionally, we used cell-selective targeting of BONCAT labeling to determine the dynamic proteomic response of an antibiotic-tolerant biofilm subpopulation to challenge with ciprofloxacin. Overall this work shows the utility of selective proteomics as applied to bacterial physiology and describes the broad biological insight obtained from that application.

## PUBLISHED CONTENT AND CONTRIBUTIONS

- (1) Ngo, J. T.; Babin, B. M.; Champion, J. A.; Schuman, E. M.; Tirrell, D. A. *ACS Chem. Biol.* **2012**, *7*, 1326–1330.
- (2) Hatzenpichler, R.; Scheller, S.; Tavormina, P. L.; Babin, B. M.; Tirrell, D. A.; Orphan, V. J. *Environ. Microbiol.* **2014**, *16*, 2568–2590.
- (3) Babin, B. M.; Bergkessel, M.; Sweredoski, M. J.; Moradian, A.; Hess, S.; Newman, D. K.; Tirrell, D. A. *Proc. Natl. Acad. Sci. U.S.A.* **2016**, *113*, E597–E605.

## TABLE OF CONTENTS

Acknowledgements . . . . .	iii
Abstract . . . . .	iv
Published Content and Contributions . . . . .	v
Table of Contents . . . . .	vi
List of Illustrations . . . . .	viii
List of Tables . . . . .	x
Nomenclature . . . . .	xi
Chapter I: Adapting and Applying BONCAT to the Study of <i>Pseudomonas</i> <i>aeruginosa</i> Physiology . . . . .	1
References . . . . .	4
Chapter II: Proteomic Analysis of <i>P. aeruginosa</i> During Anaerobic Dormancy and the Discovery of SutA, a Slow-growth Transcription Factor. . . . .	5
2.1 Summary of Contributions . . . . .	5
2.2 Abstract . . . . .	6
2.3 Introduction . . . . .	7
2.4 Results . . . . .	8
2.5 Discussion . . . . .	22
2.6 Experimental Procedures . . . . .	24
2.7 Acknowledgments . . . . .	26
2.8 Supplementary Figures . . . . .	27
References . . . . .	34
Chapter III: Structural Insight into SutA's Mechanism of Transcriptional Reg- ulation . . . . .	37
3.1 Summary of Contributions . . . . .	37
3.2 Abstract . . . . .	38
3.3 Introduction . . . . .	39
3.4 Results . . . . .	40
3.5 Discussion . . . . .	48
3.6 Future Work . . . . .	50
3.7 Experimental Procedures. . . . .	51
3.8 Supplementary Figures . . . . .	58
3.9 Supplementary Tables. . . . .	66
References . . . . .	67
Chapter IV: Proteomic Response of an Antibiotic-tolerant Biofilm Subpopu- lation to Ciprofloxacin . . . . .	69
4.1 Abstract . . . . .	69
4.2 Introduction . . . . .	70
4.3 Results . . . . .	72
4.4 Discussion . . . . .	83

4.5 Experimental Procedures . . . . .	85
4.6 Supplementary Figures . . . . .	90
4.7 Supplementary Datasets . . . . .	95
References . . . . .	96
Appendix A: Supplementary Information for Chapter 2 . . . . .	101
A.1 Supplementary Experimental Procedures . . . . .	101
A.2 Supplementary Tables . . . . .	115
A.3 Supplementary Datasets . . . . .	119
References . . . . .	122
Appendix B: Other Contributions . . . . .	124
B.1 Contributions to Ngo, et al. . . . .	124
B.2 Contributions to Hatzenpichler, et al. . . . .	127

## LIST OF ILLUSTRATIONS

<i>Number</i>	<i>Page</i>
2.1 BONCAT enables enrichment and identification of proteins synthesized during anaerobic survival. . . . .	10
2.2 Phenotypic characterization of <i>sutA</i> mutants. . . . .	12
2.3 SutA upregulation during slow growth is post-transcriptional. . . . .	13
2.4 RNA polymerase co-precipitates with SutA. . . . .	15
2.5 SutA localizes throughout the chromosome and enhances transcription of ribosomal genes. . . . .	19
2.6 SutA has broad effects on gene expression. . . . .	22
2.S1 BONCAT labeling and enrichment during anaerobic survival. . . . .	28
2.S2 Phenotype screens and $\Delta sutA$ growth characterization. . . . .	30
2.S3 RpoA co-immunoprecipitated proteins. . . . .	31
2.S4 HA-SutA and RpoA chromatin immunoprecipitation. . . . .	33
3.1 SutA cross-links to RNAP <i>in vitro</i> . . . . .	44
3.2 SutA interacts with both sides of the RNAP clamp. . . . .	46
3.3 SEC analysis of the SutA-RNAP complex. . . . .	47
3.4 Effects of SutA truncation on pyocyanin production. . . . .	48
3.S1 SutA structural predictions. . . . .	58
3.S2 BS <sup>3</sup> cross-linking. . . . .	59
3.S3 MS2 spectra for SutA-RNAP BS <sup>3</sup> cross-links. . . . .	60
3.S4 Bpa cross-linking. . . . .	61
3.S5 Bpa cross-link spectra. . . . .	62
3.S6 Protein foot-printing. . . . .	63
3.S7 Intra-RNAP BS <sup>3</sup> cross-links. . . . .	64
3.S8 2D NMR of SutA. . . . .	65
4.1 Cell state-selective labeling using the <i>rpoS</i> promoter. . . . .	74
4.2 Targeted proteomics of a biofilm subpopulation. . . . .	77
4.3 BONCAT analysis of protein synthesis during ciprofloxacin challenge. . . . .	80
4.4 Dynamic cellular responses to ciprofloxacin. . . . .	82
4.S1 Promoter-controlled expression. . . . .	90
4.S2 Enrichment from biofilms of <i>P<sub>rpoS</sub></i> and <i>P<sub>trc</sub></i> strains. . . . .	91
4.S3 Shared and unique proteomic hits. . . . .	92

4.S4	Other responses to ciprofloxacin. . . . .	94
B.1	Proteomic labeling with AnI under conditions of oxidative stress. . .	125
B.2	Tagging rate in the SoxRS system is less than 10% . . . . .	125
B.3	Tagging rate in the SoxRS system is dependent on the degree of induction of NLL-MetRS expression. . . . .	126

## LIST OF TABLES

<i>Number</i>		<i>Page</i>
3.1	BS <sup>3</sup> cross-linked peptides. . . . .	42
3.2	Bpa cross-links . . . . .	43
3.3	Chapter 3: Strains and plasmids. . . . .	66
3.4	Bpa mass modifications . . . . .	66
4.1	Chapter 4: Strains and plasmids. . . . .	86
A.1	Chapter 2: Strains and plasmids. . . . .	115
A.2	Chapter 2: Primers. . . . .	116

## NOMENCLATURE

- BS<sup>3</sup>**. Bis(sulfosuccinimidyl) suberate.
- Aha**. L-azidohomoalanine.
- Anl**. L-azidonorleucine.
- BONCAT**. Bioorthogonal noncanonical amino acid tagging.
- Bpa**. L-benzoylphenylalanine.
- CF**. Cystic fibrosis.
- ChIP**. Chromatin immunoprecipitation.
- DBCO**. Dibenzylcyclooctyne.
- GFP**. Green fluorescent protein.
- ICD**. Isocitrate dehydrogenase.
- IP**. Immunoprecipitation.
- LC-MS/MS**. Liquid chromatography-tandem mass spectrometry.
- MIC**. Minimum inhibitory concentration.
- ncAA**. Noncanonical amino acid.
- RNAP**. RNA polymerase.
- ROS**. Reactive oxygen species.
- rRNA**. Ribosomal RNA.
- TAMRA**. Tetramethylrhodamine.
- TCA cycle**. Tricarboxylic acid cycle.
- UTR**. Untranslated region.

*Chapter 1***ADAPTING AND APPLYING BONCAT TO THE STUDY OF  
*PSEUDOMONAS AERUGINOSA* PHYSIOLOGY**

The overarching goal for my doctoral work was to adapt a method for selective proteomic analysis toward the study of bacterial physiology. When I began this effort, the bioorthogonal noncanonical amino acid tagging (BONCAT) method for proteome labeling and enrichment [1] had recently been modified to allow for cell-selective analyses [2]. BONCAT relies on the cellular incorporation of a noncanonical amino acid (ncAA) into nascent proteins. Proteins that contain the ncAA are chemically distinct from the pre-existing proteome, and the presence of a functional chemical handle on the ncAA (e.g., an azide) allows for selective chemical targeting of these proteins via bioorthogonal chemistry (e.g., azide-alkyne cycloaddition). Labeled proteins can be visualized in cells or lysates via reaction with a fluorescent tag, or can be enriched via reaction with an affinity tag or solid support followed by chromatography. Enriched proteins can then be identified and quantified by liquid chromatography-tandem mass spectrometry (LC-MS/MS).

Temporal selectivity of protein labeling is achieved with a ncAA that is incorporated by wild-type translational machinery (e.g., *L*-azidohomoalanine, Aha) by controlling when cells are exposed to the ncAA. Cell-selective protein labeling is achieved through the use of a ncAA (e.g., *L*-azidonorleucine, Anl) that is only incorporated by cells that express a mutant aminoacyl-tRNA synthetase (mRS) engineered for activity toward the ncAA. By restricting expression of the mRS to cell-types of interest, BONCAT labeling can be targeted to particular cells present in complex, heterogeneous mixtures.

Through the work of my colleagues and others, BONCAT has been established as a powerful tool for studying nascent protein synthesis in a broad range of biological contexts. For example, the method has been used to study localized translation in mammalian neuronal cultures [3], newly synthesized proteins in mice [4], and to selectively target tissues in *Caenorhabditis elegans* [5], and *Drosophila melanogaster* [6]. In bacteria, time-selective labeling with Aha has revealed the dynamic proteome of *Bacillus subtilis* reviving from spores [7] and proteomics of the quorum sensing response of *Vibrio harveyi* [8], while cell-selective approaches have been used to

identify bacterial proteins important for host-cell infection [9, 10]. See Yuet and Tirrell for a comprehensive review [11].

Because BONCAT offers sensitive temporal and cellular selectivity, we thought it would be particularly well suited to address questions of bacterial slow growth and heterogeneity. Of particular interest was the opportunistic pathogen *Pseudomonas aeruginosa*, a gram-negative bacterium whose infections of the cystic fibrosis (CF) lung are chronic and recalcitrant to both the host immune system and to antimicrobial therapies. These infections are characterized by bacteria in a dormant state with low metabolic rates, a physiological condition known to contribute to antibiotic tolerance. Additionally, *P. aeruginosa* grows as biofilm microcolonies within the CF lung, a growth state in which nutrient gradients lead to phenotypic heterogeneities that contribute further to tolerance. We set out to learn more about the physiology of these states *in vitro* through selective proteomics. My contributions were twofold: (i) the application of BONCAT to bacterial systems in which low metabolic rates or cellular heterogeneity create difficulties for traditional proteomic analyses and (ii) the discovery and characterization of a new regulatory protein that helps *P. aeruginosa* to adapt to these challenging conditions.

Chapter 2 describes the application of the BONCAT method for temporally-selective proteomic analysis. In this work, we determined the nascent proteome of *P. aeruginosa* subsisting in an anoxic survival state. Analysis of proteins preferentially expressed in this state led to the discovery of a previously uncharacterized transcription factor, which we now call SutA (survival under transitions). We found SutA to be important for biofilm formation, the production of *P. aeruginosa*'s phenazine virulence factors, and the organism's ability to adapt to changing conditions. We identified an interaction between SutA and RNA polymerase, and through this interaction, its association with much of the chromosome. In particular, SutA shows high levels of association with loci encoding ribosomal components (ribosomal proteins and ribosomal RNA) and its presence in the cell enhances the expression of these genes. In addition, SutA generally shifts the gene expression profile away from genes involved in primary metabolism and toward those involved in cellular maintenance and secondary metabolisms.

Chapter 3 describes our investigations into the physical interaction between SutA and RNA polymerase. Because the primary amino acid sequence of SutA does not match any characterized proteins or domains, nothing was known about its mechanism of transcriptional regulation. We undertook a series of *in vitro* experiments to

characterize the structure of SutA and the nature of its binding to RNA polymerase. Through chemical cross-linking and protein foot-printing, we find evidence that SutA binds the  $\beta$  lobe 1 and  $\beta'$  clamp domains of RNA polymerase and, through this interaction, may elicit a conformational change of RNA polymerase. We suggest how this function may explain the physiological effects described in Chapter 2.

Chapter 4 describes the adaptation of the cell-selective BONCAT method for the study of heterogeneous *P. aeruginosa* biofilms. We take advantage of phenotypic differences between biofilm cells to restrict BONCAT labeling to an antibiotic-tolerant subpopulation. We characterize our ability to selectively enrich and identify proteins synthesized by this subpopulation. We then determine the dynamic proteomic response of these cells to antibiotic challenge with the clinical antibiotic ciprofloxacin, and place the measured proteomic changes into biological context.

## References

- (1) Dieterich, D. C.; Link, A. J.; Graumann, J.; Tirrell, D. A.; Schuman, E. M. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 9482–9487.
- (2) Ngo, J. T.; Champion, J. A.; Mahdavi, A.; Tanrikulu, I. C.; Beatty, K. E.; Connor, R. E.; Yoo, T. H.; Dieterich, D. C.; Schuman, E. M.; Tirrell, D. A. *Nat. Chem. Biol.* **2009**, *5*, 715–717.
- (3) Dieterich, D. C.; Hodas, J. J. L.; Gouzer, G.; Shadrin, I. Y.; Ngo, J. T.; Triller, A.; Tirrell, D. A.; Schuman, E. M. *Nat. Neurosci.* **2010**, *13*, 897–905.
- (4) McClatchy, D. B.; Ma, Y.; Liu, C.; Stein, B. D.; Martínez-Bartolomé, S.; Vasquez, D.; Hellberg, K.; Shaw, R. J.; Yates, J. R. *J. Proteome Res.* **2015**, *14*, 4815–4822.
- (5) Yuet, K. P.; Doma, M. K.; Ngo, J. T.; Sweredoski, M. J.; Graham, R. L. J.; Moradian, A.; Hess, S.; Schuman, E. M.; Sternberg, P. W.; Tirrell, D. A. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112*, 2705–2710.
- (6) Erdmann, I.; Marter, K.; Kobler, O.; Niehues, S.; Abele, J.; Müller, A.; Bussmann, J.; Storkebaum, E.; Ziv, T.; Thomas, U.; Dieterich, D. C. *Nat. Commun.* **2015**, *6*, 7521.
- (7) Sinai, L.; Rosenberg, A.; Smith, Y.; Segev, E.; Ben-Yehuda, S. *Mol. Cell* **2015**, *57*, 695–707.
- (8) Bagert, J. D.; van Kessel, J. C.; Sweredoski, M. J.; Feng, L.; Hess, S.; Bassler, B. L.; Tirrell, D. A. *Chem. Sci.* **2016**, *7*, 1797–1806.
- (9) Mahdavi, A.; Szychowski, J.; Ngo, J. T.; Sweredoski, M. J.; Graham, R. L. J.; Hess, S.; Schneewind, O.; Mazmanian, S. K.; Tirrell, D. A. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 433–438.
- (10) Grammel, M.; Dossa, P. D.; Taylor-Salmon, E.; Hang, H. C. *Chem. Commun.* **2012**, *48*, 1473–1474.
- (11) Yuet, K. P.; Tirrell, D. A. *Ann. Biomed. Eng.* **2014**, *42*, 299–311.

*Chapter 2*

PROTEOMIC ANALYSIS OF *P. AERUGINOSA* DURING  
ANAEROBIC DORMANCY AND THE DISCOVERY OF SUTA,  
A SLOW-GROWTH TRANSCRIPTION FACTOR.

**2.1 Summary of Contributions**

Published as:

- (1) Babin, B. M.; Bergkessel, M.; Sweredoski, M. J.; Moradian, A.; Hess, S.; Newman, D. K.; Tirrell, D. A. *Proc. Natl. Acad. Sci. U.S.A.* **2016**, *113*, E597–E605.

This work was performed in close collaboration with Dr. Megan Bergkessel. I performed the BONCAT proteomic screen but the efforts for the majority of the follow-up experiments were shared by me and Megan, as indicated by our co-first authorship of the associated publication. Contributions that were primarily my own include the synthesis of chemical probes, BONCAT labeling and enrichment, protein preparation and analysis of LC-MS/MS experiments, phenotypic screens of transposon mutants for biofilm formation, and immunoprecipitation and chromatin immunoprecipitation experiments. Megan's primary contributions include liquid chromatography measurements of phenazine concentrations, microscopy and analysis of the survival competition experiment, GFP fluorescence measurements, quantitative PCR measurements of transcript abundances, next generation sequencing and data analysis for RNA-Seq and ChIP-Seq experiments. We shared the following work equally: generation of plasmids and mutant strains, experimental planning, data interpretation, and writing of the manuscript.

## 2.2 Abstract

Microbial quiescence and slow growth are ubiquitous physiological states, but their study is complicated by low levels of metabolic activity. To address this issue, we used a time-selective proteome labeling method (bio-orthogonal non-canonical amino acid tagging; BONCAT) to identify proteins synthesized preferentially, but at extremely low rates, under anaerobic survival conditions by the opportunistic pathogen *Pseudomonas aeruginosa*. One of these proteins is a transcriptional regulator that has no homology to any characterized protein domains, and is post-transcriptionally upregulated during survival and slow growth. This small, acidic protein associates with RNA polymerase and chromatin immunoprecipitation (ChIP) followed by high-throughput sequencing suggests that it associates with genomic DNA through this interaction. ChIP signal is found both in promoter regions and throughout the coding sequences of many genes, and is particularly enriched at ribosomal protein genes and in the promoter regions of ribosomal RNA genes. Deletion of the gene encoding this protein affects expression of these and many other genes, and impacts biofilm formation, secondary metabolite production, and fitness in fluctuating conditions. On the basis of these observations, we have designated the protein SutA (survival under transitions).

### 2.3 Introduction

*Pseudomonas aeruginosa* is a cosmopolitan bacterium, notorious as a dominant opportunistic pathogen of burn wounds, medical devices, and the lungs of cystic fibrosis (CF) patients. Its genome is large and encodes an unusually high proportion of regulators [1]. Compared to *Escherichia coli*, *P. aeruginosa* possesses more  $\sigma$  factors that direct RNA polymerase (RNAP) to promoter regions (24 vs. 7), more DNA-binding activators and repressors that enhance or prevent RNAP binding and transcription (550 vs. 150) [2, 3], and more small, noncoding RNAs (ncRNAs) that modulate the stability or translation of target transcripts (200 vs. 100) [4, 5]. Much effort has been directed toward understanding the mechanisms by which this regulatory capacity governs the behaviors—such as quorum sensing, protein secretion, secondary metabolite production, and biofilm formation—that contribute to *P. aeruginosa* virulence.

The physiological states of bacteria involved in chronic infections are substantially different from those most often studied in standard laboratory experiments; chronic infections are characterized by slow growth rates imposed by limited nutrients or oxidants, or by host immune responses. Direct measurements of *in situ* microbial growth rates in the context of lung infections in CF patients have revealed doubling times of several days [6]. Measurements of expectorated sputum show that hypoxic and anoxic zones exist within infected CF airways, and can experience dramatic fluctuations in redox potential [7]; *P. aeruginosa* strains isolated from the CF lung show gene expression patterns consistent with adaptations to hypoxia [8], suggesting that a lack of oxygen may limit growth. While *P. aeruginosa* can generate energy in this environment by using nitrate as the terminal electron acceptor for respiration [9], levels of nitrate may be too low or too variable for this to represent the sole energy source in anoxic zones. *P. aeruginosa* can also remain viable for weeks in an anaerobic survival state by carrying out substrate-level phosphorylation to generate ATP, using either pyruvate (assisted by phenazines [10] or arginine as a carbon and energy source [11, 12]. The cells do not grow when limited to this type of metabolism, and little is known about how basic cellular processes are maintained.

We explored the *P. aeruginosa* anaerobic survival state by identifying the proteins that are synthesized in this energy-limited condition. Previous studies have characterized transcriptomic responses to low oxygen [13, 14] and have identified a few proteins that increase in abundance under conditions of anaerobic survival [15]. The potential for important post-transcriptional regulation under stress conditions [16,

17] led us to take a proteomic approach, and the low metabolic rates that occur during anaerobic survival meant that the quantity of protein made after the shift to anaerobic conditions would likely be small relative to the size of the pre-existing proteome. To address these challenges and specifically identify proteins associated with the anaerobic survival state, we used a time-selective proteome-labeling approach, referred to as bio-orthogonal noncanonical amino acid tagging (BONCAT) [18, 19] to enrich and identify proteins made during anaerobic survival. We identified 91 proteins that were preferentially synthesized under anaerobic survival conditions compared to aerobic growth conditions in the same medium. Phenotypic screens of mutants lacking these proteins led us to focus on a single uncharacterized protein that is expressed under multiple slow-growth conditions and plays a role in biofilm formation, virulence factor production, and survival under transitions between different conditions. We used a combination of co-immunoprecipitation, mass spectrometry, and sequencing to establish this protein as a novel regulator of transcription. The protein binds RNA polymerase, causes widespread changes in gene expression, and plays a direct role in the regulation of genes encoding ribosomal components.

## 2.4 Results

### **BONCAT Enables Enrichment and Identification of Proteins Synthesized at Low Rates During Anaerobic Survival**

The BONCAT technique relies on pulse-labeling cultures with the methionine (Met) surrogate azidohomoalanine (Aha) (Figure 2.S1A), which is incorporated into nascent proteins by a cell's endogenous translational machinery. Aha provides a chemical handle by which newly synthesized proteins can be distinguished and physically enriched from the pre-pulse proteome (Figure 2.S1B). To probe protein synthesis during anaerobic survival on arginine, we shifted an aerobic arginine culture to anaerobic conditions, allowed cells to adapt for 24 h, and then treated them with Aha (Figure 2.1A). The total amount of incorporation of Aha into cellular protein during a 16 h pulse was approximately 4-fold lower than that observed for an aerobic sample treated for only 15 min (Figure 2.1B, Figure 2.S1C-D), providing evidence of slow, but detectable, protein synthesis during anaerobic survival. Lysates from anaerobic and aerobic cultures were treated with an alkyne-biotin affinity tag, enriched for Aha-labeled proteins with streptavidin beads (Figure 2.S1F), and analyzed by liquid chromatography-tandem mass spectrometry (LC-MS/MS).

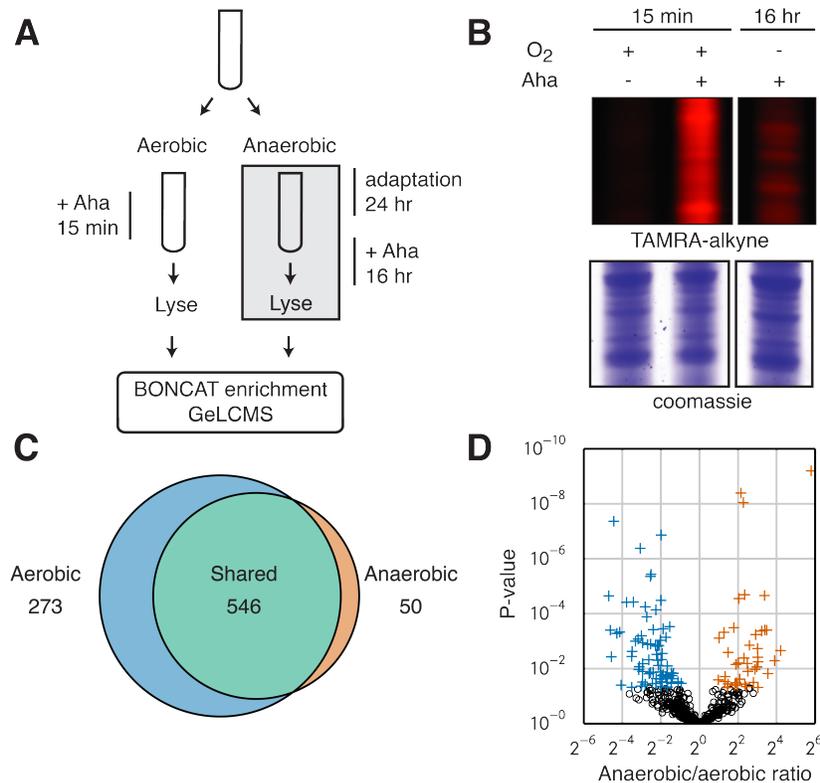
We identified 869 proteins overall; 50 were detected only in the anaerobic sample and 273 were detected only in the aerobic sample (Figure 2.1C). For the 546 pro-

teins identified in both samples, we used label-free quantification to find proteins preferentially synthesized under each set of conditions. Peptide intensities were normalized to the total peptide intensity for each run, and the ratio for each protein was calculated as the median of its peptide ratios. We found 41 and 74 proteins whose anaerobic:aerobic ratios were significantly greater than or less than 1, respectively (Figure 2.1D). Complete proteomic results are listed in Dataset A.1. The 91 proteins that were more abundant or detected only in the anaerobic sample included proteins previously implicated in anaerobic growth or survival, such as targets of the oxygen-sensing regulator Anr: NirM, CcpR, UspK, PctA, and PA14\_06000 [14, 15]. More than one third, however, are annotated as “hypothetical proteins.” We hypothesized that this list of “anaerobic hits” might contain poorly characterized proteins that play important roles in regulating slow-growth physiology. To identify general regulators, we tested the ability of transposon mutants of these genes (from a mutant library [20]) to form biofilms—another growth condition in which nutrients and oxygen are limited and cells experience low metabolic rates [21].

We looked for defects in two modes of biofilm growth: as attached biofilms on a polystyrene substrate and as colony biofilms on agar plates (Figure 2.S2A-B). Mutants for three genes showed defects in both biofilm assays: FimV, PA14\_44460, and PA14\_69770. FimV and PA14\_44460 have previously been implicated as contributors to type II secretion—a process known to be important for biofilm formation [22]. In contrast, PA14\_69770 has no homology to any characterized proteins or domains and has not been investigated to date. For this reason, we chose to study further the role of PA14\_69770 in *P. aeruginosa* under survival and slow-growth conditions. Based on its contribution to fitness during transitions to and from these states, uncovered in our studies, we refer to this protein as Suta (survival under transitions).

### **Suta Promotes Biofilm Formation, Inhibits Pyocyanin Production, and Confers a Fitness Advantage under Fluctuating Conditions**

We generated a clean deletion strain ( $\Delta sutA$ ) and an arabinose-inducible overexpression strain ( $P_{ara}:sutA$ ) to verify the results of the biofilm phenotype screens. Arabinose cannot support growth of *P. aeruginosa* when supplied as the sole carbon source, so does not act as a nutrient during induction of gene expression in this context. For all experiments involving arabinose-induced overexpression, arabinose was also added to the wild-type and  $\Delta sutA$  strains to control for any potential physiological impacts. The deletion mutant formed smooth colony biofilms that



**Figure 2.1: BONCAT enables enrichment and identification of proteins synthesized during anaerobic survival.** (A) Overall scheme of the BONCAT experiment. (B) Lysates were treated with TAMRA-alkyne and separated via SDS-PAGE to visualize Aha incorporation. Coomassie staining indicates total protein loading (See 2.S1E for entire gel). (C) Identified proteins fell into three groups: unique to the aerobic sample, shared, and unique to the anaerobic sample. (D) Protein ratios between the two samples were calculated via label-free quantification. Proteins significantly more abundant in each sample (Benjamini-Hochberg FDR,  $p < 0.05$ ) are marked with crosses.

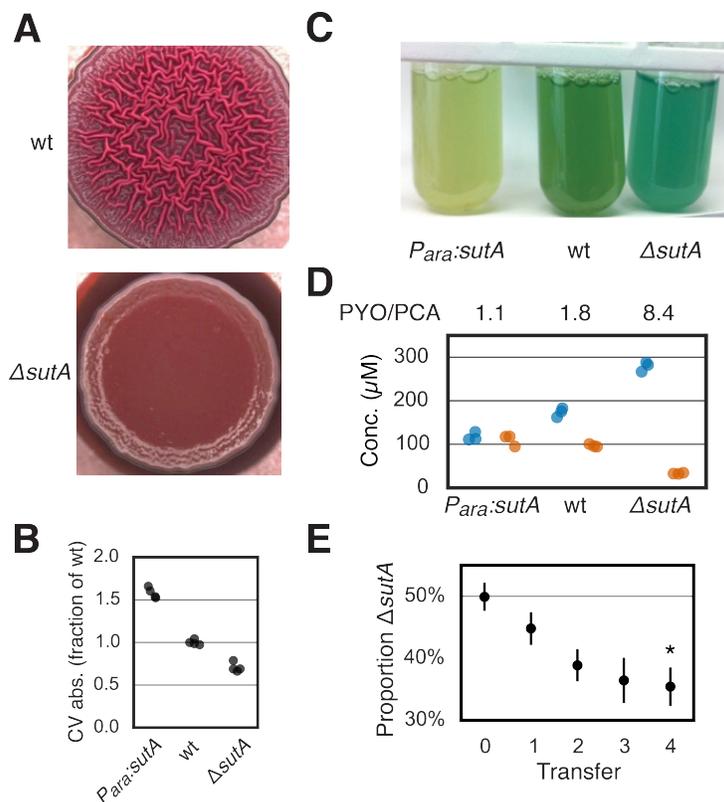
lacked the complex wrinkled structures observed in wild-type biofilms, while the overexpression strain did not show substantially different colony morphology (Figure 2.2A). The deletion strain also formed smaller biofilms, and the overexpression strain larger biofilms, on polystyrene compared to the wild type (Figure 2.2B). The biofilm deficiencies of the mutant strain were not due to a growth defect, as there were no differences in growth rates between  $\Delta sutA$  and the wild-type strain during aerobic planktonic culture in either rich or minimal media (Figure 2.S2C). There was, however, a strong effect of *SutA* on the colors of planktonic cultures;  $\Delta sutA$  cultures were more blue and  $P_{ara}:sutA$  cultures less blue than the wild type. This effect was pronounced under nutrient-poor conditions, following aerobic growth in

minimal medium containing pyruvate as a carbon source (Figure 2.2C). The blue color of high-density *P. aeruginosa* cultures is often due to the presence of the redox-active phenazine pyocyanin (PYO), which plays roles in signaling and virulence, and whose production is sensitive to various regulatory inputs [23–25]. We measured the concentrations of PYO and its metabolic precursor phenazine-1-carboxylic acid (PCA) in culture supernatants using HPLC and found that  $\Delta sutA$  produced more PYO and less PCA than the wild type, while *P<sub>ara</sub>:sutA* showed the opposite effect (Figure 2.2D). Absorbance measurements of culture supernatants gave the same results (Figure 2.S2D).

Because control of biofilm formation and phenazine production relies on integration of multiple regulatory inputs, particularly those related to changes in cell density and nutrient availability, we tested SutA's contribution to the fitness of cells exposed to changing conditions. To detect subtle effects, we competed fluorescently marked wild-type and  $\Delta sutA$  strains while alternating between aerobic growth in LB and anaerobic survival in minimal arginine medium. On average, the wild-type strain significantly outcompeted  $\Delta sutA$  after four transitions (Figure 2.2E), and in five out of six trials, the wild-type strain showed a clear advantage after two transitions (Figure 2.S2E), suggesting that SutA is important during transitions to and from the survival state.

### **SutA Upregulation During Slow Growth is Post-transcriptional**

We initially focused on SutA based on its upregulation under anaerobic survival conditions, but its roles in biofilm formation and phenazine production under aerobic conditions suggested that its expression is not solely dependent on anoxia. To assay SutA expression at both the transcript and protein levels, we generated a reporter strain carrying a fusion of the *sutA* promoter, 5' untranslated region (UTR), and 3' UTR to *gfp* (*P<sub>sutA</sub>:gfp*). Both 5' and 3' UTRs have previously been shown to impact transcript stability and translation [26], so our construct was designed to capture effects conferred by both regions. We measured GFP fluorescence per cell using flow cytometry during growth in LB and pyruvate minimal media, starting in mid-exponential phase (which takes longer to reach in pyruvate minimal media than in LB). In LB, reporter protein levels per cell were low during mid- and late-exponential phase (0 to 3 h) but increased up to eightfold in late stationary phase, while transcript levels (shown normalized to the level measured at time 0 in LB) varied less than twofold throughout the experiment (Figure 2.3, solid lines). In pyruvate medium, in which cells grow approximately fourfold slower compared to



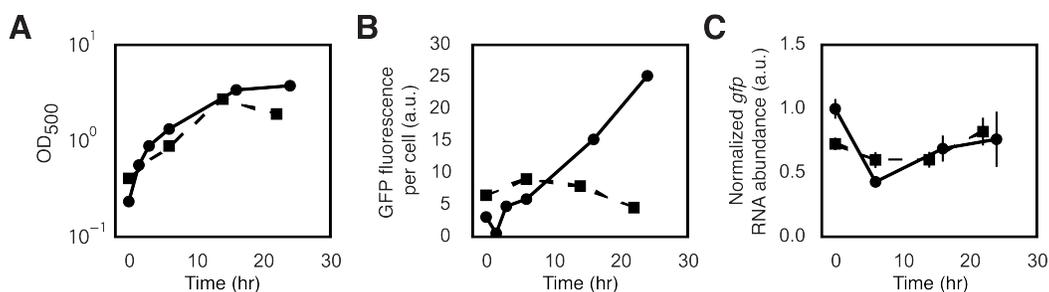
**Figure 2.2: Phenotypic characterization of *sutA* mutants.** (A) Colony biofilms were grown for 6 days at room temperature. (B) Biofilm growth on polystyrene was measured with the Crystal Violet assay ( $n = 4$ ). (C) Cultures were grown in pyruvate minimal medium to stationary phase overnight at 37 °C. (D) Concentrations of PYO (blue) and PCA (orange) in culture supernatants were measured via HPLC. Average molar ratios are indicated above the plot ( $n = 3$ ). (E) Co-cultures of wild-type and  $\Delta sutA$  strains were subjected to repeated rounds of anaerobic survival followed by outgrowth to mid-exponential phase in LB. After each outgrowth, the proportion of  $\Delta sutA$  was measured by fluorescence microscopy. Error bars show standard error ( $n = 6$ ). The asterisk indicates a significant difference from the initial time point (paired Student's t-test,  $p < 0.05$ ).

LB and remain in exponential phase for a longer time (0 to 14 h) (see also Figure 2.S2C), GFP fluorescence per cell was higher than in LB during exponential growth, and increased slightly with culture density before decreasing in late stationary phase. As in LB, normalized transcript levels showed little variation (Figure 2.3, dashed lines).

To verify that changes in fluorescence measurements reflected regulation of transcription and translation and were not due to accumulation of GFP, we constructed an analogous reporter that encoded a fusion of the promoter, 5' UTR, and 3' UTR

of the ribosomal protein gene *rpsG* to *gfp* ( $P_{rpsG}:gfp$ ). As expected, per cell GFP expression was high in exponential phase and decreased sevenfold in stationary phase (Figure 2.S2F-H). In contrast to the *sutA* reporter construct, transcript and protein levels followed the same trend.

These results indicate that SutA upregulation occurs in conditions that cause slow growth, and does not require a lack of oxygen. Because slow growth in pyruvate minimal medium resulted in constitutive moderate expression of SutA and because we could clearly observe a phenazine phenotype resulting from SutA mutation in this medium, we chose to use late exponential phase in pyruvate minimal medium for further study of the functions of SutA.



**Figure 2.3: SutA upregulation during slow growth is post-transcriptional.** A  $P_{sutA}:gfp$  cassette was transposed into a neutral locus of the wild-type strain. (A) Optical density, (B) per-cell GFP fluorescence, and (C) *gfp* transcript abundance were measured throughout growth in LB (circles, solid lines) and pyruvate minimal medium (squares, dashed lines). Error bars represent the standard error of biological replicates ( $n = 3$ ), and in some cases are smaller than the marker. RNA abundances were normalized by *oprI*. RNA and GFP measurements are relative to the value for the  $P_{sutA}:gfp$  strain in LB at time 0.

### SutA Interacts with RNA Polymerase

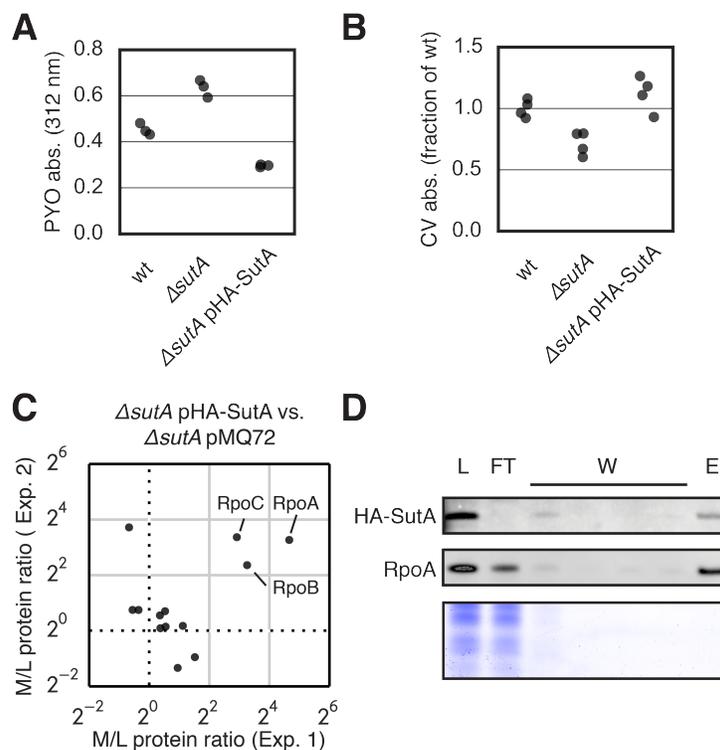
To gain insight into how SutA brings about the observed phenotypic changes, we sought to identify interacting protein partners. We generated an N-terminal hemagglutinin-tagged copy of SutA (HA-SutA), and verified that expression of this protein from the pMQ72 plasmid backbone in the  $\Delta sutA$  background complemented the phenazine (Figure 2.4A) and biofilm (Figure 2.4B) phenotypes. We performed an immunoprecipitation (IP) against the HA epitope in this strain and in the  $\Delta sutA$  strain carrying the empty pMQ72 vector following induction with arabinose in late exponential phase in pyruvate minimal medium. We identified co-precipitating proteins via LC-MS/MS analysis of the eluent fraction. Proteins co-precipitated with HA-SutA or from the empty vector control were digested with trypsin and

reacted with “medium” or “light” dimethyl labels, respectively. Peptides from both immunoprecipitations were mixed and ratios directly quantified by LC-MS/MS. In two experiments, we identified three proteins that were enriched at least fivefold in the strain expressing HA-SutA compared to the empty vector control: the  $\alpha$ ,  $\beta$ , and  $\beta'$  subunits of RNAP (RpoA, RpoB, and RpoC) (Figure 2.4C). We also detected co-precipitation of RpoA with HA-SutA in the IP eluent fraction by Western blot (Figure 2.4D). The presence of some RpoA signal in the unbound (“FT”) fraction suggests that not all cellular RNAP is tightly bound by SutA under the condition tested. We also performed the experiment in reverse by immunoprecipitating RNAP from the same cell lysates with an anti-RpoA antibody and identifying co-precipitated proteins via LC-MS/MS. When co-precipitated proteins were ordered by total peptide intensities, HA-SutA ranked above known RNAP-binders  $\sigma^{70}$ , NusA, and Rho (Figure 2.S3, Dataset A.2).

### **SutA Associates with Genomic Loci and Enhances Transcription of Ribosomal Genes**

To investigate the context of the interaction between SutA and RNAP and the effects it might have on gene expression, we performed a chromatin immunoprecipitation (ChIP) -Seq experiment and an RNA-Seq experiment. The ChIP-Seq experiment was performed with the same strains and conditions used to detect the interaction with RNAP: the  $\Delta sutA$  strain carrying HA-SutA on the pMQ72 arabinose-inducible plasmid and the  $\Delta sutA$  strain carrying the pMQ72 empty vector as a control, both grown to late exponential phase in pyruvate minimal medium in the presence of arabinose. We cross-linked protein–DNA complexes with formaldehyde, sonicated chromosomal DNA to generate fragments 0.5 to 1 kb in length, performed immunoprecipitations against the HA epitope or against RpoA, and sequenced the co-precipitated DNA. For the RNA-Seq experiment, we sequenced rRNA-depleted RNA extracted from the wild-type,  $\Delta sutA$ , and  $P_{ara}:sutA$  strains using the same growth medium and time point as for the ChIP-Seq experiment.

Because our IP experiment suggested that not all cellular RNAP was associated with SutA, we first sought to determine whether the interaction between SutA and RNAP occurs while RNAP is engaged in transcription, which should result in efficient formaldehyde crosslinking of SutA to genomic DNA, through concurrent interactions with RNAP. Immunoprecipitation of HA-SutA led to an average recovery of 4% of input DNA compared to 0.2% in IPs from the empty vector control strain that did not encode HA-SutA (Figure 2.S4 A), indicating that SutA likely interacts with



**Figure 2.4: RNA polymerase co-precipitates with Suta.** (A) Absorbance measurements of culture supernatants and (B) Crystal Violet (CV) measurements of biofilm formation. (C) LC-MS/MS detection and quantification of proteins co-immunoprecipitated with HA-SutA. Each axis represents the protein abundance ratio as determined by dimethyl quantification between proteins co-precipitated from the pHA-SutA (medium; M) or pMQ72 control (light; L) strains. The three main subunits of RNAP are indicated. (D) Immunoprecipitation fractions were analyzed for the presence of HA-SutA and RpoA via Western blots and for total protein via Coomassie staining (bottom). L: lysate, FT: flow-through, W: washes, E: eluent.

RNAP while RNAP is interacting with genomic DNA. Over 1,400 of the approximately 6,200 annotated genes showed a statistically significant enrichment in the HA-SutA IP compared to the empty vector IP, though the enrichment was greater than twofold for only 85 genes (Dataset A.3). We next assessed the relationship between Suta and RNAP occupancies at genomic loci by comparing average per-gene reads per kilobase per million reads mapped (rpkm) from each IP. We saw a moderately strong correlation between the associations of Suta and RpoA across all genes (Figure 2.5A, Pearson's  $r = 0.77$ ), suggesting that Suta and RNAP tend to co-localize throughout the chromosome. This degree of correlation with RNAP ChIP signal is similar to what has been observed for NusG in *E. coli* ( $r = 0.86$ ) and GreA in *Bacillus subtilis* ( $r = 0.86$ ), both of which bind RNAP during transcription

elongation [27, 28]. When the ChIP data were divided into 100 bp tiles across the entire chromosome, the correlation between RNAP signal and HA-SutA signal had an  $r$  value of 0.66, which is lower than the value previously calculated in *E. coli* for DksA ( $r = 0.79$ ) but higher than that for  $\sigma^{70}$  ( $r = 0.57$ ), which dissociates from polymerase prior to transcription elongation [29]. We noted that a subset of genes had ratios of SutA ChIP signal to RpoA ChIP signal that were substantially higher than the mean for all genes, and found that many of these genes encoded ribosomal proteins (Figure 2.5A-B).

We next asked whether RNAP association at genomic loci was affected by the presence of SutA. We compared average per-gene ChIP signals for RpoA between the strain expressing HA-SutA and the strain carrying the empty vector. We found a very high correlation in per-gene RpoA ChIP signals between these two strains (Figure 2.S4B, Pearson's  $r = 0.94$ ), suggesting that changes in the distribution of polymerase caused by the presence of SutA are subtle, or limited to a small number of loci. Although the differences in rpkm per gene were not statistically significant on an individual gene basis, we did note some departures from the overall high correlation. In particular, both ribosomal RNA (rRNA) and transfer RNA (tRNA) loci tended to show higher RpoA ChIP signals in the strain expressing HA-SutA compared to the strain lacking SutA (Figure 2.5C, Figure 2.S4D).

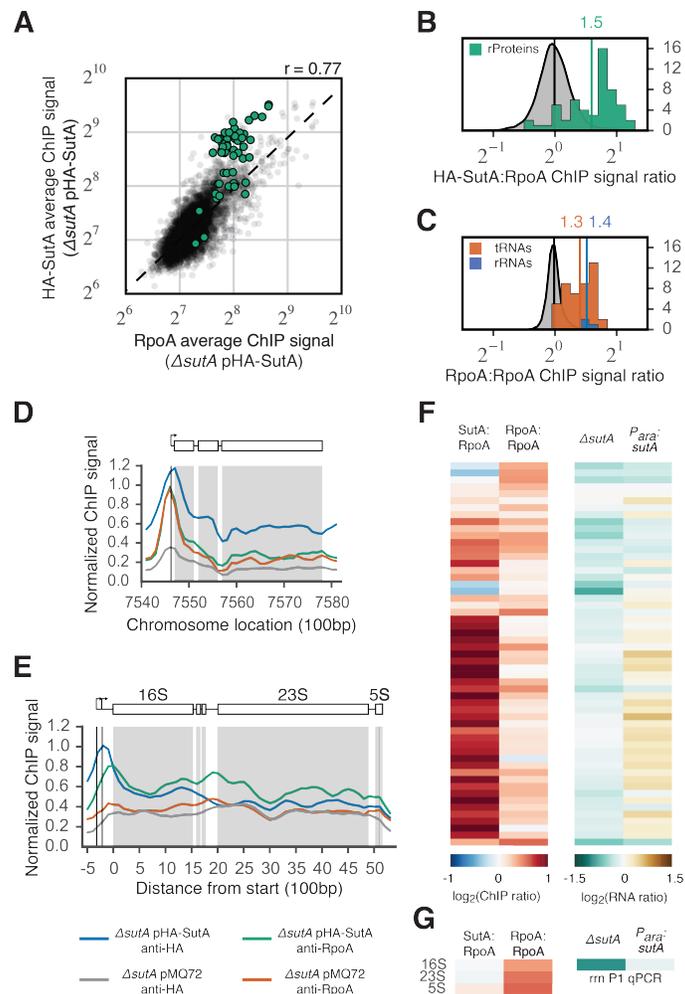
To establish a higher-resolution view of SutA and RNAP associations at ribosomal protein and rRNA loci, we examined ChIP-Seq reads per 100 bp tile across the relevant loci. We adapted the “apparent occupancy” metric described previously for displaying ChIP-chip data [27]. Because some non-specific immunoprecipitation of DNA is expected, the normalized read counts observed at the least expressed genes in the genome were used to define a baseline signal representing no true occupancy, and the counts observed at the highest peaks in each sample that were associated with protein coding genes were used to define a maximum signal for that sample. All count values in each sample were then scaled from 0 to 1 based on the calculated baseline and maximum values for that sample. The count values for the IP from the empty vector strain are included for comparison, and are scaled to the baseline and maximum values calculated for the HA-SutA IP to best facilitate the comparison (the dynamic range for the empty vector IP was small, as expected for a control IP in which association is non-specific) (see Supplemental Experimental Methods and Datasets A.4 and A.5 for more information).

Ribosomal protein loci exhibited distinct peaks in RNAP and SutA signal near their

transcription start sites (Figure 2.5D, Figure 2.S4C). The SutA peak was shifted very slightly downstream from the RpoA peak, and the ratio of SutA signal to RpoA signal was high over promoter and coding regions, consistent with what was observed in the per-gene analysis. The presence of SutA did not result in a significant difference in RpoA signal at any individual ribosomal protein gene locus, but across all ribosomal protein genes, there appears to be a trend toward increased RpoA signal in the presence of SutA (Figure 2.5F). Because the sequences of the four rRNA operons are nearly identical, these loci were aligned and the signals for homologous 100bp tiles from each operon were averaged (Figure 2.5E). While the rRNA genes did not show high levels of HA-SutA ChIP signal relative to RpoA ChIP signal in our per-gene analysis, this higher-resolution view shows that a very strong peak of SutA signal is centered just upstream of the start of the 16S gene, near the predicted P2 transcription start site, with a lower ratio of SutA to RpoA signal across the coding region. This view also shows a statistically significant increase in the RpoA signal at the rRNA promoter region in the presence of SutA, which was missed in our per-gene analysis. These two features are distinct from the observations for the ribosomal protein loci.

We then investigated whether the presence of SutA at ribosomal protein and rRNA genomic loci, and the changes in RNAP localization to rRNA in particular, might impact their expression. To assess the effects of SutA on ribosomal protein gene mRNA levels, we queried our RNA-Seq dataset. We measured small but statistically significant differences in mRNA abundance among the three strains for a majority of the ribosomal protein genes (46 of 55 genes, FDR-adjusted p-value < 0.05) (Dataset A.3). In general, they were expressed at higher levels in the *Para:sutA* strain, and at lower levels in the  $\Delta sutA$  strain, compared to the wild-type strain (Figure 2.5F). Because the stability of mature ribosomal RNA makes it a poor indicator of rRNA transcription rates, and because rRNA was intentionally depleted from our RNA-Seq samples prior to library construction, we used qPCR against the 16S leader sequence as a proxy for levels of new rRNA synthesis. The  $\Delta sutA$  strain had levels of the 16S leader that were twofold lower compared to either the wild-type strain or the overexpression strain (Figure 2.5G, Figure 2.S4E). Taken together, the ChIP and RNA abundance measurements suggest that the presence of SutA has a direct and positive effect on the transcription of both ribosomal protein and ribosomal RNA genes, but that the nature of the interactions with these two types of loci may be distinct. Extensive work by many laboratories (reviewed in [30]) has shown that regulation of rRNA transcription occurs primarily at the level of initiation while

regulation of ribosomal protein gene transcription occurs mostly during elongation. Consistent with this regulatory paradigm, our ChIP data suggest association of SutA primarily in the promoter regions of rRNA genes but throughout the coding regions of ribosomal protein genes. Also potentially consistent with these two modes of regulation, we see a decrease in RpoA ChIP signal in the absence of SutA for rRNA genes but much less so for ribosomal protein genes. Further study will be required to elucidate the mechanistic details of these two possible regulatory modes.



**Figure 2.5: SutA localizes throughout the chromosome and enhances transcription of ribosomal genes.** (A) ChIP signals (rpkm) for HA-SutA vs. RpoA for each gene. Genes encoding ribosomal proteins are highlighted (green) (Pearson's  $r = 0.77$ ). (B) The distribution of HA-SutA:RpoA ChIP signal ratios from the  $\Delta sutA$  pHA-SutA strain for all genes (gray probability density plot) and for ribosomal protein genes (green histogram). (C) The distribution of the ratios of RpoA ChIP signal from  $\Delta sutA$  pHA-SutA vs.  $\Delta sutA$  pMQ72 for all genes (gray probability density plot), tRNAs (orange histogram), and rRNAs (blue histogram). The mean ratios for each subset are indicated above. (D, E) Normalized ChIP signals from each IP at the *rpsLG-fusA1* ribosomal protein operon (D) and for ribosomal RNA operons (E). Legend describing strains and IPs for each trace is below. (F, G) Heat maps for ribosomal protein genes (F) and rRNA (G) showing ChIP signal ratios as calculated in (B) and (C) and transcript abundance ratios for  $\Delta sutA$  and  $P_{ara}::sutA$  strains, each compared to the wild-type strain as determined by RNA-Seq (F) or qPCR (G).

### **SutA Localizes to Many Non-ribosomal Genes and Has Broad Effects on Gene Expression**

Ribosomal proteins and rRNAs are notable as classes of genes that had high levels of SutA association and whose transcript levels were significantly changed. However, the influence of SutA was not limited to these loci; much of the chromosome (approximately 20% of all 100 bp regions) showed statistically significant enrichment for the HA-SutA IP compared to the empty vector IP. To explore the general pattern of association of SutA with genomic loci, we identified a “high ChIP signal” subset of 230 transcriptional units that (i) had high-quality peaks in both RpoA and SutA ChIP signals near their starts (defined as having an apparent occupancy greater than 0.25 for RpoA and 0.20 for SutA) and (ii) showed a statistically significant enrichment in the HA-SutA ChIP signal compared to the empty vector ChIP signal. For those that had annotated transcriptional start sites and were not among the ribosomal protein and RNA genes discussed above ( $n = 171$ ), we averaged ChIP signal values from 500 bp upstream to 1000 bp downstream of that location to generate aggregate traces of the associations of RNAP and HA-SutA across non-ribosomal loci (Figure 2.6A). The average pattern of RpoA and SutA association across these transcriptional units was similar to that observed for the ribosomal protein genes: RpoA association was centered at the transcriptional start site and a broader peak of HA-SutA was centered slightly downstream. This aggregate includes upstream regions that drive transcription of diverging transcription units as well as those for which adjacent transcription units are on the same strand, so the breadth of the observed peaks may reflect limits of the resolution of our ChIP technique as well as contributions from binding to adjacent transcriptional units.

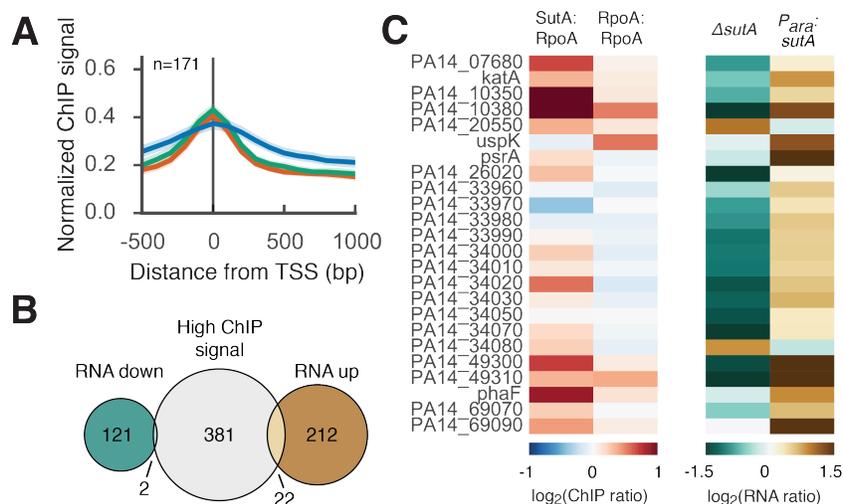
We next investigated whether SutA association at non-ribosomal transcriptional units was also associated with increased expression. To focus on likely direct effects, we examined the 24 genes that were among the “high ChIP signal subset” and also showed greater than two-fold changes in transcript levels. 22 of these (92%) had higher transcript levels in the overexpression strain than in the deletion strain (Figure 2.6B-C), suggesting, as was observed for the ribosomal protein and rRNA genes, that the presence of SutA at these genomic loci tends to enhance their transcription. Higher-resolution views of specific loci reinforced the observations from the aggregate analysis: transcription units exhibited a broad peak of HA-SutA association centered downstream of the peak of RpoA association. PA14\_10380 is predicted to encode a protein that is structurally similar to bacteriocins and is among the highest ranked-genes both in terms of SutA association and differential expres-

sion between the  $\Delta sutA$  and the  $P_{ara}:sutA$  strains (Figure 2.S4F) [31]. PA14\_21220 encodes the universal stress protein UspK (Figure 2.S4G), and PA14\_26020 encodes an aminopeptidase (Figure 2.S4H). In each of these cases, the apparent occupancy of RpoA in the promoter region is higher in the SutA-containing strain.

Many of the genes that were differentially expressed in the SutA mutants were not among the genes that showed the highest ChIP signal and many genes that had high ChIP signal did not show large SutA-dependent changes in gene expression (Figure 2.6B). This pattern is likely due to several factors. First, because the presence of SutA generally enhances transcription at loci to which it is recruited, decreased expression in the presence of SutA may be due largely to the shift of free RNAP to highly expressed loci that are upregulated by SutA (e.g., rRNA). Our data show several transcriptional units that recruit significantly more RNAP in the absence of SutA (as evidenced by higher RpoA ChIP peaks in the strain lacking HA-SutA, and no significant SutA association in the HA-SutA ChIP experiment) and that have increased expression in the  $\Delta sutA$  strain; PA14\_40800 and PA14\_40100-40110, divergently transcribed, are two examples (Figure 2.S4I). Second, the list of genes that are likely directly regulated by SutA includes the components of the ribosome as well as known master regulators such as the stationary phase transcription factor *psrA* [32]. Increased expression of these genes is likely to cause widespread secondary effects, which may explain why some genes that are upregulated in the presence of SutA do not show strong HA-SutA ChIP signal. Third, as suggested by our analysis of rRNA and ribosomal protein genes, SutA may affect different aspects of transcription for different genes (e.g., initiation vs. elongation), with different patterns of ChIP signals and expression levels resulting. Further work is required to fully understand the impacts of SutA on different genes and different phases of gene expression.

Finally, to take a broad view of the effects of SutA, both direct and indirect, on the physiological state of the cell, we grouped the genes that differed more than twofold between the  $\Delta sutA$  and the  $P_{ara}:sutA$  strains according to their functional designations from the Clusters of Orthologous Groups (COG) categories [33], and asked whether any groups were differentially represented compared to the genome as a whole (Figure 2.S4J). In general, genes that were upregulated in the presence of SutA tended to have functions related to energy generation and maintenance; these genes included proteases, oxidoreductases, and alternate metabolism genes. Conversely, genes involved in growth and carbohydrate and amino acid metabolism

were significantly underrepresented. Genes that were downregulated were more likely to be involved in defense mechanisms, signaling, and motility. For the full set of results, see Dataset A.3 and GEO accession number GSE66181.



**Figure 2.6: SutA has broad effects on gene expression.** (A) Average ChIP signals around transcriptional start sites (TSS) for genes in the “high ChIP signal” subset. Shaded regions around each trace represent the 95% confidence interval for the mean ( $n = 171$ ). Traces represent:  $\Delta sutA$  pHA-SutA, anti-HA (blue);  $\Delta sutA$  pHA-SutA, anti-RpoA (green); and  $\Delta sutA$  pMQ72, anti-RpoA (orange). The direction of transcription is from left to right. (B) Numbers of genes in the “high ChIP signal” subset and genes whose expression changed more than twofold between the  $\Delta sutA$  and  $P_{ara}:sutA$  strains. (C) Heat maps (as in Figure 2.5F-G) for genes found in both subsets.

## 2.5 Discussion

While microbes have spent the majority of their evolutionary history enduring slow-growth conditions, relatively little is known about their physiology in these states. In part, this knowledge gap arises from technical challenges—slow metabolic rates and high phenotypic heterogeneity can lead to increased noise and decreased signal for many biomolecules of interest. Yet slow growth and survival states are of great relevance in many clinical and environmental contexts, and new tools are needed for their study. As illustrated here, the BONCAT method, which enables enrichment of newly synthesized proteins from large pre-existing proteomes, is well suited to the exploration of slow-growth modes of microbial life.

We used the BONCAT method to discover a previously unknown RNAP-binding factor, which we have named SutA. We found SutA to be upregulated post-transcriptionally in various growth limiting conditions. Through its interaction with RNAP, SutA lo-

calizes to many genes throughout the chromosome and elicits broad transcriptional changes. Some of these changes are likely direct effects; for example, SutA associates strongly with loci encoding ribosomal components and the transcription of these loci is reduced in the absence of SutA. Other changes may be due to secondary effects resulting from changes in the pool of free polymerase or from changes in downstream regulation by directly affected genes. Our broad analysis of transcriptional changes suggest that cells expressing SutA prioritize the expression of genes required for survival, and our phenotypic studies show that SutA is important for the establishment of biofilms, the regulation of phenazine production, and transitions to and from growth-limited states.

Understanding the molecular mechanism by which SutA effects these changes will require further study, but our observations suggest some intriguing comparisons to the well-studied regulator, DksA. DksA acts with the small molecule alarmone ppGpp during nutritional downshifts to destabilize open promoter complexes, especially at ribosomal RNA promoters. This activity reduces rRNA transcription in response to a decreased availability of nucleotides [34]. DksA can also influence elongation; it may help prevent the transition from a paused to an arrested state [35]. Interestingly, SutA appears to affect many of the same genes and phenotypes as DksA, but in the opposite direction. While DksA has been shown in both *E. coli* and *P. aeruginosa* to repress expression of ribosomal protein and rRNA genes [34, 36, 37], SutA enhances expression of these genes. Both DksA and SutA show high ChIP signal across the coding regions of highly expressed protein-coding genes, including ribosomal protein genes, and a lower signal across the coding regions of the rRNA genes. However, unlike DksA, SutA shows a high peak of ChIP signal at the promoters of rRNA genes, consistent with the observations that SutA enhances rRNA expression while DksA represses it [29]. Disruptions of DksA or SutA in *Pseudomonas* species also appear to cause opposing phenotypes: disruption of DksA causes a decrease in pyocyanin production and an increase in biofilm persistence [38, 39], while deletion of SutA causes overproduction of pyocyanin and a decrease in biofilm accumulation. Taken together, these observations suggest that a subset of genes, including the ribosomal RNA and ribosomal protein genes, are sensitive to some modulation of RNAP activity, and DksA and SutA tend to modulate this activity in opposite ways.

In our BONCAT experiment, we detected new synthesis of DksA in the aerobic exponential growth condition but not in the anaerobic survival condition. This is

consistent with a previous report that DksA is undetectable by Western blot during stationary phase in *P. aeruginosa* [36] and suggests that the repression by DksA of rRNA and ribosomal protein gene expression is downregulated during protracted slow growth. DksA is advantageous in the context of actively growing cells because it protects against “traffic jams” of stalled RNAP that obstruct the completion of DNA replication [40] and allows limited cellular resources to be directed towards expression of genes important for ameliorating the limitations (e.g., amino acid biosynthetic genes) [41]. However, for cells that are dividing infrequently or not at all, and that are limited for basic energy resources rather than specific metabolites, these functions may be counterproductive. Instead, the most adaptive response may be to maintain transcription, even at low levels, of core machinery in order to retain a capacity for cellular maintenance and to allow for a rapid upregulation of biosynthetic pathways when conditions improve. Our results suggest that SutA contributes to this type of response, and they set the stage for future biochemical and structural studies.

Recent reports have described RNAP-binding regulators that broadly affect transcription in different organisms under a range of conditions, suggesting that this is an important and diverse mode of regulation. For example, the non-essential  $\delta$  subunit of *Bacillus subtilis* RNAP [39] and the recently discovered AtfA from *Acinetobacter spp.* [42] are both small proteins that, like SutA, contain highly acidic domains and broadly impact transcription, but unlike SutA, are expressed during exponential phase. CarD is a mycobacterial protein that has recently been crystallized in a complex with RNAP; unlike SutA it is essential and appears to localize primarily to promoter regions, but like SutA it broadly serves to stimulate transcription. One characteristic of all of these proteins is that they lack homologs in *E. coli*, the model organism from which much of our knowledge of bacterial transcriptional regulatory mechanisms has been derived. Each has a different phylogenetic distribution; SutA is found only in selected families of the Alteromonadales and Pseudomonadales orders of Gammaproteobacteria. This growing body of work, including the results described here, demonstrates that regulation of RNAP is diverse, and even in well-studied, clinically important pathogens, basic regulatory mechanisms governing slow growth remain to be discovered.

## 2.6 Experimental Procedures

For detailed descriptions of all experimental procedures, see Appendix A. Strains and plasmids used are listed in Table A.1.

**Strains and growth conditions.** Rich medium was Luria-Bertani (LB) broth. Minimal medium was phosphate buffered, and contained 40 mM carbon source [10]. In experiments involving *P<sub>ara</sub>:sutA*, all cultures were grown in the presence of 20–25 mM arabinose. Where necessary, plasmids were maintained with the appropriate antibiotics. Aerobic growth was carried out with shaking at 37 °C. Anaerobic survival was carried out in Balch tubes in an anaerobic chamber (Coy, Grass Lake, Michigan) without shaking at 37 °C. Growth for colony morphology assays was carried out at room temperature. Genetic manipulations used standard procedures.

**Biofilm measurements.** Crystal Violet and colony morphology assays were carried out as previously described [43, 44].

**Phenazine measurements.** Phenazine concentrations in culture supernatants were determined by HPLC as previously described [25] or estimated by measuring absorbance at 312 nm.

**Individual gene expression measurements.** Per-cell GFP measurements were made using the Accuri c6 flow cytometer, and RNA measurements were made by qRT-PCR. Primers are listed in Table A.2.

**Proteomics.** BONCAT labeling, chemistry, and enrichment were performed as previously described [45]. Label-free quantitation was used for the initial screen. Relative protein abundances for immunoprecipitations were quantified via dimethyl labeling [46].

**IP and ChIP.** Cultures of  $\Delta sutA$  pMQ72 or  $\Delta sutA$  pMQ72-HA-SutA were grown to late exponential phase in pyruvate minimal medium containing 20 mM arabinose and 50  $\mu$ g/ml gentamicin. HA-SutA or RpoA was purified with anti-HA agarose beads (Thermo Scientific) or protein A/G beads (Santa Cruz Biotechnology) and an anti-RpoA antibody, respectively. Fractions were saved for Western blot analysis and eluents were analyzed via LC-MS/MS. For ChIP, cultures were grown as above, cross-linked with 1% formaldehyde, lysed via sonication, and either HA-SutA or RpoA was immunoprecipitated. Protein digestion and DNA cleanup were performed as previously described [47].

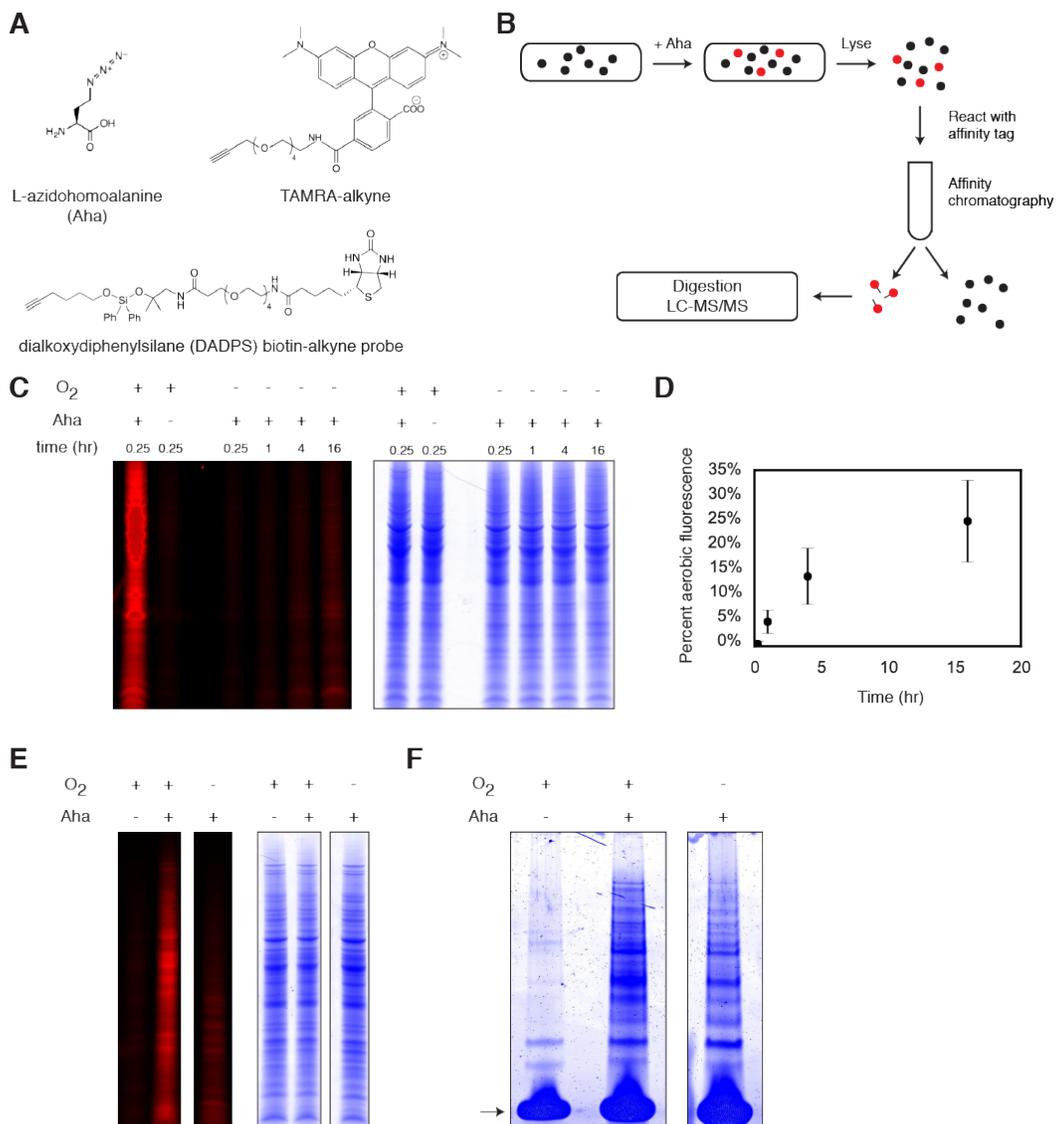
**Sequencing library preparation and sequencing.** For RNA-Seq, cultures of wild-type,  $\Delta sutA$ , and *P<sub>ara</sub>:sutA* strains were grown to late exponential phase in pyruvate minimal medium containing 25 mM arabinose. Total RNA was extracted using the RNeasy Mini Kit (Qiagen) and rRNA was depleted using the Magnetic Gram Neg-

ative Bacteria RiboZero Kit (Epicentre). For ChIP-Seq, immunoprecipitated DNA was further fragmented using DS Fragmentase (NEB). Both types of libraries were prepared using the relevant Library Prep kits for Illumina (NEB). Sequencing was performed to a depth of 10–15 million reads per sample on an Illumina HiSeq2500 machine, and data analysis was performed using standard open source software, or as described in more detail in SI. Sequencing was performed on biological triplicates.

## **2.7 Acknowledgments**

We thank Geoff Smith and Roxana Eggleston-Rangel for technical assistance with LC-MS/MS and Dr. Igor Antoshechkin for assistance with sequencing. We thank Dr. Olaf Schneewind for his gift of the anti-RpoA antibody. We appreciate constructive feedback on the manuscript from members of the Newman and Tirrell labs and Dr. Richard Gourse, as well as helpful comments from the editor and reviewers of *The Proceedings of the National Academy of Sciences*.

## 2.8 Supplementary Figures



**Figure 2.S1: BONCAT labeling and enrichment during anaerobic survival.**

(A) Chemical compounds used for the BONCAT experiment, in-gel fluorescence detection, and protein enrichment. (B) General scheme of a BONCAT experiment. Cells are treated with Aha to initiate protein labeling. Newly synthesized proteins (red circles) are chemically distinct from pre-existing proteins (black circles) and can be reacted with an alkyne-biotin affinity tag. These proteins can be enriched via streptavidin affinity chromatography followed by cleavage of the tag, yielding a mass modification at Aha residues (black lines). Enriched proteins are digested and analyzed by LC-MS/MS. (C) Time course of Aha labeling during anaerobic survival on arginine. Cultures surviving anaerobically were treated with 1 mM Aha for the indicated time. The left two lanes show aerobically growing cultures. In-gel fluorescence of TAMRA (left) indicates Aha incorporation and Coomassie staining (right) indicates total protein loading. Images are of the same gel. (D) Quantification of relative Aha incorporation. Four regions of each lane from the gel in (C) were measured. For each lane, integrated fluorescence intensity was divided by Coomassie intensity to normalize to protein loading. Values from the anaerobic lanes were then divided by the normalized fluorescence from the aerobic culture. Error bars show the standard deviation for 4 regions from each lane. (E) The full gel lanes shown in 2.1B. Images are from the same gel. (F) Eluent fractions following BONCAT enrichment. The three samples shown in (E) were reacted with an alkyne-biotin affinity tag, bound to streptavidin beads, washed, and eluted. Eluents were concentrated and separated via SDS-PAGE. Streptavidin leached from the agarose beads is indicated with an arrow. The right two lanes were cut into eight pieces, digested, and analyzed by LC-MS/MS.

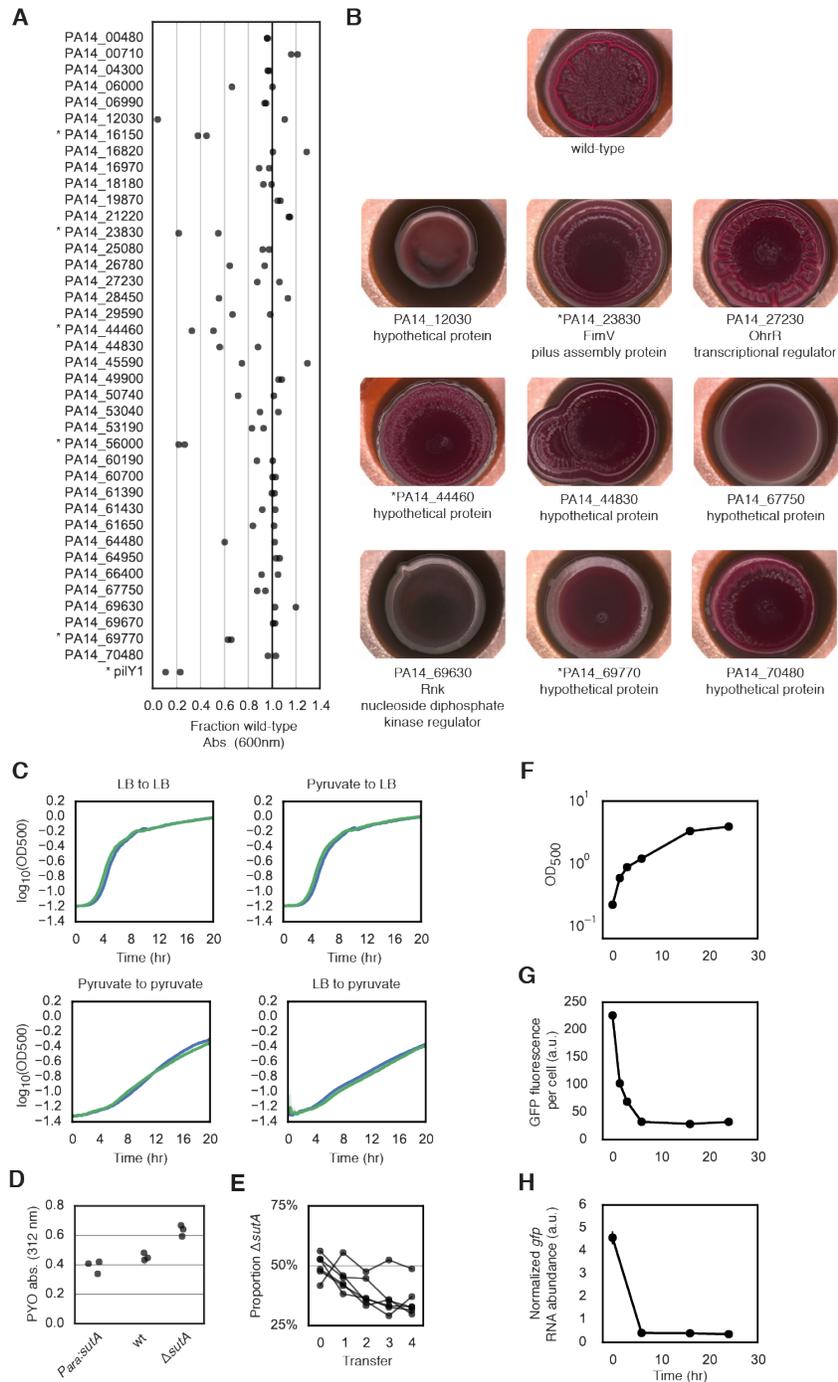


Figure 2.S2: **Phenotype screens and  $\Delta sutA$  growth characterization.** (A) Absorbance of Crystal Violet following biofilm growth on polystyrene well plates. Absorbance values were divided by the value for wild type. Each circle indicates the average value for experiments performed on different days, each with three to four biological replicates. Asterisks indicate mutants whose absorbance ratios were significantly less than 1 in both experiments (Student's t-test,  $p < 0.05$ ). The *pilY1* mutant is a control strain known to have a Crystal Violet screen defect. (B) Transposon mutants that exhibited colony biofilm phenotypes different from the wild-type strain. The phenotype screen was performed in duplicate. Representative images are shown. Mutants that were also defective in the Crystal Violet screen are marked with an asterisk. (C) Growth curves for wild-type (green) and  $\Delta sutA$  (blue) strains in LB or pyruvate minimal medium. Cultures were grown overnight in the first medium and then diluted into the second medium. For dilution into LB, cultures were diluted to an  $OD_{500}$  of 0.001. For dilution into pyruvate, cultures were diluted to an  $OD_{500}$  of 0.005. Each line represents the mean of 8 replicates; 95% confidence intervals for the mean are obscured by the thickness of the lines. (D) Absorbance measurements at 312 nm of culture supernatants from wild-type,  $\Delta sutA$ , and *P<sub>ara</sub>:sutA* strains. (E) Competition assay results for all six individual replicates. (F-H) A *P<sub>rpsG</sub>:gfp* cassette was transposed into a neutral locus of the wild-type strain. (F) Optical density, (G) per-cell GFP fluorescence, and (H) *gfp* transcript abundance were measured throughout growth in LB (circles, solid lines). Error bars represent the standard error of biological replicates ( $n = 3$ ), and in some cases are smaller than the marker. RNA abundances were normalized by *oprI*. RNA and GFP measurements are relative to the value for wild type *P<sub>sutA</sub>:gfp* in LB at time 0 (see Figure 2.3).

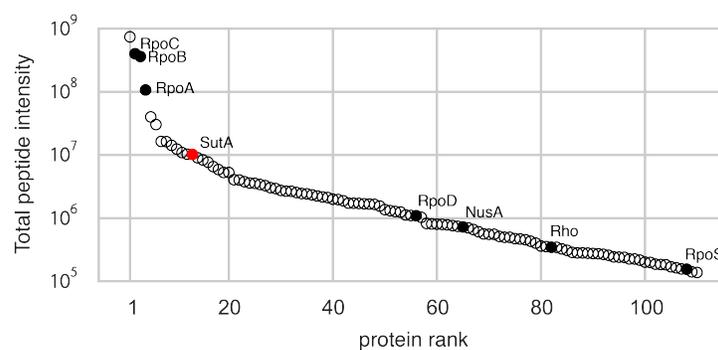


Figure 2.S3: **RpoA co-immunoprecipitated proteins.** Total peptide intensities for proteins that co-precipitated with RpoA. Proteins are ranked by intensity from left to right. The  $\alpha$ ,  $\beta$ , and  $\beta'$  subunits of RNAP (RpoA, RpoB, and RpoC respectively), as well as the sigma factors RpoD and RpoS, the elongation factor NusA, and the termination factor Rho are shown in black. SutA is shown in red.

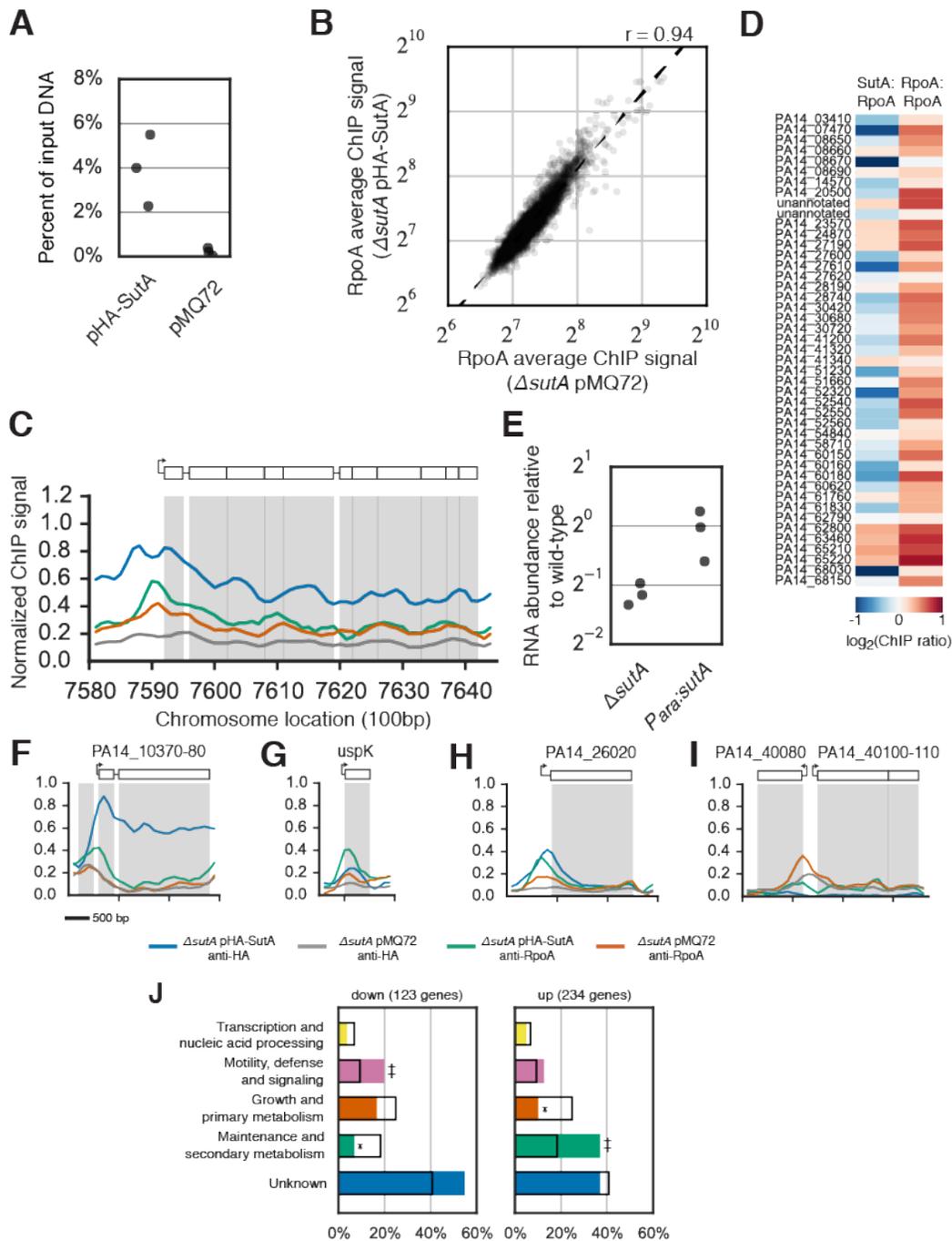


Figure 2.S4: **HA-SutA and RpoA chromatin immunoprecipitation.** (A) DNA yields from chromatin immunoprecipitations against the HA epitope from  $\Delta sutA$  pHA-SutA and  $\Delta sutA$  pMQ72 relative to input DNA were estimated by quantitative PCR for an intergenic region that was not enriched in the HA-SutA ChIP samples. (B) Average RPKM mapped for all genes from the RpoA immunoprecipitations from  $\Delta sutA$  pHA-SutA and  $\Delta sutA$  pMQ72 (Pearson's  $r = 0.94$ ). (C) Normalized and scaled ChIP signals for HA immunoprecipitation from  $\Delta sutA$  HA-SutA (blue) and  $\Delta sutA$  pMQ72 (gray), and for RpoA immunoprecipitation from  $\Delta sutA$  pHA-SutA (green) and  $\Delta sutA$  pMQ72 (orange) across a chromosomal region containing the S10 (*rpsJ*) ribosomal protein operon. (D) ChIP and RNA-Seq results for tRNA genes. Heatmaps show ratios for HA-SutA ChIP RPKM values compared to RpoA ChIP RPKM values from  $\Delta sutA$  pHA-SutA (left column) and RpoA ChIP RPKM values between  $\Delta sutA$  pHA-SutA and  $\Delta sutA$  pMQ72. tRNAs encoded within rRNA operons are excluded. Because many tRNAs have substantial sequence similarity with each other, only sequencing reads that could be mapped uniquely are displayed, and only tRNAs with at least 10 unique RPKM in the RpoA immunoprecipitation from  $\Delta sutA$  pHA-SutA are shown (45 of 62 tRNA genes). (E) qRT-PCR measurements for the 16S leader sequence in the  $\Delta sutA$  and  $P_{ara}:sutA$  strains compared to the wild-type strain. Circles show individual measurements. These data were averaged to generate the expression heatmap shown in Figure 2.5G. (F-I) Normalized ChIP signals at selected genetic loci; scale bar represents 500 bp. Traces are colored as in (C). (J) COG distributions for genes up- and downregulated by SutA, compared to the entire genome. The percentage of genes in each category is indicated with colored bars. Open black bars represent the proportion of the entire genome in each category. Markers indicate categories that are significantly over- ( $\ddagger$ ) or underrepresented (\*) (Fisher's exact test,  $p < 0.001$ ).

## References

- (1) Stover, C. K. et al. *Nature* **2000**, *406*, 959–964.
- (2) Potvin, E.; Sanschagrín, F.; Levesque, R. C. *FEMS Microbiol. Rev.* **2008**, *32*, 38–55.
- (3) Salgado, H.; Gama-Castro, S.; Peralta-Gil, M.; Díaz-Peredo, E.; Sánchez-Solano, F.; Santos-Zavaleta, A.; Martínez-Flores, I.; Jiménez-Jacinto, V.; Bonavides-Martínez, C.; Segura-Salazar, J.; Martínez-Antonio, A.; Collado-Vides, J. *Nucleic Acids Res.* **2006**, *34*, D394–D397.
- (4) Raghavan, R.; Groisman, E. A.; Ochman, H. *Genome Res.* **2011**, *21*, 1487–1497.
- (5) Wurtzel, O.; Yoder-Himes, D. R.; Han, K.; Dandekar, A. A.; Edelheit, S.; Greenberg, E. P.; Sorek, R.; Lory, S. *PLoS Pathog.* **2012**, *8*, e1002945.
- (6) Kopf, S. H.; Sessions, A. L.; Cowley, E. S.; Reyes, C.; Sambeek, L. V.; Hu, Y.; Orphan, V. J.; Kato, R.; Newman, D. K. *Proc. Natl. Acad. Sci. U.S.A.* **2016**, *113*, E110–E116.
- (7) Cowley, E. S.; Kopf, S. H.; LaRiviere, A.; Ziebis, W.; Newman, D. K. *mBio* **2015**, *6*, e00767–15.
- (8) Hoboth, C.; Hoffmann, R.; Eichner, A.; Henke, C.; Schmoltdt, S.; Imhof, A.; Heesemann, J.; Hogardt, M. *J. Infect Dis.* **2009**, *200*, 118–130.
- (9) Quinn, R. A.; Lim, Y. W.; Maughan, H.; Conrad, D.; Rohwer, F.; Whiteson, K. L. *mBio* **2014**, *5*, e00956–13.
- (10) Glasser, N. R.; Kern, S. E.; Newman, D. K. *Mol. Microbiol.* **2014**, *92*, 399–412.
- (11) Eschbach, M.; Schreiber, K.; Trunk, K.; Buer, J.; Jahn, D.; Schobert, M. *J. Bacteriol.* **2004**, *186*, 4596–4604.
- (12) Wauven, C. V.; Piérard, A.; Kley-Raymann, M.; Haas, D. *J. Bacteriol.* **1984**, *160*, 928–934.
- (13) Alvarez-Ortega, C.; Harwood, C. S. *Mol. Microbiol.* **2007**, *65*, 153–165.
- (14) 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. *Environ. Microbiol.* **2010**, *12*, 1719–1733.
- (15) Schreiber, K.; Boes, N.; Eschbach, M.; Jaensch, L.; Wehland, J.; Bjarnsholt, T.; Givskov, M.; Hentzer, M.; Schobert, M. *J. Bacteriol.* **2006**, *188*, 659–668.
- (16) Repoila, F.; Majdalani, N.; Gottesman, S. *Mol. Microbiol.* **2003**, *48*, 855–861.
- (17) Storz, G.; Vogel, J.; Wassarman, K. M. *Mol. Cell* **2011**, *43*, 880–891.

- (18) Dieterich, D. C.; Link, A. J.; Graumann, J.; Tirrell, D. A.; Schuman, E. M. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 9482–9487.
- (19) Dieterich, D. C.; Lee, J. J.; Link, A. J.; Graumann, J.; Tirrell, D. A.; Schuman, E. M. *Nat. Protoc.* **2007**, *2*, 532–540.
- (20) Liberati, N. T.; Urbach, J. M.; Miyata, S.; Lee, D. G.; Drenkard, E.; Wu, G.; Villanueva, J.; Wei, T.; Ausubel, F. M. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 2833–2838.
- (21) Stewart, P. S.; Franklin, M. J. *Nat. Rev. Microbiol.* **2008**, *6*, 199–210.
- (22) Michel, G. P. F.; Aguzzi, A.; Ball, G.; Soscia, C.; Bleves, S.; Voulhoux, R. *Microbiology* **2011**, *157*, 1945–1954.
- (23) Lau, G. W.; Hassett, D. J.; Ran, H.; Kong, F. *Trends Microbiol.* **2004**, *10*, 599–606.
- (24) Wang, Y.; Kern, S. E.; Newman, D. K. *J. Bacteriol.* **2010**, *192*, 365–369.
- (25) Dietrich, L. E. P.; Price-Whelan, A.; Petersen, A.; Whiteley, M.; Newman, D. K. *Mol. Microbiol.* **2006**, *61*, 1308–1321.
- (26) Fröhlich, K. S.; Vogel, J. *Curr. Opin. Microbiol.* **2009**, *12*, 674–682.
- (27) Mooney, R. A.; Davis, S. E.; Peters, J. M.; Rowland, J. L.; Ansari, A. Z.; Landick, R. *Mol. Cell* **2009**, *33*, 97–108.
- (28) Kusuya, Y.; Kurokawa, K.; Ishikawa, S.; Ogasawara, N.; Oshima, T. *J. Bacteriol.* **2011**, *193*, 3090–3099.
- (29) Zhang, Y.; Mooney, R. A.; Grass, J. A.; Sivaramakrishnan, P.; Herman, C.; Landick, R.; Wang, J. D. *Mol. Cell* **2014**, *53*, 766–778.
- (30) Dennis, P. P.; Ehrenberg, M.; Bremer, H. *Microbiol. Mol. Biol. Rev.* **2004**, *68*, 639–668.
- (31) Michiels, J.; Dirix, G.; Vanderleyden, J.; Xi, C. *Trends Microbiol.* **2001**, *9*, 164–168.
- (32) Kang, Y.; Lunin, V. V.; Skarina, T.; Savchenko, A.; Schurr, M. J.; Hoang, T. T. *Mol. Microbiol.* **2009**, *73*, 120–136.
- (33) Tatusov, R. L. et al. *BMC Bioinformatics* **2003**, *4*, 41.
- (34) Paul, B. J.; Barker, M. M.; Ross, W.; Schneider, D. A.; Webb, C.; Foster, J. W.; Gourse, R. L. *Cell* **2004**, *118*, 311–322.
- (35) Perederina, A.; Svetlov, V.; Vassilyeva, M. N.; Tahirov, T. H.; Yokoyama, S.; Artsimovitch, I.; Vassilyev, D. G. *Cell* **2004**, *118*, 297–309.
- (36) Perron, K.; Comte, R.; Van Delden, C. *Mol. Microbiol.* **2005**, *56*, 1087–1102.
- (37) Lemke, J. J.; Sanchez-Vazquez, P.; Burgos, H. L.; Hedberg, G.; Ross, W.; Gourse, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 5712–5717.

- (38) Blaby-Haas, C. E.; Furman, R.; Rodionov, D. A.; Artsimovitch, I.; de Crécy-Lagard, V. *Mol. Microbiol.* **2011**, *79*, 700–715.
- (39) López-Sánchez, A.; Jiménez-Fernández, A.; Calero, P.; Gallego, L. D.; Govantes, F. *Environ. Microbiol. Rep.* **2013**, *5*, 679–685.
- (40) Tehranchi, A. K.; Blankschien, M. D.; Zhang, Y.; Halliday, J. A.; Srivatsan, A.; Peng, J.; Herman, C.; Wang, J. D. *Cell* **2010**, *141*, 595–605.
- (41) Gummesson, B.; Lovmar, M.; Nyström, T. *J. Biol. Chem.* **2013**, *288*, 21055–21064.
- (42) Withers, R.; Doherty, G. P.; Jordan, M.; Yang, X.; Dixon, N. E.; Lewis, P. J. *Mol. Microbiol.* **2014**, *93*, 1130–1143.
- (43) Dietrich, L. E. P.; Okegbe, C.; Price-Whelan, A.; Sakhtah, H.; Hunter, R. C.; Newman, D. K. *J. Bacteriol.* **2013**, *195*, 1371–1380.
- (44) Müsken, M.; Di Fiore, S.; Dötsch, A.; Fischer, R.; Häussler, S. *Microbiology* **2010**, *156*, 431–441.
- (45) Mahdavi, A.; Szychowski, J.; Ngo, J. T.; Sweredoski, M. J.; Graham, R. L. J.; Hess, S.; Schneewind, O.; Mazmanian, S. K.; Tirrell, D. A. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 433–438.
- (46) Boersema, P. J.; Rajmakers, R.; Lemeer, S.; Mohammed, S.; Heck, A. J. R. *Nat. Protoc.* **2009**, *4*, 484–494.
- (47) Gilbert, K. B.; Kim, T. H.; Gupta, R.; Greenberg, E. P.; Schuster, M. *Mol. Microbiol.* **2009**, *73*, 1072–1085.

*Chapter 3***STRUCTURAL INSIGHT INTO SUTA'S MECHANISM OF  
TRANSCRIPTIONAL REGULATION****3.1 Summary of Contributions**

This work is a continuation of the work described in Chapter 2 and my collaboration with Dr. Megan Bergkessel continued accordingly. Contributions that were primarily my own include chemical cross-linking and foot-printing experiments, preparation and analysis of LC-MS/MS experiments, and analytical chromatography. We shared the following work equally: generation of plasmids and mutant strains, experimental planning, purification of SutA and RNA polymerase, data interpretation, and writing of the manuscript. We expect that follow-up experiments described at the end of this chapter will lead to a published manuscript.

### 3.2 Abstract

We recently reported the discovery of SutA, a small, acidic transcription factor expressed during slow growth in *Pseudomonas aeruginosa*. The primary sequence of SutA does not match any characterized protein domains and nothing is known about SutA's mechanism of transcriptional regulation. We performed a series of *in vitro* cross-linking and protein foot-printing experiments to investigate the structural interaction between SutA and the RNAP complex. We find that SutA interacts with the  $\beta$  and  $\beta'$  RNAP subunits in the presence and absence of nucleic acids, and through this interaction, is involved in a conformational change of RNAP. We propose a model for SutA's mechanism of transcriptional regulation that incorporates these new structural data and the previously observed effects of SutA on the transcription of ribosomal RNA and ribosomal protein genes.

### 3.3 Introduction

We previously described the discovery of SutA, a new type of transcriptional regulator in *Pseudomonas aeruginosa* [1]. SutA is a small, acidic protein that binds RNA polymerase (RNAP). The physiological role of SutA is broad; its expression is upregulated in slow growth conditions, and it enhances biofilm formation, regulates the production of phenazines, and contributes to survival during fluctuating conditions. SutA's effects on transcription are also broad; SutA colocalizes with RNA polymerase throughout much of the chromosome, and exhibits particularly strong association with the genes encoding ribosomal components [ribosomal RNA (rRNA) and ribosomal proteins (rProteins)]. Based on chromatin immunoprecipitation sequencing (ChIP-Seq) and RNA-Seq measurements, SutA generally has a positive influence on the transcription of genes with which it associates (e.g., rRNA are two-fold less abundant in the absence of SutA). However, nothing is known about SutA's mechanism of transcriptional regulation.

The transcription of ribosomal components is the primary function of bacterial RNAP. Under growth conditions, rRNA makes up an estimated 85% of cellular RNA, and rProteins 10-20% of cellular protein [2]. Accordingly, the regulation of this transcription in response to changing environmental conditions is critical and complex. rRNA transcription in particular is subject to a variety of regulatory mechanisms, including the binding of protein factors to DNA upstream of promoter regions (e.g., Fis and H-NS), the direction of RNAP to rRNA promoters by  $\sigma$ -factors, regulation of RNAP initiation and elongation steps by small proteins (e.g., DksA) and small molecules [e.g., guanosine tetraphosphate (ppGpp)] involved in the stringent response, and by "antitermination," which prevents premature release of RNAP during elongation of rRNA operons (reviewed in [3, 4]). rProtein transcription is also controlled by the stringent response regulators DksA and ppGpp [5], but the consequent translation of rProtein transcripts adds an additional layer of regulation to those genes, including recent evidence for translation-coupled transcriptional regulation of elongation [6].

The set of small proteins that bind RNAP and modulate its activity is varied. These include proteins that bind the RNAP secondary channel like the well-studied stringent response regulator DksA and its homologs (the *P. aeruginosa* genome encodes four DksA-like proteins), GreA and GreB, and Rnk [7]; the elongation and antitermination Nus factors; and a variety of other factors like the nonessential  $\delta$  subunit from *Bacillus subtilis* [8], the small acidic protein AtfA from *Acinetobacter*

*spp.* [9], and the widely conserved transcription factor CarD [10].

To better understand the mechanism by which SutA regulates transcription and to place SutA in the context of known transcriptional regulators, we sought to establish a structural model of its interaction with RNAP. We performed a set of *in vitro* protein cross-linking and foot-printing experiments to find the domains of the RNAP complex to which SutA binds. We use these and other *in vitro* experiments to provide evidence for the role of SutA as a modulator of RNAP conformation, and suggest a model by which SutA enhances transcription during periods of slow growth. We finish with recommendations for future work.

### 3.4 Results

#### **SutA Interacts with the $\beta$ and $\beta'$ Subunits of RNAP.**

SutA lacks sequence similarity to any characterized proteins or domains, so the molecular mechanism by which it affects transcription is difficult to predict. It is a small (11.2 kDa) protein with a striking number of acidic residues, particularly in the N-terminal third of its sequence. Of its 105 amino acids, 32 are negatively charged and only 13 positively charged, giving a predicted pI of 3.87 and predicted charge of  $-19$  at pH 7. Much of the protein is predicted to be disordered [11] and sequence-based structural analysis with Phyre [12] predicts a short alpha helical domain between residues Ala<sup>59</sup> and Ser<sup>76</sup> (Figure 3.S1A). Circular dichroism (CD) measurements of purified SutA are consistent with these predictions, with strong local minima observed near 200 nm and at 222 nm, corresponding to unstructured and  $\alpha$ -helical conformations, respectively (Figures 3.S1B). To determine the regions of the RNAP complex with which SutA interacts, we undertook two *in vitro* cross-linking approaches: an unbiased cross-linking approach using bis(sulfosuccinimidyl) suberate (BS<sup>3</sup>), and a directed approach that relies on the site-selective incorporation of the UV-activated amino acid cross-linker L-benzoylphenylalanine (Bpa) (Figure 3.1A).

BS<sup>3</sup> is a homobifunctional cross-linker, with two amine-reactive NHS-esters bridged by a hexamethylene linker. BS<sup>3</sup> reacts specifically with terminal amines on lysine side chains or protein N-termini that are in close proximity within a complex. This approach has previously been applied to study the structure of protein complexes, including eukaryotic RNA Pol II [13]. We added BS<sup>3</sup> to a mixture of purified SutA and RNAP in a 10:1 molar ratio. SDS-PAGE analysis showed reduction in intensity of the SutA and RNAP subunit bands and the appearance of high molecular weight complexes upon cross-linker addition (Figure 3.1B). With or without SutA, two

large cross-linked products were observed: one that did not migrate through the gel, and a large, but mobile product that likely corresponds to a single RNAP complex. Gel intensity analysis revealed that the addition of SutA to the cross-linking reaction resulted in preferential formation of the smaller product (70% with SutA vs. 35% without) (Figure 3.S2A).

Cross-linked complexes were digested and analyzed via liquid chromatography-tandem mass spectrometry (LC-MS/MS) and spectra were searched for matches to cross-linked peptides. Following stringent filtering of spectra from two independent experiments (see Materials and Methods, Figure 3.S2B-C), we identified 16 cross-links: nine between RNAP residues, three between SutA residues, and three between SutA and RNAP (Table 3.1). The three cross-links between SutA and RNAP were between Lys<sup>62</sup> or Lys<sup>69</sup> of SutA and Lys<sup>116</sup> of the  $\beta$  subunit and between Lys<sup>95</sup> of SutA and Lys<sup>40</sup> of the  $\beta'$  subunit. Each was identified via high-quality matches of MS/MS spectra to fragment ion masses from each peptide, including fragment ions that contained the cross-linked sites (Figure 3.S3).  $\beta$  Lys<sup>116</sup> and  $\beta'$  Lys<sup>40</sup> are located in the “ $\beta$  lobe 1” and the “ $\beta'$  clamp” domains, respectively, which lie on opposite sides of the clamp that closes around downstream template DNA during transcription elongation [14].

Due to its reaction with amines and its relatively long linker (11.4 Å), cross-linking with BS<sup>3</sup> can be performed with wild-type proteins and allows for the discovery of protein domains that are in close proximity but not necessarily in direct contact. However, the reaction relies on lysine residues which are conspicuously absent from much of SutA's sequence. We therefore took a complementary, targeted cross-linking approach using the cross-linking amino acid Bpa. Bpa can be site-selectively incorporated into a protein by stop-codon reassignment using a mutant aminoacyl-tRNA synthetase/tRNA pair [15]. Upon UV activation, the diphenyl ketone can react nonspecifically with a nearby carbon. Because Bpa is a linker-free cross-linker and due to the short-lived radical intermediate, we expect Bpa cross-links to reflect short-range interactions between SutA and RNAP.

We purified nine SutA mutants with Bpa incorporated at positions throughout the protein sequence (Leu<sup>6</sup>, Leu<sup>11</sup>, Leu<sup>22</sup>, Leu<sup>54</sup>, Gln<sup>61</sup>, Phe<sup>74</sup>, Ile<sup>84</sup>, Val<sup>89</sup>, or Tyr<sup>100</sup>). We mixed each SutA mutant with RNAP under the same conditions used for the BS<sup>3</sup> cross-linking (10:1 molar ratio). We exposed mixtures to 365 nm light, separated resulting species via SDS-PAGE, and looked for RNAP subunits with increased mass corresponding to cross-linking to SutA. We observed cross-linking between SutA

Table 3.1: **BS<sup>3</sup> cross-linked peptides.** Cross-links detected following BS<sup>3</sup> cross-linking of RNAP-SutA complexes. The first four columns indicate the proteins and amino acid positions involved in each cross-link. Evidence columns represent the number of spectra matched to each cross-linked amino acid pair in each replicate. Maximum Score Difference (SD) is reported as calculated by Protein Prospector. Distance refers to the inter  $\alpha$ -carbon distance. Cross-links between SutA and RNAP are bolded.

Protein 1	AA 1	Protein 2	AA 2	Ev. Rep. 1	Ev. Rep. 2	Max. SD	Dist. (Å)
RpoB	116	RpoB	481	NA	8	32.8	18.4
SutA	95	SutA	99	2	NA	30.9	NA
<b>RpoB</b>	<b>116</b>	<b>SutA</b>	<b>69</b>	<b>5</b>	<b>9</b>	<b>26.1</b>	NA
RpoB	265	RpoB	284	5	1	23.9	17.2
SutA	69	SutA	80	31	6	22.8	NA
<b>RpoB</b>	<b>116</b>	<b>SutA</b>	<b>62</b>	<b>NA</b>	<b>18</b>	<b>17.3</b>	NA
RpoB	1144	RpoB	1160	9	20	15.3	13.3
RpoB	284	RpoC	1047	NA	3	14.6	40.1
RpoB	650	RpoC	678	NA	2	14.5	27.6
RpoB	1144	RpoB	1215	5	NA	13.1	33.3
RpoC	50	RpoC	87	4	6	12.8	18.1
<b>RpoC</b>	<b>40</b>	<b>SutA</b>	<b>95</b>	<b>2</b>	<b>1</b>	<b>12.4</b>	NA
RpoB	600	RpoB	631	6	NA	10.9	17.7
RpoB	1154	RpoB	1160	5	20	9.7	8.3
SutA	60	SutA	69	1	4	8.4	NA
SutA	1	SutA	80	4	NA	8.3	NA

Leu<sup>54</sup> and the RNAP  $\beta$  subunit in a UV dose-dependent manner (Figure 3.S4A) and did not observe cross-linking following incubation with wild-type SutA (Figure 3.1C, left lane) or in the absence of UV treatment for any SutA mutant (Figure 3.S4C). We surveyed the panel of SutA mutants and observed cross-linking between positions 54 and 84 to the  $\beta$  subunit with nearly complete yield, while positions 6, 11, and 22 cross-linked to both  $\beta$  and  $\beta'$  subunits with lower yield (Figure 3.1C). No mutants cross-linked to the  $\alpha$  subunit (Figure 3.S4B), and the other SutA mutants tested showed minimal evidence of cross-linking to any RNAP subunit. Because we suspect SutA to interact with RNAP during transcription, we performed the experiment on a pre-formed complex of RNAP and a ssRNA-ssDNA dimer similar to one previously used to constrain RNAP in a transcription elongation complex [16]. Based on gel analysis, cross-linked products were qualitatively the same in the presence (Figure 3.1C) or absence (Figure 3.S4C) of the nucleic acids.

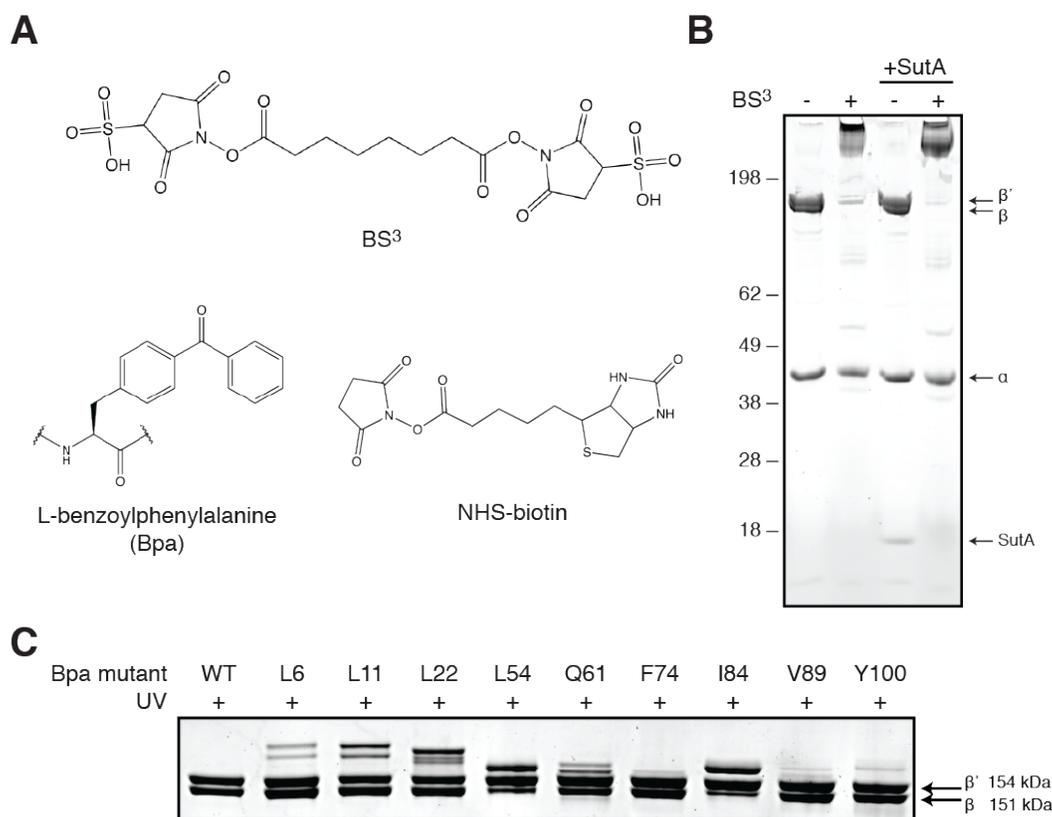
We cut cross-linked products from each lane, digested the proteins, and analyzed

peptides via LC-MS/MS. We searched for masses uniquely identified in each cross-linked sample and not found from a control sample of uncross-linked  $\beta$  and  $\beta'$  subunits. Because the radical-mediated cross-linking of Bpa is nonspecific, searching for particular modifications is challenging, but we were able to identify three cross-linked products (Table 3.2, Figure 3.S5). We detected cross-links between SutA Leu<sup>6</sup>, Leu<sup>11</sup>, and Leu<sup>22</sup> and peptides from the  $\beta'$  subunit clamp domain.

Table 3.2: **Bpa cross-links** Modified peptides detected by MS following cross-linking of SutA Bpa mutants to RNAP.

SutA position	RNAP subunit	Peptide	AA positions
L6	RpoC	KGQLLNDE	156-163
L11	RpoC	QYFEALE	165-170
L22	RpoC	KRM(ox)LQE	296-301

To complement these cross-linking results, we performed a protein foot-printing experiment [17]. Samples containing RNAP with or without SutA were reacted with *N*-hydroxysuccinimidobiotin (NHS-biotin) (Figure 3.1A), which covalently modifies solvent-exposed lysines, leading to a biotin mass modification readily detectable by MS. Proteins were separated via SDS-PAGE and the  $\beta$  and  $\beta'$  subunits were digested and analyzed via LC-MS/MS. More than 80% of the lysines in each protein were identified and approximately half of those were also found in their biotinylated state. To identify regions of RNAP whose solvent accessibility changes in the presence of SutA, we compared the intensity of peptides containing modified lysines between the two experimental conditions (i.e., with or without SutA). We identified eight residues whose modification was reduced at least 1.5-fold in the presence of SutA (Figure 3.S6). We grouped these residues into three categories: residues found near SutA cross-links: Lys<sup>116</sup> and Lys<sup>119</sup> of the  $\beta$  subunit; residues along the main channel: Lys<sup>1257</sup> and Lys<sup>1277</sup> of the  $\beta$  subunit and Lys<sup>332</sup> and Lys<sup>1231</sup> of the  $\beta'$  subunit; and residues on the secondary channel face: Lys<sup>1215</sup> of the  $\beta$  subunit and Lys<sup>996</sup> of the  $\beta'$  subunit. We also found two residues with higher modified intensities in the presence of SutA: Lys<sup>207</sup> of the  $\beta$  subunit along the main channel and Lys<sup>603</sup> of the  $\beta'$  subunit on the secondary channel face. We note that a change in biotinylation of a given residue can indicate either SutA binding at that location or a change in RNAP conformation that alters the accessibility of that residue.



**Figure 3.1: SutA cross-links to RNAP *in vitro*.** (A) Chemicals used for cross-linking and foot-printing experiments. (B) SDS-PAGE analysis of RNAP and SutA before and after cross-linking with BS<sup>3</sup>. The gel is annotated with positions of molecular weight markers (kDa, left) and RNAP subunits (right). (C) Cross-linked RNAP subunits following incubation with SutA Bpa mutants and UV irradiation. The position of the SutA residue replaced by Bpa is indicated above.

### SutA Binds Across the RNAP Clamp.

The structure of *P. aeruginosa* RNAP has not been solved, so to visualize cross-linking and foot-printing results, we mapped cross-linked residues to the SutA sequence (Figure 3.2A) and to homologous residues in the *Escherichia coli* RNAP structure (PDB: 3LU0, Figure 3.2B) [18]. The primary sequences of *E. coli* RNAP subunits are more than 84% similar to their *P. aeruginosa* homologs. Distances between  $\alpha$ -carbons of intra-RNAP BS<sup>3</sup> cross-linked residues ranged from 8-40Å, a range similar to that observed for cross-links between RNA Pol II lysines in a previous BS<sup>3</sup> cross-linking experiment (Table 3.1, Figure 3.S7) [13]. Distances longer than expected based on the length of the cross-linker can be explained by (i) differences between the structures of the *E. coli* and *P. aeruginosa* complexes, (ii) dynamics of the complex in solution, or (iii) a conformational change of RNAP in

the presence of SutA.

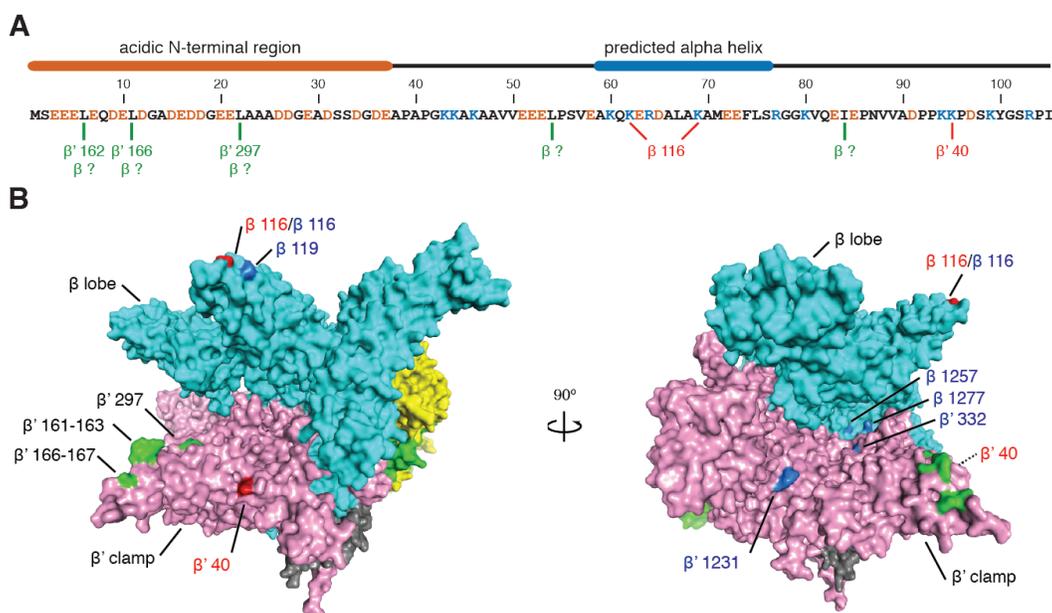
The three BS<sup>3</sup> cross-links between SutA and RNAP (Figure 3.2, red) point to SutA's interaction with both sides of the RNAP clamp ( $\beta$  Lys<sup>116</sup> and  $\beta'$  Lys<sup>40</sup>) that closes around downstream DNA during elongation (forming the “closed clamp” or DNA “open complex”). The two BS<sup>3</sup> cross-links to the  $\beta$  lobe 1 coincide with two obscured residues detected in the foot-printing experiment:  $\beta$  Lys<sup>116</sup> and  $\beta$  Lys<sup>116</sup> (Figure 3.2B, blue). Taken together, these provide strong evidence for a direct interaction between the central region of SutA and the  $\beta$  lobe 1 domain. Near complete cross-linking of SutA Bpa mutants Leu<sup>54</sup> and Ile<sup>84</sup> to the  $\beta$  subunit further corroborate these results.

Additional obscured residues on the  $\beta$  and  $\beta'$  subunits lie along the RNAP main channel that accommodates downstream template DNA. If the C-terminal half of SutA bridges the clamp, the acidic, and likely unstructured N-terminus could be in position to obscure these residues. However, Bpa cross-links between positions in the N-terminal portion of SutA (Leu<sup>6</sup>, Leu<sup>11</sup>, and Leu<sup>22</sup>) place these residues against the  $\beta'$  clamp, just outside of the main channel (Figure 3.2, green). Additionally, our previous observations that SutA associates with the chromosome across the gene coding regions [1] suggests that its interaction with RNAP can occur during elongation while DNA is present in the main channel. While these observations do not rule out the occupation of the main channel by the SutA N-terminus, they suggest the alternative hypothesis that main-channel lysines are instead obscured by a conformational change of RNAP that reduces their solvent accessibility.

### **SutA is Involved with a Conformational Change of RNAP.**

To explore the possibility that SutA induces a conformational change of RNAP, we explored the intra-RNAP cross-links captured by our BS<sup>3</sup> experiment. A majority of the detected cross-links spanned residues separated by a distance equal to or less than expected by the length of the cross-linker and the lysine side chains (approx. 25 Å) (Table 3.1). Of the three cross-links that spanned larger distances,  $\beta$  Lys<sup>284</sup>- $\beta'$  Lys<sup>1047</sup> (40 Å) and  $\beta$  Lys<sup>650</sup>- $\beta'$  Lys<sup>678</sup> (28 Å), span the RNAP clamp (Figure 3.S7).

Following our initial purification of SutA and RNAP, to confirm that their interaction was preserved in the *in vitro* conditions, we evaluated binding using analytical size exclusion chromatography (SEC) of RNAP alone, SutA alone, or RNAP and SutA mixed together (Figure 3.3A). SDS-PAGE analysis of RNAP elution fractions from the mixed sample provided evidence for SutA binding under these conditions (Figure



**Figure 3.2: SutA interacts with both sides of the RNAP clamp.** (A) SutA sequence, predicted structural domains, and residues found to cross-link to RNAP. Acidic residues are orange and basic residues are blue. Domains predicted by Phyre are indicated above. Residues found to cross-link to RNAP via BS<sup>3</sup> (red) or Bpa (green) are indicated below. Particular positions of cross-linking are indicated by RNAP subunit and residue position. Bpa cross-links for which interacting residues were not identified are shown with question marks. (B) Views of the RNAP core enzyme showing mapped locations of lysines involved in BS<sup>3</sup> cross-links to SutA (red), regions involved in Bpa cross-links (green), and lysines with reduced modification by NHS-biotin in the presence of SutA (blue). Subunits are colored as follows:  $\alpha_I$  (yellow),  $\alpha_{II}$  (green),  $\beta$  (cyan),  $\beta'$  (pink), and  $\omega$  (gray). The structure was adapted from PDB:3LU0 [18].

3.3B); SutA was present in early eluent fractions only when RNAP was included. Interestingly, we noticed that RNAP in the presence of SutA exhibited later elution compared to the core enzyme alone (1.33 mL vs. 1.28 mL), suggestive of a possible conformational change of RNAP upon SutA binding. This change is consistent with a decrease in the Stokes radius (i.e., a more compact complex conformation). Similarly, we observed a difference in the preferred cross-linked product following BS<sup>3</sup> cross-linking of RNAP with and without SutA. We cannot unambiguously assign these gel bands to molecular species, but note that the inclusion of SutA led to a majority of the smaller product.

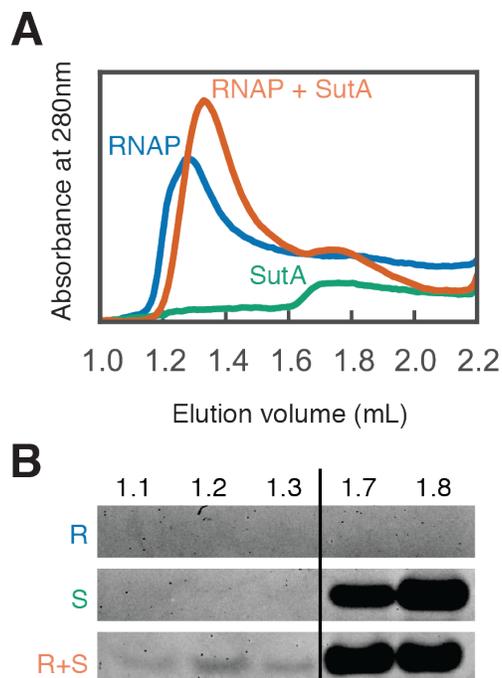


Figure 3.3: **SEC analysis of the SutA-RNAP complex.** (A) UV absorbance traces of size exclusion chromatography of RNAP (blue), SutA (green), and the mixed complex (orange). (B) Fractions (0.1 mL total volume, with start volumes listed above each lane) from each chromatographic separation in (A) were concentrated, separated via SDS-PAGE, and stained with Coomassie. Images show the SutA region of each gel.

### The C-terminal, RNAP-binding Domain is Required for SutA Function.

To test whether the sites of interaction identified in the cross-linking experiments are relevant *in vivo*, we expressed SutA mutants composed of either the N-terminal or the C-terminal portion of SutA in a SutA deletion strain ( $\Delta sutA$ ) and evaluated each mutant's ability to phenocopy a strong phenotype of the deletion strain: overproduction of the small, colored phenazine, pyocyanin (PYO). We grew each strain in minimal medium and measured absorbance of the culture supernatant at 312 nm as a proxy for PYO abundance (Figure 3.4). The C-terminal fragment complemented the PYO overproduction phenotype of the  $\Delta sutA$  strain while the acidic N-terminal fragment did not, suggesting that SutA's C-terminal portion, which contains all residues captured by BS<sup>3</sup> and the Bpa cross-links to the  $\beta$  subunit, is more important for SutA's function than the N-terminal portion.

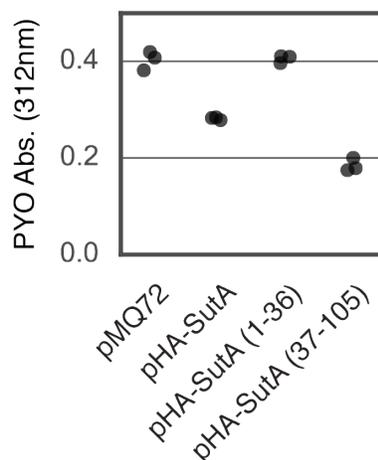


Figure 3.4: **Effects of SutA truncation on pyocyanin production.** The *sutA* deletion mutant was transformed with the indicated plasmids. Absorbance of each culture supernatant serves as a proxy for pyocyanin concentration.

### 3.5 Discussion

Here we provide structural context for the interaction between RNAP and SutA, a small, acidic, previously uncharacterized transcription factor. Previous work exploring SutA's physiological effects showed strong association of SutA with the loci encoding ribosomal genes, where it enhances their transcription. We also previously noted striking contrasts with the function of the small protein DksA that binds in the RNAP secondary channel. As part of the stringent response, DksA is involved in downregulating transcription of ribosomal components during nutrient downshifts. Mutation of *dksA* in *Pseudomonas spp.* has opposite effects on biofilm formation and pyocyanin production than we have observed in the *sutA* deletion strain.

DksA has complex allosteric effects on the activities of RNAP, so anticipating how SutA might cause opposing effects is difficult [19]. However, size exclusion chromatography indicates that SutA may affect RNAP conformation, and cross-linking and foot-printing results allow us to speculate about what these conformational changes might entail. The coincidence of the cross-linked (Lys<sup>116</sup>) and obscured residues (Lys<sup>116</sup> and Lys<sup>119</sup>) on the  $\beta$  lobe 1 domain suggest that this is a site of direct interaction between the predicted alpha helical portion of SutA and RNAP. The cross-link between the C-terminal region of SutA and Lys<sup>40</sup> in the  $\beta'$  clamp raises the possibility that SutA might bridge the cleft between the  $\beta$  lobe 1 and  $\beta'$  clamp domains. Though the distance between the cross-linked residues in these domains

is approximately 70 Å in the open clamp conformations of RNAP observed in most crystal structures [18, 20], the mobility of the  $\beta'$  clamp is well documented [21] and the main channel contracts upon DNA binding and transition to the transcription elongation complex [22]. Without atomic-level structural information about SutA and *P. aeruginosa* RNAP, we cannot determine whether SutA requires (or forces) a closed-clamp conformation to bridge this gap. However, our observation that lysine residues within the RNAP main channel are less accessible to chemical modification in the presence of SutA is consistent with clamp closure. Identification of Bpa cross-links between residues within the N-terminal domain of SutA to the  $\beta'$  clamp near, but outside the main channel suggest that the N-terminus itself does not obscure those residues. We interpret the partial cross-linking of Leu<sup>6</sup>, Leu<sup>11</sup>, Leu<sup>22</sup> positions to both  $\beta$  and  $\beta'$  subunits to reflect the mobility of the presumably unstructured N-terminal acidic domain.

In addition to rotation of the clamp domain, other movements of large, mobile RNAP domains in the *E. coli* enzyme, such as the  $\beta$ i9 and  $\beta'$ i6 domains, have been observed or inferred [18, 23]. Several of the lysines for which the accessibility to chemical modification changes in the presence of SutA ( $\beta'$  Lys996 and Lys1231, which exhibit decreased modification in the presence of SutA; and  $\beta$  Lys207 and  $\beta'$  Lys603, which exhibit increased modification) are near or part of the  $\beta'$ i6 domain, suggesting that this domain could occupy a different position in the presence of SutA. This domain does not appear in any crystal structures, likely due to its mobility, so alternate conformations are not well characterized. However, it has been suggested that the position of the  $\beta'$ i6 domain can affect the ability of DksA to act on RNAP [23]. Likewise, an open clamp conformation is thought to facilitate DksA's interaction with polymerase [19, 24]. If SutA interacts preferentially with RNAP in a closed clamp conformation, this could contribute to the opposite effects on gene expression and phenotype observed between the *dksA* and *sutA* mutants. Future efforts to obtain additional SutA structural information and to characterize its effects on RNAP *in vitro* will help to distinguish among these possibilities.

A model of SutA as a protein that can alter the clamp position of RNAP is reminiscent of what has been observed for the elongation factor NusG [25], though their exact points of interaction are likely different. There are several other clear differences between the roles and activities of NusG and the potential roles for SutA. NusG is essential, and is critical for facilitating interactions between RNAP and other complexes [26], while SutA is nonessential, even under the conditions in

which it is upregulated, and appears to interact only with the polymerase. NusG is thought to be broadly recruited to most or all genes in *E. coli*, and accumulates as transcription along the gene proceeds [27], while SutA appears to be preferentially recruited to some genes more than others, and its peak of ChIP signal appears just slightly downstream of the peak of RNAP ChIP signal associated with the promoter. Nevertheless, our data suggest that SutA enhances expression of the genes to which it is recruited and one possibility is that like NusG, it does so primarily by enhancing transcription elongation at protein-coding genes (e.g., rProtein genes). At rRNA genes, in contrast, SutA associates primarily to the promoter region, suggesting that its activity may be different for these genes, and as with DksA, may have a primary role in regulating initiation. A key question for future studies will be whether and how SutA may interact or compete with a variety of other RNAP-binding regulators, including the rRNA antitermination complex or  $\sigma$ -factors, for example. More detailed structural resolution and measurements of the direct effects of SutA on RNAP activity will also be required to adequately test the proposed model.

### 3.6 Future Work

An atomic-level structure of SutA alone or in complex with RNAP would greatly enhance our understanding of its mechanism of transcriptional regulation. Thus far, attempts to obtain a co-crystal structure of SutA and RNAP have been unsuccessful. However, preliminary NMR measurements of SutA alone have been promising. Heteronuclear single quantum coherence (HSQC) measurements of  $^{15}\text{N}$ -labeled SutA show well-dispersed  $^1\text{H} - ^{15}\text{N}$  peaks (Figure 3.S8). Three-dimensional NMR measurements with  $^{15}\text{N}$ ,  $^{13}\text{C}$ -labeled SutA should be useful in determining its solution structure. Additionally, experiments on labeled SutA mixed with RNAP may provide information about any SutA rearrangements that occur upon binding.

Our model of SutA as a modulator of the RNAP clamp brings with it predictions about specific effects on transcription. For example, at rRNA promoters, the RNAP-DNA open complex (clamp closed) is notoriously unstable; under rich nutrient conditions, this instability allows RNAP to move quickly off of promoter regions and proceed to elongation, while in nutrient-limited conditions, factors like DksA further destabilize the RNAP-DNA open complex at rRNA promoters and reduce transcription. The effect of SutA on open complex stability at the promoters of rRNA and other genes can be tested explicitly with *in vitro* transcription experiments. Furthermore, results from our ChIP experiments provide a list of other genomic loci to which SutA associates strongly (e.g., rProtein and other protein-coding genes).

Effects of SutA on RNAP behaviors such as elongation rate or premature termination can also be examined at these loci.

### 3.7 Experimental Procedures.

**Strain construction.** See Table 3.3 for a complete list of strains used. Cloning was performed using standard methods. Enzymes and supplies were purchased from New England Biolabs. For overexpression and purification, SutA was cloned from *P. aeruginosa* genomic DNA, appended with an N-terminal 6x Histidine tag followed by a TEV cleavage site (MRGSHHHHHHENLYFQS), and cloned into pQE80L (Qiagen) to generate DKN1697. Plasmids for overexpression of mutants for Bpa incorporation were generated from DKN1697 by replacing the codon at each indicated position with an amber stop codon (TAG) via PCR amplification and blunt-end ligation to create BMB14X, where X indicates the position in the amino acid sequence that encodes for Bpa. Plasmids for overexpression of SutA truncation mutants were generated from DKN1640 [1] by removing the coding region for the N-terminal (positions 1-36; DKN1688) or C-terminal (positions 37-105; DKN1687) amino acids via PCR amplification and blunt-end ligation.

**Media and growth conditions.** All cultures were grown at 37 °C with shaking unless otherwise noted. Liquid media were LB (5 g yeast extract, 10 g tryptone, 10 g NaCl per liter), 2xYT (10 g yeast extract, 16 g tryptone, and 5 g NaCl per liter), or M9 (12.8 g Na<sub>2</sub>HPO<sub>4</sub> · 7 H<sub>2</sub>O, 3 g KH<sub>2</sub>PO<sub>4</sub>, 0.5 g NaCl, 1 g NH<sub>4</sub>Cl, 1 mM MgSO<sub>4</sub>, 0.1 mM CaCl<sub>2</sub> per liter).

**SutA purification.** DKN1697 cells were grown in the presence of 200 µg ampicillin. A 20 mL culture grown overnight in LB was distributed between two flasks each containing one liter of 2xYT and grown to  $OD_{600} = 0.6$ . Protein expression was induced by addition of 1 mM isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG) and expression was allowed to continue for 4 h.

For Bpa incorporation, BMB14X cells were co-transformed with pEVOL-pBpF and approx. 20 colonies were scraped from the agar plate and grown at 33 °C in LB to a  $OD_{600} = 0.6$ . Cultures were treated with 1 mM Bpa (Iris-Biotech) and 1 mM IPTG and incubated in the dark for 20 h.

For <sup>15</sup>N labeled SutA, DKN1697 cells were grown for 8 h in 5 ml LB then added to 50 ml M9 medium containing 1 g/L <sup>15</sup>NH<sub>4</sub>Cl. These starter cultures were grown overnight, then added to 4x 1 L of the same M9 medium. Cells were grown to  $OD_{600} = 0.8$  and protein expression was induced with 1 mM IPTG and allowed to

continue for 6 h.

For all, cells were pelleted and frozen at  $-80\text{ }^{\circ}\text{C}$ . Pellets were resuspended in lysis buffer (40 mM  $\text{NaH}_2\text{PO}_4$ , 300 mM NaCl, pH 8) containing 5 mM imidazole, 1 mg/mL lysozyme, and cOmplete mini protease inhibitor, EDTA free and lysed by probe sonication. The lysate was treated with Benzonase Nuclease on ice for 30 min and centrifuged. Soluble protein was applied to His-Pur Ni-NTA (Thermo Scientific) washed three times with lysis buffer containing 20 mM imidazole and eluted three times with lysis buffer containing 250 mM imidazole. Eluents were combined, loaded onto an Amicon 10 kDa centrifugal filter (EMD Millipore), and buffer exchanged to TEV-digestion buffer (50 mM Tris pH 8.0, 0.5 mM EDTA, and 1 mM DTT). The 6xHis-tag was cleaved by addition of His-tagged TEV protease in a 1:50 mass ratio and incubation overnight at  $4\text{ }^{\circ}\text{C}$ . The digested sample was reapplied to His-Pur Ni-NTA, and washed with lysis buffer containing 20 mM imidazole; SutA eluted in this wash step, while the cleaved peptide tag and His-tagged TEV protease remained bound to the resin. The cleaved protein product includes the native SutA sequence with an additional N-terminal serine. SutA fractions were pooled and concentrated on an Amicon 10 kDa centrifugal filter, applied to a Superdex 75 10/300 column (GE Healthcare), buffer exchanged to SutA storage buffer (25 mM Tris pH 8, 100 mM NaCl, 20% glycerol, and 2 mM  $\beta$ -mercaptoethanol), and stored at  $-80\text{ }^{\circ}\text{C}$ .

Due to the lower yield and additional contaminant proteins following expression in M9 medium, an additional ion exchange step was used to further purify SutA for NMR studies. Following TEV cleavage and removal of the tag, SutA was buffer exchanged into IEX buffer (20 mM *N*-methylpiperazine) with 100 mM NaCl and loaded onto a 5 mL Q FF column (GE Healthcare). Contaminants were washed with IEX buffer with 100 mM NaCl, and SutA was eluted via a linear gradient of IEX buffer from 100 to 600 mM NaCl. Purified protein was buffer exchanged into 10 mM Tris pH 7.0 with 100 mM NaCl via SEC as described above.  $^{15}\text{N}$  incorporation was greater than 97% as verified by whole-protein MS.

***P. aeruginosa* RNAP purification.** RNAP was purified from the *P. aeruginosa*  $\Delta sut A$  strain essentially as previously described ([28] and references therein). Cells were grown in 6 L of Terrific Broth medium to an  $OD_{600}$  of approximately 1.0. Cells were washed with TBS and pellets were frozen at  $-80\text{ }^{\circ}\text{C}$ . Cell pellets were resuspended in 90 mL RNAP lysis buffer (50 mM Tris pH 8.0, 100 mM NaCl, 1 mM EDTA, and cOmplete Ultra EDTA-free protease inhibitor tablets (Roche)) contain-

ing 40 Kunitz units DNaseI and cells were lysed by passage through an EmulsiFlex-C3 (Avestin). Lysates were clarified by centrifugation at 12,000g, and nucleic acids and acidic proteins were precipitated by addition of a 10% polyethyleneimine (polymin P; Sigma-Aldrich) solution at pH 7.9 to a final concentration of 0.5%. Precipitated protein was pelleted, washed with TGEB (10 mM Tris pH 8.0, 5% glycerol, 0.1 mM EDTA, 10 mM  $\beta$ -mercaptoethanol) plus 0.3 M NaCl, and the RNAP fraction was eluted with TGEB plus 1 M NaCl. Residual polymin P was removed by ammonium sulfate precipitation (2M). The ammonium sulfate pellet was resuspended in TGEB and loaded onto a 50 mL Heparin Sepharose 6 Fast Flow column (GE Healthcare). The column was washed with 2 column volumes of TGEB plus 0.3 M NaCl, and RNAP was eluted with a step to TGEB plus 0.6 M NaCl. The elution fraction was precipitated with 2 M ammonium sulfate, and resuspended into approximately 1 mL of TGEB plus 0.5 M NaCl. Low molecular weight contaminants were removed via size exclusion chromatography on a HiPrep 16/60 Sephacryl S-300 HR column (GE Healthcare). Fractions containing RNAP were diluted in TGEB to a final NaCl concentration of 0.3 M and loaded onto a HiTrap Q FF 5 mL column (GE Healthcare). RNAP was eluted into TGEB with a gradient between 0.3 M and 0.5 M NaCl over 20 column volumes. RNAP was dialyzed into RNAP storage buffer (20 mM Tris-HCl pH 8.0, 0.1 mM EDTA, 10 mM  $\beta$ -mercaptoethanol, 100 mM NaCl, 20% glycerol), concentrated to 1.4 mg/mL and frozen at  $-80^{\circ}\text{C}$ . The total yield was approximately 2.9 mg of high purity core enzyme.

**BS<sup>3</sup> cross-linking.** Bis(sulfosuccinimidyl) suberate (BS<sup>3</sup>) d<sub>0</sub> and d<sub>4</sub> isotopologs were purchased from Thermo Scientific. RNAP and SutA were mixed in a 1:10 molar ratio (0.5  $\mu\text{M}$  RNAP, SutA 5.0  $\mu\text{M}$ ) in 10 mM HEPES pH 8, 100 mM potassium acetate and incubated on ice for 1.5 h. Cross-linking was initiated by addition of 5 mM of a 4:1 molar ratio of BS<sup>3</sup> d<sub>0</sub>:d<sub>4</sub> and the reaction was incubated on ice for 2 h. Cross-linking was quenched by addition of ammonium bicarbonate to a final concentration of 50 mM. In-solution digestion, HPLC desalting, and LC-MS/MS were performed as described below. The experiment was performed with two replicates.

**Bpa cross-linking.** RNAP and SutA mutants were mixed in a 1:10 molar ratio (0.5  $\mu\text{M}$  RNAP, SutA 5.0  $\mu\text{M}$ ) in 25 mM Tris, pH 8, 100 mM NaCl and incubated on ice for 1 h. Samples were irradiated with 500 mW/cm<sup>2</sup> of 365 nm light for 120 s (or as otherwise indicated). Samples were separated via SDS-PAGE, stained with

Coomassie, and imaged.

**Protein foot-printing.** RNAP (control sample) or RNAP and SutA mixed in a 1:10 molar ratio (0.5  $\mu$ M RNAP, SutA 5.0  $\mu$ M) (SutA sample) were incubated in 10 mM HEPES pH 8, 100 mM potassium acetate on ice for 1.5 h. Modification of lysines was initiated by addition of 100  $\mu$ M NHS-biotin, and samples were incubated at room temperature for 30 min. The reaction was quenched by addition of ammonium bicarbonate to a final concentration of 50 mM. Samples were separated via SDS-PAGE and GeLCMS was performed on the bands corresponding to the  $\beta$  and  $\beta'$  subunits of RNAP. HPLC desalting, and LC-MS/MS were performed as described below. The experiment was performed with two replicates.

**Mass spectrometry.** Liquid chromatography-mass spectrometry were essentially carried out as previously described [29]. Protein foot-printing experiments were performed on a nanoflow LC system, EASY-nLC 1000 coupled to a hybrid linear ion trap Orbitrap Classic mass spectrometer (Thermo Scientific) equipped with a nanoelectrospray ion source (Thermo Scientific) with the following modifications: For the EASY-nLC II system, solvent A consisted of 97.8% H<sub>2</sub>O, 2% ACN, and 0.2% formic acid and solvent B consisted of 19.8% H<sub>2</sub>O, 80% ACN, and 0.2% formic acid. For the LC-MS/MS experiments, digested peptides were directly loaded at a flow rate of 500 nL/min onto a 16-cm analytical HPLC column (75  $\mu$ m ID) packed in-house with ReproSil-Pur C<sub>18</sub>AQ 3  $\mu$ m resin (120 Å pore size, Dr. Maisch, Ammerbuch, Germany). The column was enclosed in a column heater operating at 30 °C. After 30 min of loading time, the peptides were separated with a 50 min gradient at a flow rate of 350 nL/min. The gradient was as follows: 0–30% B (50 min), and 100% B (10 min). The Orbitrap was operated in data-dependent acquisition mode to automatically alternate between a full scan ( $m/z=400$ –1600) in the Orbitrap and subsequent 10 CID MS/MS scans in the linear ion trap. CID was performed with helium as collision gas at a normalized collision energy of 35% and 30 ms of activation time.

BS<sup>3</sup> and Bpa cross-linking experiments were run on the Orbitrap Elite, equipped with a nanoUPLC. Solvent A and B, and column were the same as described above. The gradient was as follows: 2% B for five min, 2-40% B (60 min), and 100% B (10 min). The Orbitrap was operated in data-dependent acquisition mode to automatically alternate between a full scan ( $m/z=300$ -1600) in the Orbitrap and subsequent 5 HCD MS/MS scans in the Orbitrap. Normalized collision energy was 40% and activation time was 100 ms. Resolution on MS was set to 120,000 and

MS/MS was 15,000.

**Analysis of BS<sup>3</sup> cross-links.** Raw files were converted to peak lists with ProteoWizard [27] and analyzed with Protein Prospector online, version 5.12.4 following reported protocols with modifications below [30]. The protein database contained the sequences for SutA, RpoA, RpoB, RpoC, and RpoZ. 80 peaks from each spectrum were searched using a tolerance of 10 ppm for precursor ions and 25 ppm for product ions. Enzyme specificity was GluC, and up to two missed cleavages per peptide were allowed. Carbamidomethylation of cysteines was specified as a constant modification, and protein N-terminal acetylation, oxidation of methionine, and dead-end modification with the cross-linker at lysine positions and protein N-termini were set as variable modifications. Additionally, incorrect monoisotopic peak assignments were considered as variable modifications. The analysis was run twice for each set of peak lists to search for both cross-linker isotopologs. Raw files were independently searched using MaxQuant for precursor mass pairs, differing by 4.02 Da, that represent cross-links made by both linker isotopologs.

Cross-links detected by Protein Prospector were matched against the mass pair list to remove cross-links not present as 4.02 Da offset mass pairs. For cross-links detected between RNAP proteins, we used a reported structural model of the *E. coli* RNAP complex (PDB: 3LU0) to calculate the inter  $\alpha$ -carbon distance between amino acids [18]. We used this calculated distance as a metric to distinguish “quality” cross-links from all others. Based on the length of the linker, the maximum inter  $\alpha$ -carbon distance between lysines cross-linked by BS<sup>3</sup> is 24.6 Å, so we considered cross-links with distances near or below this value to be reasonable. Like the study by Trnka et al., we found Score Difference to be the best discriminant for making this distinction. A Score Difference cutoff of 5.6 (similar to the value of 8.5 found by Trnka et al.) separated high-distance and low-distance cross-links (Figure 3.S2C) yielding an FDR of < 0.05 (see Figure 3.S2B for the ROC curve for this classification model). The final criteria for assigning quality cross-links were: (i) found as a precursor mass pair, (ii) Score Difference greater than 5.6, and (iii) matched by at least two spectra. These cross-links were aggregated to determine the number of spectra from each replicate and the maximum Score Difference for each amino acid linkage (Table 3.1). The best spectra used to identify each cross-link between SutA and RNAP are shown in Figure 3.S3.

**Analysis of Bpa cross-links.** Raw files were searched using MaxQuant against a protein database containing the sequences for SutA, RpoB, and RpoC and a

contaminant database (246 sequences). GluC was specified as the digestion enzyme with up to two missed cleavages. Carbamidomethylation of cysteine was set as a fixed modification and protein N-terminal acetylation and methionine oxidation were variable modifications. We used the “matchedFeatures” output file to search for unique precursor masses in each run (e.g., found in a particular Bpa mutant run, but not in any other, including the uncross-linked RpoB and RpoC sample). We also searched raw files for variable mass modifications at any position corresponding to the expected SutA cross-linked peptide (Table S2) using MS-GF+ [31]. We cross-referenced the list of modified peptides (MSGF Score > 0) with the list of unique masses to find cross-linked peptides from RNAP.

**Analysis of foot-printing.** Raw files were searched using MaxQuant against a protein database containing the sequences for purified RpoB and RpoC and a contaminant database (246 sequences). Trypsin was specified as the digestion enzyme with up to two missed cleavages. Carbamidomethylation of cysteine was set as a fixed modification and protein N-terminal acetylation and methionine oxidation were variable modifications. We also included a variable modification of lysine to search for biotinylated residues. For quantification, the raw files and the list of identified peptides were imported into Skyline version 3.1, and subset for high quality peak matches among all runs (isotopic dot product score > 0.75) [32]. For each replicate, peptide intensity ratios were calculated for each peptide ion between the SutA and control samples. To account for variations in LC-MS/MS loading, all peptide intensity ratios for each experiment were normalized so that the median ratio was 1. Peptides whose intensities changed by 1.5-fold between the SutA and control samples in both replicates were classified as “obscured” (SutA < control) or “revealed” (SutA > control).

**Analytical size exclusion chromatography.** RNAP (0.5  $\mu$ M) and SutA (5.0  $\mu$ M) were incubated together or separately for 2 h at 4 °C in 40 mM Tris pH 8.0, 30 mM NaCl, 5% glycerol, and 3 mM  $\beta$ -mercaptoethanol, then separated on a Superdex Increase 3.2/300 column (GE Healthcare).

**Software analysis and data presentation.** This section describes software packages that were not mentioned above. Data processing and statistical analysis were performed with Python version 2.7.9 with NumPy version 1.9.2, SciPy version 0.15.1, and Pandas version 0.16.1. Data were plotted with Matplotlib version 1.4.3 [33] and Seaborn version 0.5.1. Gel images were analyzed with ImageJ 64-bit version 1.45 [34]. RNAP structural analysis was performed with Biopython version

1.65 [35] and structural visualization was performed with Open-Source PyMOL version 1.3. Figures were assembled in Adobe Illustrator CS5.

### 3.8 Supplementary Figures

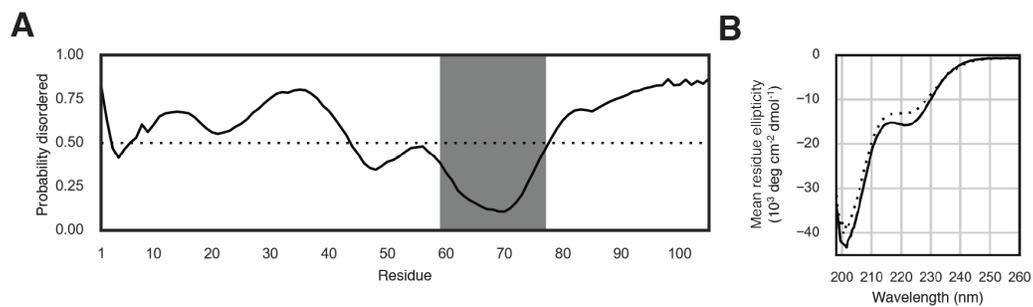


Figure 3.S1: **SutA structural predictions.** (A) Disordered regions were predicted using DisEMBL [11]. The region containing the predicted alpha helix is indicated with a gray box. (B) CD measurements of purified SutA at 4 °C (solid) and 22 °C (dotted).

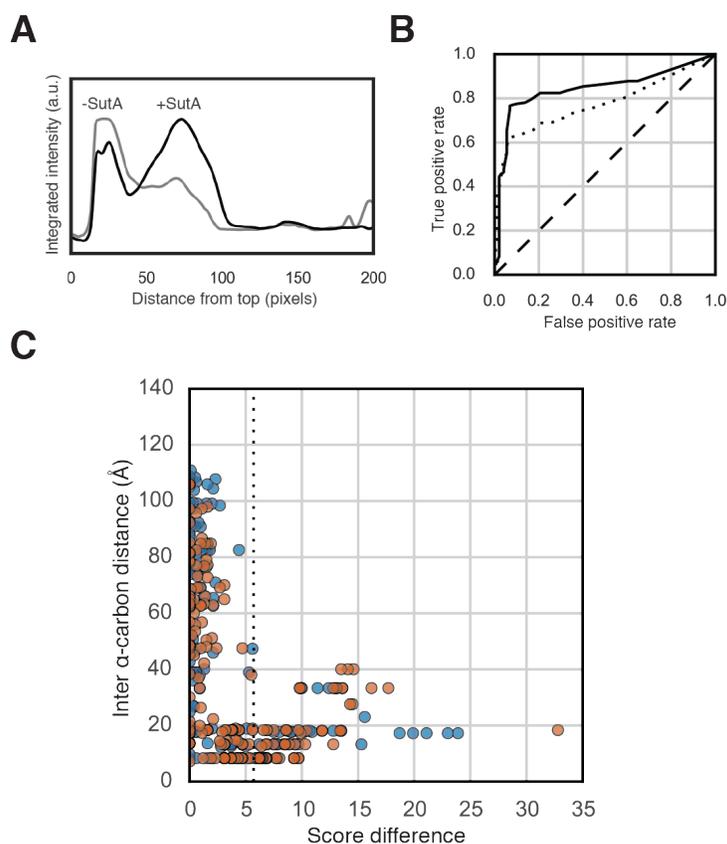


Figure 3.S2: **BS<sup>3</sup> cross-linking.** (A) Analysis of high molecular weight RNAP cross-linking products without (gray) and with SutA (black). Integrated density is plotted for the upper region of lanes from Figure 3.1B. (B) ROC curve for the “quality” cross-link model before (dotted) or after (solid) subsetting for paired precursor ions (i.e., found with both BS<sup>3</sup> isotopologs). (C) Inter  $\alpha$ -carbon distance vs. Score Difference for cross-links from replicate 1 (blue) and replicate 2 (orange). The dotted line indicates the Score Difference cutoff used.

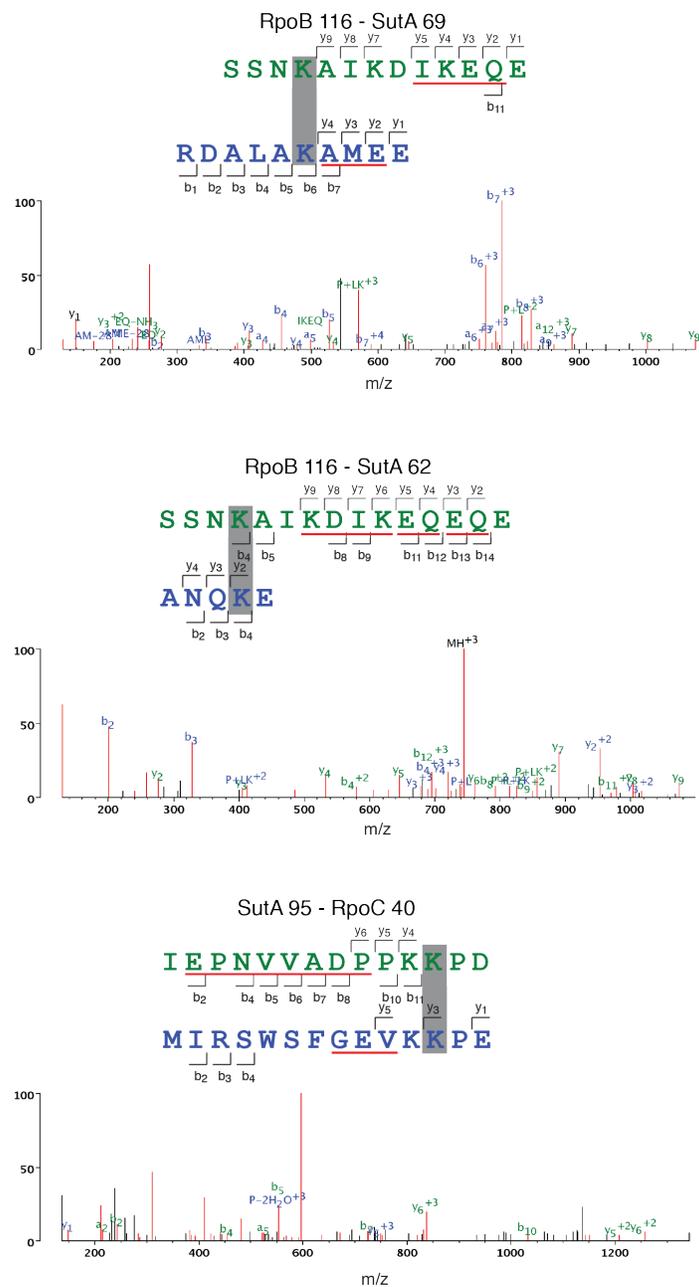


Figure 3.S3: **MS2 spectra for SutA-RNAP BS<sup>3</sup> cross-links.** HCD spectra for the highest scoring cross-links between SutA and RNAP. Matched fragment ions (b and y) are indicated above and below the peptide sequences and inter-peptide fragment ions by red underlines. The location of each cross-link is indicated by a gray box.

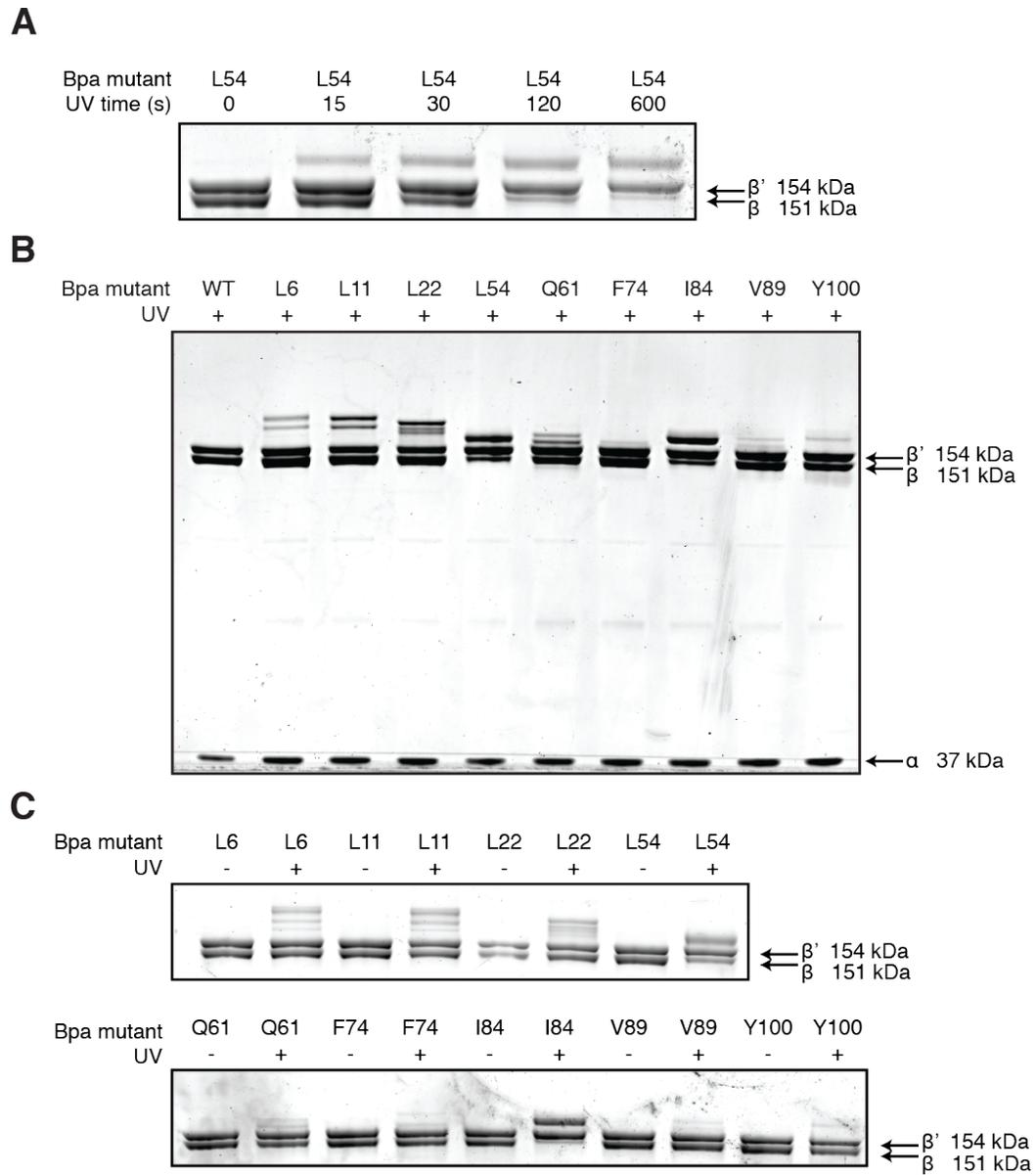


Figure 3.S4: **Bpa cross-linking.** SDS-PAGE analysis of Bpa cross-linking. (A) A time course of UV irradiation of RNAP mixed with SutA L54Bpa mutant. (B) Full gel from Figure 3.1C. (C) RNAP incubated with SutA mutants and irradiated in the absence of nucleic acids.

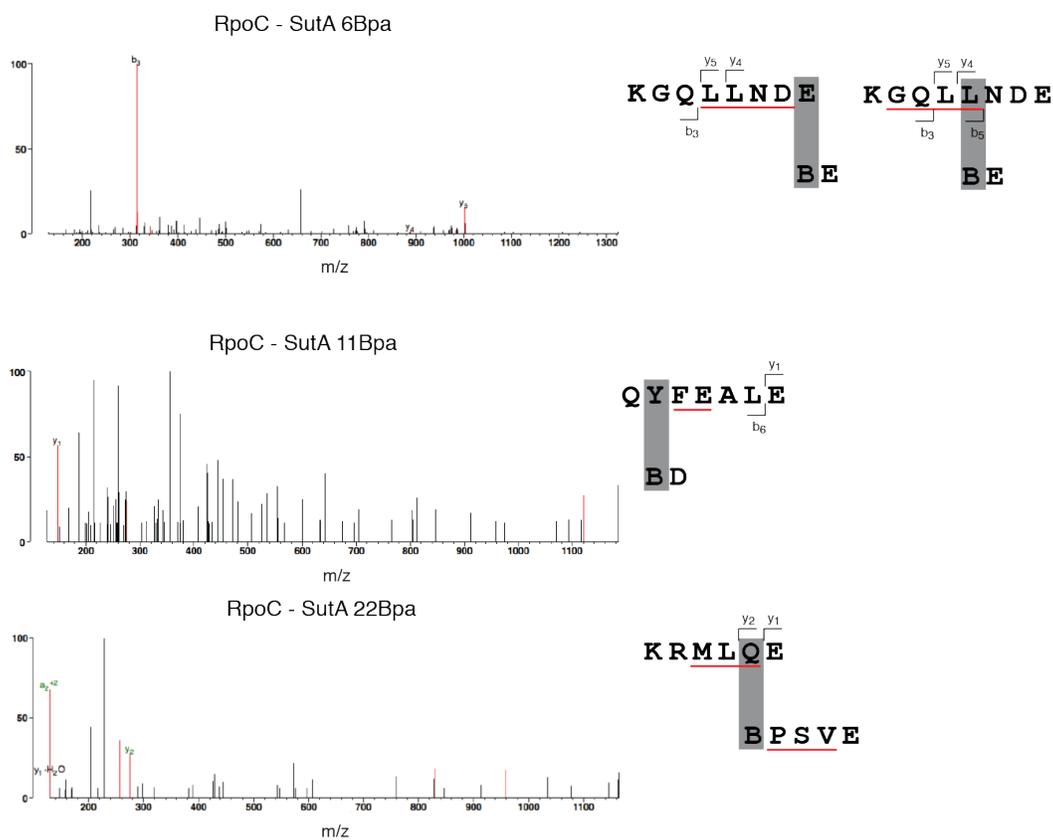


Figure 3.S5: **Bpa cross-link spectra.** HCD spectra for the best cross-links between SutA Bpa mutants and RNAP. Matched fragment ions (b and y) are indicated above and below the peptide sequences and inter-peptide fragment ions by red underlines. The location of each cross-link is indicated by a gray box. When evidence for multiple cross-links between the same peptides were found, all cross-link locations are shown.

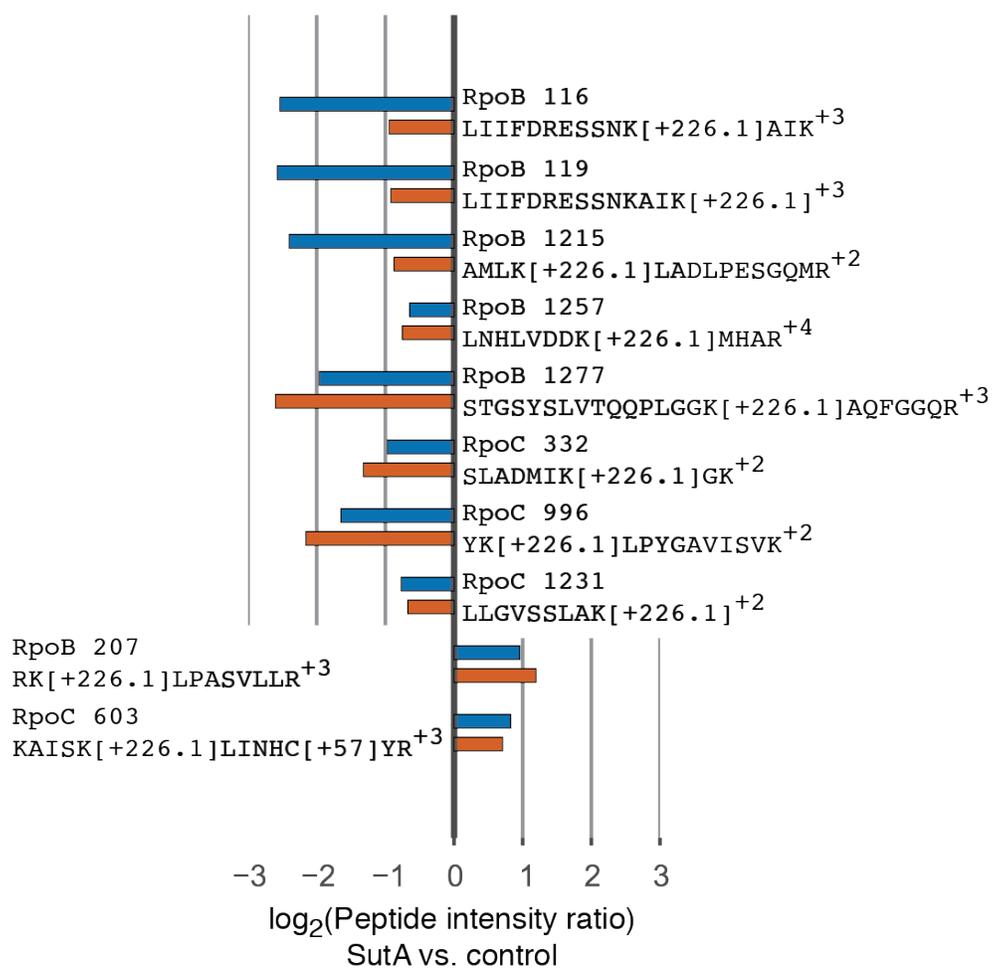


Figure 3.S6: **Protein foot-printing.** Lysine residues determined to be obscured (top) or revealed (bottom) in the presence of SutA. Peptide intensity ratios between RNAP with SutA vs. RNAP alone are shown for each replicate (blue or orange). The modified residue, modified peptide sequence, and charge state are listed next to the bars for each intensity ratio.

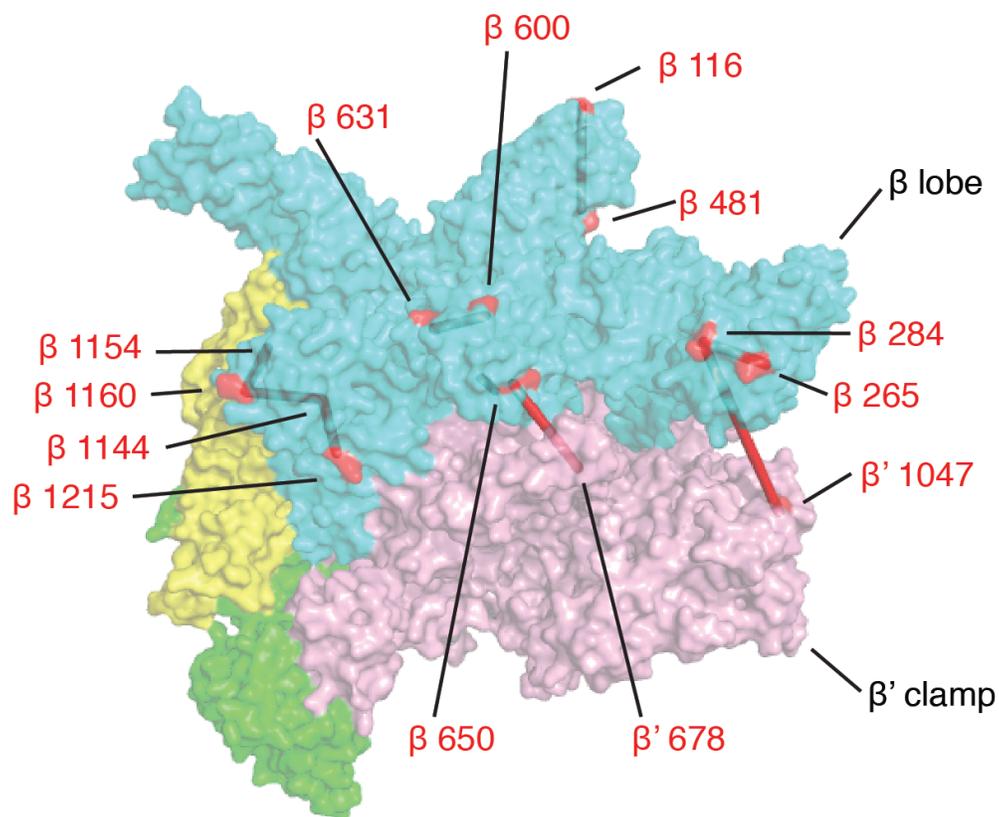


Figure 3.S7: **Intra-RNAP BS<sup>3</sup> cross-links.** Cross-links were mapped onto the *E. coli* structure. Cross-linked lysines are colored red, and inter  $\alpha$ -carbon distances are displayed as red bars. Eight detected intra-RNAP cross-links are shown; the ninth is located on the opposite face of the structure. Subunits are colored as follows:  $\alpha_I$  (yellow),  $\alpha_{II}$  (green),  $\beta$  (cyan), and  $\beta'$  (pink).

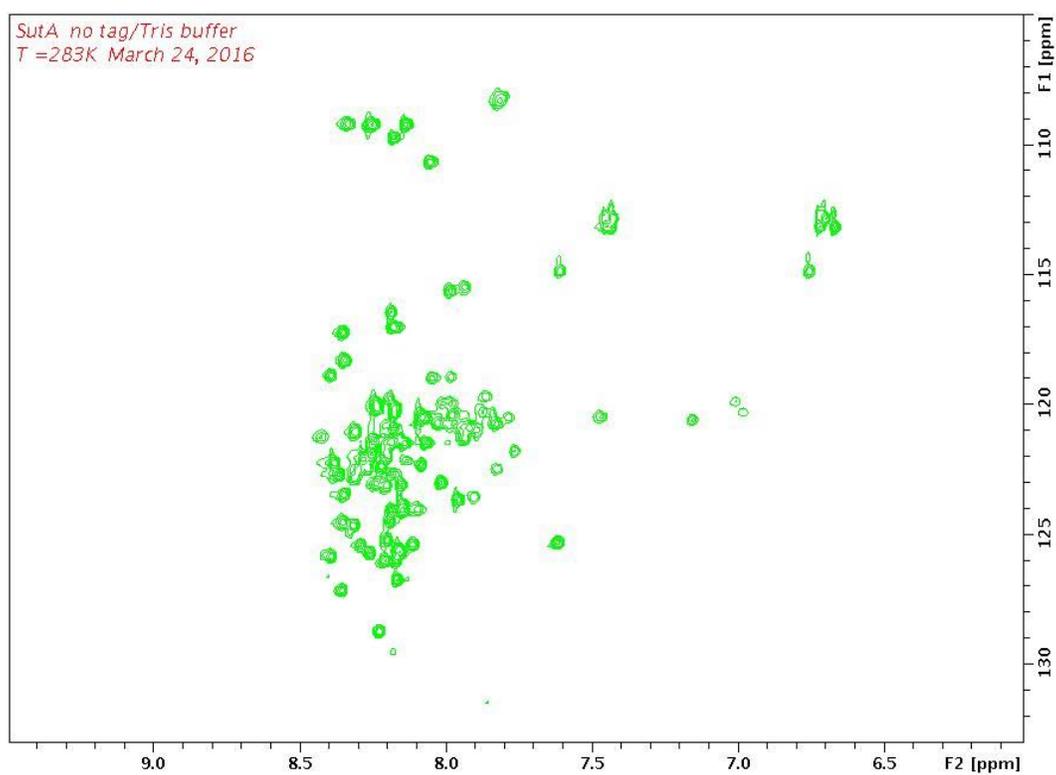


Figure 3.S8: **2D NMR of SutA.** HSQC spectrum of  $^{15}\text{N}$ -labeled SutA. The horizontal and vertical axes represent  $^1\text{H}$  and  $^{15}\text{N}$  chemical shifts, respectively. Provided by Dr. Ben Ramirez.

### 3.9 Supplementary Tables.

Table 3.3: **Chapter 3: Strains and plasmids.** Strains and plasmids used in this study. Plasmids are stored as *E. coli* strains carrying the plasmid, and requests should be for the *E. coli* strain.

<i>Pseudomonas aeruginosa</i> Strains		
Name	Genotype	Source
DKN263	<i>P. aeruginosa</i> UCBPP-PA14	
DKN1625	UCBPP-PA14 $\Delta sutA$	[1]
<i>Escherichia coli</i> Strains		
Name	Genotype	Source
DKN1640	Mach1 pMQ72_HA-sutA	[1]
DKN1687	Mach1 pMQ72_HA-sutA_Nterm	This Study
DKN1688	Mach1 pMQ72_HA-sutA_Cterm	This Study
DKN548	DH5 $\alpha$ pMQ72	George O'Toole
DKN1697	BL21 DE3, pQE80L-6xHis-TEV-SutA	This Study
BMB14X	BL21 DE3, pQE80L-6xHis-TEV-SutA with XBpa	This Study
pEVOL-pBpF	BL21 DE3, p15A-pBpa synthetase and tRNA	[15]

Table 3.4: **Bpa mass modifications** Mass modifications that were used to search for Bpa cross-links. "B" represents the location of Bpa in the peptide sequence.

SutA Mutant	Peptide	Mass
L6	BE	398.1478
L11	BD	384.1321
L22	BAAAD	597.2434
L54	BPSVE	681.3010
I84	BE	398.1478

## References

- (1) Babin, B. M.; Bergkessel, M.; Sweredoski, M. J.; Moradian, A.; Hess, S.; Newman, D. K.; Tirrell, D. A. *Proc. Natl. Acad. Sci. U.S.A.* **2016**, *113*, E597–E605.
- (2) Dennis, P. P.; Bremer, H. *EcoSal Plus* **2013**, *1*.
- (3) Paul, B. J.; Ross, W.; Gaal, T.; Gourse, R. L. *Ann. Rev. Genet.* **2004**, *38*, 749–770.
- (4) Condon, C.; Squires, C.; Squires, C. L. *Microbiol. Rev.* **1995**, *59*, 623–645.
- (5) Lemke, J. J.; Sanchez-Vazquez, P.; Burgos, H. L.; Hedberg, G.; Ross, W.; Gourse, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 5712–5717.
- (6) Belogurov, G. A.; Artsimovitch, I. *Ann. Rev. Microbiol.* **2015**, *69*, 49–69.
- (7) Zenkin, N.; Yuzenkova, Y. *Biomolecules* **2015**, *5*, 1195–1209.
- (8) López de Saro, F. J.; Woody, A. Y.; Helmann, J. D. *J. Mol. Biol.* **1995**, *252*, 189–202.
- (9) Withers, R.; Doherty, G. P.; Jordan, M.; Yang, X.; Dixon, N. E.; Lewis, P. J. *Mol. Microbiol.* **2014**, *93*, 1130–1143.
- (10) Srivastava, D. B.; Leon, K.; Osmundson, J.; Garner, A. L.; Weiss, L. A.; Westblade, L. F.; Glickman, M. S.; Landick, R.; Darst, S. A.; Stallings, C. L.; Campbell, E. A. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 12619–12624.
- (11) Linding, R.; Jensen, L. J.; Diella, F.; Bork, P.; Gibson, T. J.; Russell, R. B. *Structure* **2003**, *11*, 1453–1459.
- (12) Kelley, L. A.; Sternberg, M. J. E. *Nat. Protoc.* **2009**, *4*, 363–371.
- (13) Chen, Z. A.; Jawhari, A.; Fischer, L.; Buchen, C.; Tahir, S.; Kamenski, T.; Rasmussen, M.; Lariviere, L.; Bukowski-Wills, J.-C.; Nilges, M.; Cramer, P.; Rappsilber, J. *EMBO J.* **2010**, *29*, 717–726.
- (14) Lane, W. J.; Darst, S. A. *J. Mol. Biol.* **2010**, *395*, 671–685.
- (15) Chin, J. W.; Martin, A. B.; King, D. S.; Wang, L.; Schultz, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 11020–11024.
- (16) Liu, B.; Zuo, Y.; Steitz, T. A. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112*, 2006–2010.
- (17) Shell, S. M.; Hess, S.; Kvaratskhelia, M.; Zou, Y. *Biochemistry* **2005**, *44*, 971–978.
- (18) Opalka, N.; Brown, J.; Lane, W. J.; Twist, K.-A. F.; Landick, R.; Asturias, F. J.; Darst, S. A. *PLoS Biol.* **2010**, *8*, e1000483.
- (19) Lennon, C. W.; Ross, W.; Martin-Tumasch, S.; Touloukhonov, I.; Vrentas, C. E.; Rutherford, S. T.; Lee, J.-H.; Butcher, S. E.; Gourse, R. L. *Genes Dev.* **2012**, *26*, 2634–2646.

- (20) Murakami, K. S.; Masuda, S.; Campbell, E. A.; Muzzin, O.; Darst, S. A. *Science* **2002**, *296*, 1285–1290.
- (21) Vassilyev, D. G.; Vassilyeva, M. N.; Perederina, A.; Tahirov, T. H.; Artsimovitch, I. *Nature* **2007**, *448*, 157–162.
- (22) Chakraborty, A.; Wang, D.; Ebright, Y. W.; Korlann, Y.; Kortkhonjia, E.; Kim, T.; Chowdhury, S.; Wigneshweraraj, S.; Irschik, H.; Jansen, R.; Nixon, B. T.; Knight, J.; Weiss, S.; Ebright, R. H. *Science* **2012**, *337*, 591–595.
- (23) Furman, R.; Tsodikov, O. V.; Wolf, Y. I.; Artsimovitch, I. *J. Mol. Biol.* **2013**, *425*, 82–93.
- (24) Lennon, C. W.; Gaal, T.; Ross, W.; Gourse, R. L. *J. Bacteriol.* **2009**, *191*, 5854–5858.
- (25) Sevostyanova, A.; Belogurov, G. A.; Mooney, R. A.; Landick, R.; Artsimovitch, I. *Mol. Cell* **2011**, *43*, 253–262.
- (26) Burmann, B. M.; Schweimer, K.; Luo, X.; Wahl, M. C.; Stitt, B. L.; Gottesman, M. E.; Rösch, P. *Science* **2010**, *328*, 501–504.
- (27) Chambers, M. C. et al. *Nat. Biotechnol.* **2012**, *30*, 918–920.
- (28) Kuznedelov, K.; Semenova, E.; Knappe, T. A.; Mukhamedyarov, D.; Srivastava, A.; Chatterjee, S.; Ebright, R. H.; Marahiel, M. A.; Severinov, K. *J. Mol. Biol.* **2011**, *412*, 842–848.
- (29) Kalli, A.; Hess, S. *Proteomics* **2012**, *12*, 21–31.
- (30) Trnka, M. J.; Baker, P. R.; Robinson, P. J. J.; Burlingame, A. L.; Chalkley, R. J. *Mol. Cell Proteomics* **2014**, *13*, 420–434.
- (31) Kim, S.; Pevzner, P. A. *Nat. Commun.* **2014**, *5*, 5277.
- (32) Schilling, B.; Rardin, M. J.; MacLean, B. X.; Zawadzka, A. M.; Frewen, B. E.; Cusack, M. P.; Sorensen, D. J.; Bereman, M. S.; Jing, E.; Wu, C. C.; Verdin, E.; Kahn, C. R.; Maccoss, M. J.; Gibson, B. W. *Mol. Cell Proteomics* **2012**, *11*, 202–214.
- (33) Hunter, J. D. *Comput. Sci. Eng.* **2007**, *9*, 90–95.
- (34) Schneider, C. A.; Rasband, W. S.; Eliceiri, K. W. *Nat. Methods* **2012**, *9*, 671–675.
- (35) Hamelryck, T.; Manderick, B. *Bioinformatics* **2003**, *19*, 2308–2310.

*Chapter 4*PROTEOMIC RESPONSE OF AN ANTIBIOTIC-TOLERANT  
BIOFILM SUBPOPULATION TO CIPROFLOXACIN**4.1 Abstract**

Chronic biofilm infections are of particular concern due to their increased tolerance to antibiotics. The study of biofilms *in vitro* is complicated by phenotypic heterogeneity which results in biofilm subpopulations with differential responses to antibiotics. To better understand the physiology of tolerance, we adapt the bioorthogonal noncanonical amino acid tagging (BONCAT) method for selective proteomics to target a *Pseudomonas aeruginosa* biofilm subpopulation. We target protein labeling to interior regions of biofilm microcolonies through the use of the endogenous *rpoS* promoter and show successful enrichment and identification of newly synthesized proteins from this region. We perform a pulse-labeling experiment to measure the dynamic proteomic response of this tolerant subpopulation to supra-MIC treatment with the fluoroquinolone ciprofloxacin. We find upregulation of proteins involved in flagellar motility and purine synthesis as well as a substantial rearrangement of enzymes involved in central carbon metabolism.

## 4.2 Introduction

Bacteria living as surface-associated biofilms exhibit increased tolerance to a wide-variety of stresses as compared to their planktonic counterparts [1]. Of clinical importance is the increased tolerance to antibiotics that prohibits the elimination of chronic biofilm infections. This phenotypic tolerance is distinct from genotypic resistance, though the evolution of resistance can be exacerbated by the persistence of cells that survive treatment [2]. The opportunistic pathogen *Pseudomonas aeruginosa* serves as a model system for both biofilm physiology as well as antibiotic tolerant infection. *P. aeruginosa* is a primary contributor to chronic infections of the cystic fibrosis lung, where it forms biofilms that are recalcitrant to the host immune system and treatment by antibiotics. Tolerance of these biofilm infections has been well characterized within the host [3], and through *in vitro* studies of biofilms [4].

Moreover, detailed analyses of bacterial biofilms grown *in vitro* have revealed the role of spatial heterogeneity in their response to antibiotics; specific subpopulations survive treatment while others do not [1, 5]. Similar to their differential effects on fast growing vs. slow growing planktonic cells, drug classes like fluoroquinolones (DNA replication) [6], aminoglycosides (translation) [7], and  $\beta$ -lactams (peptidoglycan synthesis) [8] that target active processes kill growing cells within biofilm regions that have greater access to exogenous nutrients. Conversely, polymyxins and detergents that disrupt cellular membranes preferentially kill dormant cells in the interior of biofilm microstructures [6]. Explanations for the spatial segregation of antibiotic responses include the reduced penetration of small molecule antibiotics, lowered metabolic rates, and altered physiology [1, 5].

Measurements of mRNA or protein abundances have offered unbiased views of physiological responses to antibiotics [9–11] but a variety of challenges limit the investigation of tolerant biofilm subpopulations. Because only a subpopulation of cells exhibit tolerance, any analysis ought to distinguish tolerant cells from those that do not survive treatment. Laser capture micro dissection has been used to isolate biofilm cells from spatially distinct regions of biofilms and quantitative PCR (qPCR) and DNA microarray analyses have been used to quantify differences in mRNA transcript abundances [12, 13]. This approach has not been applied to understand the heterogeneous response to antibiotics. Proteomic measurements have been widely used to better understand biofilm physiology [14], but selective approaches have been limited. Physical separation is challenging due to the small length scales

involved, and manual selection of subpopulations of interest can introduce human error into the analysis. In addition, an open question is what differences exist, if any, between the instantaneous response to antibiotic stress and the long-term phenotypic adaptation.

An important recent technological advance is the application of pulsed stable isotope labeling with amino acids (pSILAC) to quantify changes in protein expression following adaptation of biofilm cells to challenge with the clinical polymixin antibiotic colistin [15]. By pulsed addition of an amino acid isotoplog, pSILAC provides a means to distinguish—by mass—proteins synthesized before and after the pulse [16]. Chua, *et al.* treated biofilms with colistin for 8 h thereby allowing non-tolerant cells to die, and then labeled new protein synthesis with an extended (48 h) amino acid isotoplog pulse, ensuring that labeled proteins were made by the tolerant subpopulation of interest. This approach revealed an importance for type IV-mediated motility in the resistance to colistin.

To address the challenges of phenotypic heterogeneity and dynamic responses, we employed the bioorthogonal noncanonical amino acid tagging (BONCAT) method for selective proteomics [17, 18]. BONCAT relies on the cellular incorporation of a non-canonical amino acid (ncAA) that bears a bioorthogonal chemical handle. Following incorporation, labeled proteins can be reacted to an affinity tag and enriched from the pool of unlabeled proteins. Enriched proteins can be identified and quantified via liquid chromatography-tandem mass spectrometry (LC-MS/MS). Like pSILAC, BONCAT allows for temporal selectivity; proteins synthesized during the ncAA pulse are chemically distinct from pre-existing proteins. However, a key advantage of the enrichment-based proteomics method is that proteins of interest can be physically separated from the rest of the proteome. MS-based protein identification is sensitive to the complexity of the sample, such that proteins of low abundance often go unidentified, so reducing sample complexity can aid in the identification of proteins of interest. We and others have shown the exquisite temporal sensitivity of BONCAT-based enrichment in the context of dynamic proteome changes [19, 20]. In bacteria, ncAA pulse times of a few minutes have been used to quantify dynamic processes in *Vibrio harveyi* [21, 22], *Escherichia coli* [23], and *Bacillus subtilis* [24].

BONCAT labeling can be targeted with greater precision to cell types of interest through the use of a ncAA that is not incorporated by endogenous translational machinery. However, cells expressing a mutant aminoacyl-tRNA synthetase (mRS)

that has been engineered to charge this ncAA will be labeled. Such noncanonical synthetases have been developed for the methionine surrogates azidonorleucine [25] and 2-aminooctynoic acid [26] and the phenylalanine surrogate azidophenylalanine [27]. By restricting expression of the mRS to cell types of interest, protein labeling can be targeted to a subpopulation of cells within a complex heterogeneous system. In bacteria, cell targeting can be accomplished by genetically restricting the mRS gene to a species of interest (e.g., bacteria in the presence of host cells [26, 28]) or by placing mRS expression under endogenous control of a cell-state specific promoter (e.g., reactive oxygen stress in *E. coli* [29]).

Here, we describe an adaptation of the BONCAT method for cell- and time-resolved analysis of protein synthesis in heterogeneous bacterial biofilms. We direct cell selective labeling of protein synthesis with azidonorleucine (Anl) through controlled expression of its corresponding mutant methionyl-tRNA synthetase (NLL-MetRS). We use this approach to analyze the time course of proteomic responses to ciprofloxacin stress by a biofilm subpopulation of *P. aeruginosa*.

### 4.3 Results

#### The *rpoS* Promoter Enables Cell-state Selective Labeling in Planktonic Cells

To selectively target antibiotic-tolerant biofilm cells, we aimed to restrict labeling by placing NLL-MetRS expression under control of an endogenous, cell-state selective promoter. Because regions more tolerant to many antibiotics contain cells with decreased metabolic rates, we reasoned that the use of a promoter whose activity increases during planktonic stationary phase, when metabolic rates are similarly decreased, might provide the desired selectivity. Cellular levels of the alternative sigma factor  $\sigma^{54}$  are upregulated in response to a variety of stresses. In *P. aeruginosa* and other bacteria  $\sigma^{54}$ —encoded by the gene *rpoS*—is upregulated during the transition from exponential to stationary phase during planktonic growth [30]. We hypothesized that the *rpoS* promoter would enable selective protein labeling.

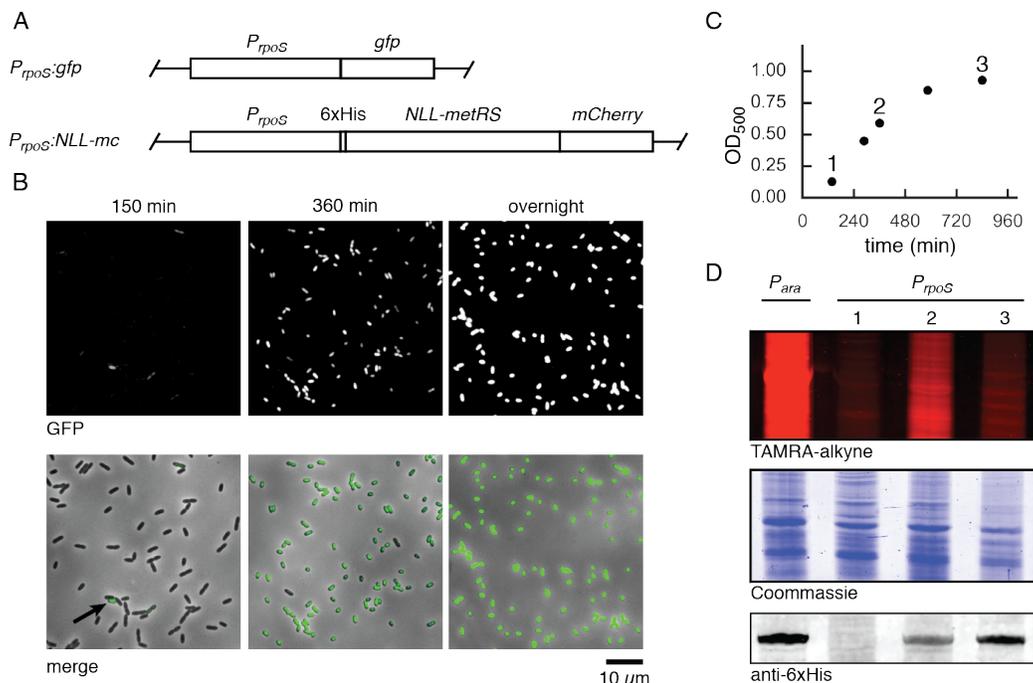
We first evaluated the activity of the *rpoS* promoter in planktonic cells. We cloned the 1 kb region upstream of the endogenous *rpoS* gene 5' to *gfp* and transposed this expression cassette to the *Tn7* locus in *P. aeruginosa* PA14 ( $P_{rpoS}:gfp$ ) (Figure 4.1A). Fluorescent imaging of  $P_{rpoS}:gfp$  throughout growth from early exponential phase (150 min following dilution) to late stationary phase (overnight) revealed the expected increase of promoter activity at higher cell densities (Figure 4.1B). Additionally, we noticed cellular heterogeneities in the expression of GFP. At the early

time point, only a small subpopulation of cells was GFP positive. The GFP-positive fraction increased in the early exponential time point and after overnight growth essentially all cells were expressing GFP. In contrast, wild-type cells exhibited no fluorescence (Figure 4.S1A) and when GFP expression was placed under control of the strong, constitutive *trc* promoter [31], all cells were GFP-positive at all time points (Figure 4.S1B).

Encouraged by these results, we generated a strain in which expression of an NLL-MetRS:mCherry translational fusion was controlled by the *rpoS* promoter,  $P_{rpoS}:nll-mc$  (Figure 4.1A). We appended this protein with an N-terminal 6x histidine tag to allow for Western blot detection of protein expression. We grew this strain from early exponential phase to stationary phase and treated samples of the culture with Anl for 15 min at three points throughout growth (Figure 4.1C). As a positive control, we treated PA14 containing a plasmid expressing NLL-MetRS under control of the  $P_{ara}$  arabinose-inducible promoter with arabinose and Anl during exponential phase. Consistent with our GFP measurements, Western blotting showed growth phase-dependent expression of NLL-MetRS when controlled by the *rpoS* promoter (Figure 4.1D). To detect Anl incorporation, we reacted cell lysates with alkyne-tetramethylrhodamine (TAMRA) under copper catalyzed click conditions, separated proteins via SDS-PAGE, and imaged fluorescence. In early exponential phase, when NLL-MetRS was not present, labeling was not detected. Labeling was strongest in late exponential phase, when NLL-MetRS expression was moderate, and low but detectable in stationary phase. For all conditions Coomassie staining was used to verify equal protein loading. The rate of Anl incorporation is dependent on both the presence of the NLL-MetRS and the overall rate protein synthesis, so we interpret the observed lower levels of Anl labeling in stationary phase as a reflection of the decreased rates of protein synthesis in this state.

### **Spatially Targeted Proteomics in Biofilms**

To test for subpopulation targeting in biofilms, we cultured biofilms on glass coverslips under constant media flow in flow cells. Four day old biofilms were treated with Anl for 1.5 h and Anl incorporation was visualized by treating fixed biofilms with dibenzocyclooctyne (DBCO)-TAMRA. The strained-alkyne present in DBCO allows for copper-free azide-alkyne cycloaddition and removes the requirement for simultaneous diffusion of the ligand, copper catalyst, and reductant throughout biofilm microcolonies. A strain expressing NLL-MetRS constitutively ( $P_{trc}:nll-mc$ ) was labeled throughout biofilm structures, while  $P_{rpoS}:nll-mc$  exhibited labeling



**Figure 4.1: Cell state-selective labeling using the *rpoS* promoter.** (A) *P. aeruginosa* was engineered to express GFP or an NLL-MetRS:mCherry translational fusion under control of the endogenous *rpoS* promoter. Expression cassettes were transposed to the neutral *Tn7* neutral chromosomal locus. (B) Representative images of GFP fluorescence of the *P<sub>rpoS</sub>:gfp* strain throughout growth. GFP fluorescence (top) and GFP-bright field merge (bottom). Arrow indicates a GFP-positive cell in the early time point. (C) Growth curve of *P<sub>rpoS</sub>:nll-mc*. At each labeled time point, an aliquot was removed and incubated with 1 mM AnI for 15 min. (D) Lysates were reacted with alkyne-TAMRA and separated via SDS-PAGE to visualize AnI incorporation. Coomassie staining of the same gel indicates equal protein loading. Lysates were also probed by Western blot for the 6x-histidine tag on NLL-MetRS.

only in the base of structures, close to the glass coverslip. Wild-type cells exhibited minimal background fluorescent signal (Figure 4.2A).

To evaluate our ability to detect proteins preferentially expressed by the labeled subpopulation, we compared proteomes enriched from *P<sub>trc</sub>:nll-mc* and *P<sub>rpoS</sub>:nll-mc* strains. To obtain adequate protein yield, biofilms of each strain were grown for four days in silicone tubing and treated with AnI for 1.5 h. We verified AnI incorporation in cell lysates via SDS-PAGE (Figures 4.S2A), reacted lysates with DBCO-biotin (Figures 4.S2B), enriched proteins on streptavidin beads, and analyzed proteins via LC-MS/MS. We detected 908 total proteins among two replicates from each strain. Proteins enriched from *P<sub>rpoS</sub>:nll-mc* made up a subset of total protein identifications (80%) (Figure 4.2B). Based on fluorescence and Western blot detections (Figure

4.S2A-B), AnI incorporation was not significantly lower in these samples, so the decreased complexity of the sample is likely due to the targeted analysis of a subset of cells. For proteins identified from both strains, we quantified their relative abundances using label-free quantification (LFQ) and found 15 and 24 proteins to be at least two-fold more or less abundant in the *P<sub>rpoS</sub>:nll-mc* strain, respectively (Figure 4.2C). Combined with proteins uniquely identified from either strain, a total of 24 and 206 proteins were more abundant in the *rpoS* or *trc* samples, respectively. Full proteomic results are listed in Dataset 4.1.

To take a functional view of region-specific expression profiles, we categorized proteins by their PseudoCAP classification [32] and searched for categories that were significantly overrepresented in each list of “hits” compared to all proteins identified. Proteins classified as transcriptional regulators were significantly more abundant in the *trc* sample, a set that includes regulators involved in low oxygen response (Anr) [33], quorum sensing (LasR) [34], flagellin synthesis (FliA) [35], and the global regulator Vfr [36]. A protein of particular interest is AmgR, a response regulator whose deletion causes increased sensitivity to the clinical aminoglycoside tobramycin [37].

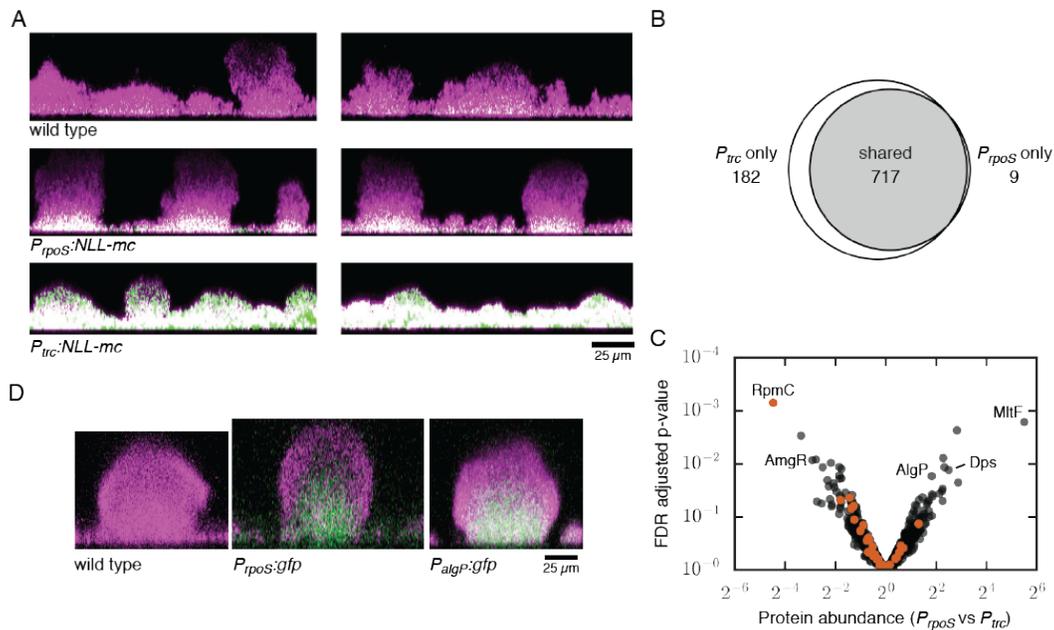
We found ribosomal proteins to be significantly less abundant in the *rpoS* region ( $p < 0.05$ ); the median relative abundance of 36 quantified ribosomal proteins was 0.8 fold lower in *rpoS* samples compared to the *trc* samples (Figure 4.2D, Figure 4.S2C). Furthermore, the protein with the lowest relative abundance in the *rpoS* samples was ribosomal protein RpmC (protein L29 of the 60S subunit, 22-fold less abundant). These results are consistent with measurements showing greater translational activity in upper regions of flow cell biofilms [38] and the higher levels of ribosomal transcripts in the region containing metabolically active cells found in a microarray comparison of regions within colony biofilms [13].

Of the few proteins significantly more abundant in the *rpoS* sample, more than half were annotated as hypothetical, unclassified, or unknown, a significant enrichment of that PseudoCAP classification as compared to all identified proteins. The proteins found whose functions are known include those involved in antibiotic resistance, stress protection, and alginate regulation (highlighted in Figure 4.2C). MltF, annotated as PA14\_15720, was the protein most enriched in the *rpoS* subpopulation and is known to play a role in resistance to the  $\beta$ -lactams piperacillin, cefotaxime and ceftazidime; disruption of the gene leads to reduced minimum inhibitory concentrations for each antibiotic in planktonic cultures [39]. Biofilm resistance to  $\beta$ -lactams

has been linked previously to the upregulation of the  $\beta$ -lactamase AmpC by peripheral cells in response to antibiotic treatment [8]. The identification of MltF in the absence of antibiotic stimulation and within the biofilm interior, gives evidence for a complementary approach to tolerance in which cells are preemptively prepared for antibiotic stress. We also identified Dps, a nonspecific DNA-remodeling protein that confers protection against a variety of stresses, including starvation, peroxide treatment, UV irradiation, and others [40, 41]. In *E. coli*, *dps* transcription is RpoS dependent, and Dps is one of the most abundant proteins in stationary phase cells [42].

Another DNA binding protein, AlgP, is one of many known regulators that control synthesis of the exopolysaccharide alginate [43]. Alginate synthesis is not required for biofilm formation in vitro [44] but its upregulation is one of the hallmarks of the “mucoïd” phenotype observed in some CF isolates. Besides its structural role as a component of the extracellular polymeric substance (EPS) that surrounds biofilm cells, alginate may play a role in defense against inflammation caused by host cells in the CF lung [43]. To validate the ability of our targeted proteomics approach to provide information about region-specific protein expression, we generated a strain that expresses GFP under control of the *algP* promoter (*P<sub>algP</sub>:gfp*). After four days of growth, GFP fluorescence in *P<sub>algP</sub>:gfp* biofilms was localized to cells within the biofilm interior (Figure 4.2D). This pattern of expression matched the localization of GFP fluorescence in *P<sub>rpoS</sub>:gfp* biofilms and the localization of AnI labeling observed in *P<sub>rpoS</sub>:nll-mc* (Figure 4.2A).

As a caveat, we note that RpoS itself was equally abundant in the *rpoS* and *trc* samples. The design of the expression cassette in *P<sub>rpoS</sub>:nll-mc* places NLL-MetRS under transcriptional control of any regulatory regions that lie 1 kb upstream of the endogenous *rpoS* gene. However, much of the control of RpoS protein levels is known to be post-transcriptional, depending on the action of sRNAs, modified translation rates, and tuned degradation [45]. Additionally, NLL-MetRS has a C-terminal fusion to mCherry which may increase its intracellular stability and may further disconnect levels of the mutant synthetase and RpoS itself. We conclude that *P<sub>rpoS</sub>:nll-mc* cells with high levels of NLL-MetRS are not necessarily cells with high levels of RpoS protein. However, our imaging results from planktonic and biofilm growth states show that the *P<sub>rpoS</sub>:nll-mc* strain can be used to target proteomic analysis to the cellular subpopulation of interest.



**Figure 4.2: Targeted proteomics of a biofilm subpopulation.** (A) AnI incorporation in wild-type,  $P_{rpoS}:nll-mc$ , and  $P_{trc}:nll-mc$  biofilms was visualized by reacting fixed biofilms with DBCO-TAMRA (green). Biofilms were counter stained with SYTO9 (magenta). Colocalization of fluorescent signals is displayed as white. Cross-sections were reconstructed from confocal image stacks. (B) Proteins identified following BONCAT enrichment from  $P_{rpoS}:nll-mc$  and  $P_{trc}:nll-mc$  strains. (C) Quantification of relative protein abundances for proteins found following enrichment from both strains. Ribosomal proteins are shown in orange. Proteins discussed in the text are indicated by name. (D) Spatial distribution of GFP expression under control of the  $rpoS$  or  $algP$  promoters in live biofilms. GFP is shown in green and SYTO62 counterstain in magenta.

### BONCAT Enrichment of Proteins Throughout Ciprofloxacin Treatment.

To identify the subpopulation-specific response to ciprofloxacin, we designed an experiment to capture dynamic changes to the proteome throughout the course of antibiotic challenge. Fluorescent imaging of biofilms treated with ciprofloxacin has shown a progression of cell death over the course of 13 h [6]. Cell death, visualized by propidium iodide staining, began between 4 and 9 h of treatment, and was restricted to peripheral regions of biofilm microstructures. Protein synthetic activity, measured by detection of expression of an unstable GFP variant, continued in interior biofilm populations even after 13 h of treatment. We replicated this time course of antibiotic challenge by treating four day-old  $P_{rpoS}:nll-mc$  biofilms with ciprofloxacin. To achieve temporal selectivity, we pulse-labeled biofilms with AnI at 0, 4, or 13 h after ciprofloxacin was added. Each pulse was for 1.5 h to distinguish

newly synthesized proteins from the preexisting proteome. To serve as a “no cipro.” control, we also labeled untreated biofilms for 1.5 h with Anl (Figure 4.3A).

We tracked the number of viable cells recovered from biofilms throughout the time course of treatment and found two stages of killing (Figure 4.3B). Compared to the untreated control, biofilms treated for 1.5 or 5.5 h exhibited an approximately 50-fold loss in viable cells, while those treated for 14.5 h exhibited a 500-fold loss. Treatment with Anl had no effect on the number of viable cells. In contrast, levels of Anl incorporation from the same samples showed a decrease with longer ciprofloxacin treatment, but varied less than two-fold (Figure 4.3C). This apparent discrepancy between viable cell counts and protein synthetic activity is likely due to a number of factors. Since we have shown that the *P<sub>rpoS</sub>:nll-mc* strain allows targeted labeling of a small subpopulation of biofilm cells in the region known to tolerate ciprofloxacin treatment, we expect that cells incorporating Anl will be viable throughout treatment. Second, viability counts by plating do not necessarily reflect the number of translationally active cells, exemplified by the body of literature documenting so called viable but non-culturable cells [46] particularly in the context of antibiotic persistence [47]. Finally, the total amount of Anl incorporation in a given time is dependent on both the number of translationally active cells as well as the overall rate of translation, as shown above for *P<sub>rpoS</sub>:nll-mc* labeling in planktonic cultures; a lower level of Anl incorporation reflects either a reduction in active cells, a reduction in the protein synthesis of those cells, or some combination.

We performed BONCAT enrichment on lysates from each experimental condition (performed in triplicate) and identified proteins by LC-MS/MS. We identified more than 1200 proteins among all runs. Protein abundances, estimated by LFQ, were well correlated between experimental replicates (Figure 4.3D). We used principal component (PC) analysis to visualize the variance among replicates and experimental conditions (Figure 4.3E) and found that, in general, biological replicates clustered with one another and that ciprofloxacin treated samples were separated from the untreated control samples. For each time point, we quantified differences in protein abundances compared to the untreated control (Figure 4.3F). Consistent with the correlation analysis, fewer proteins were significantly changed in the 1.5 h treatment condition (73 proteins) than in the 5.5 h (187 proteins) or 14.5 h (204 proteins) treatment conditions (Figure 4.S3). Correlation, PC, and quantification analyses are consistent with the classification of the proteomic data into two subgroups: proteins whose rates of synthesis change immediately upon ciprofloxacin

exposure, and a later response characterized by a greater number of changes, many of which are shared between the 5.5 h and 13.5 h groups.

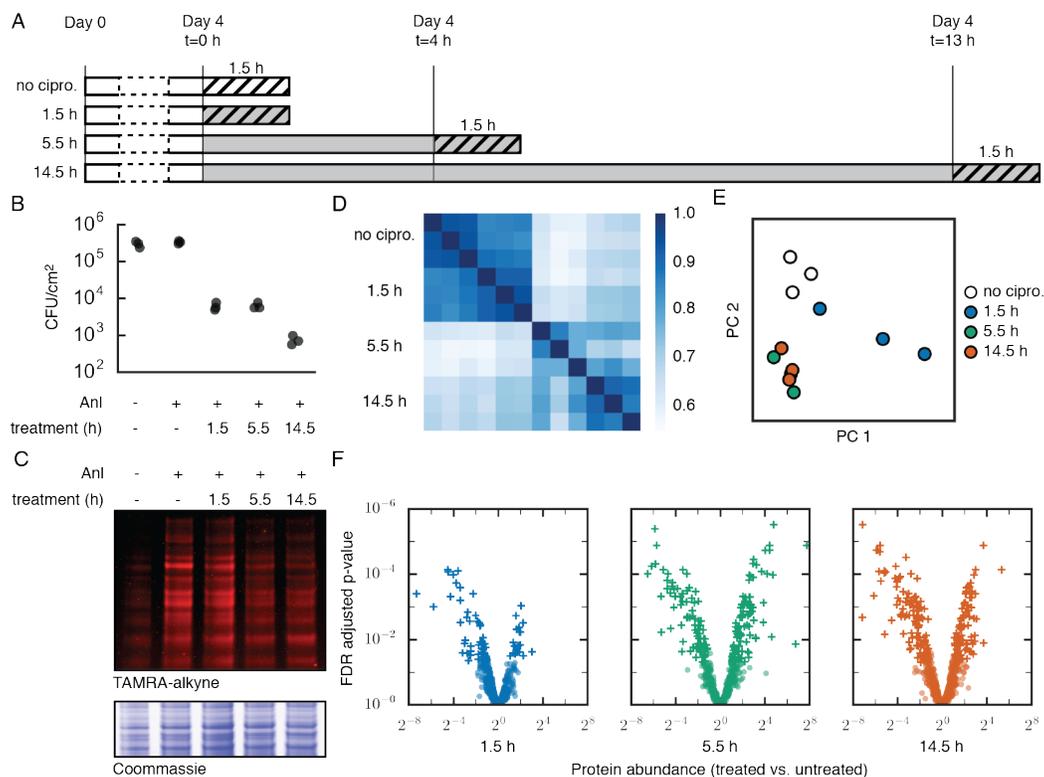
For the following analyses, we subset proteins identified at each time point into groups significantly more or less abundant in the treated vs. untreated control. Each group includes proteins whose relative abundances were quantified (fold-change > 2 and FDR adjusted p-value < 0.05) as well as proteins identified in one condition (at least two of three replicates) and not identified in the other (not found in any replicates). Full proteomic results are listed in Dataset 4.2.

### **The Dynamic Proteomic Response to Ciprofloxacin.**

We first compared our results to other studies of the *P. aeruginosa* response to ciprofloxacin. The responses of planktonic *P. aeruginosa* to sub-MIC to supra-MIC ciprofloxacin treatments (0.01 to 1.0  $\mu\text{g}/\text{mL}$ ) have been characterized via microarray measurements of transcript abundances [10, 48, 49]. While the design of these experiments differed, a small consensus of responses to ciprofloxacin has emerged. These changes include the upregulation of proteins involved in the SOS response and DNA repair (e.g., RecA) and the pyocin synthesis regulator, PrtN. We found that our dataset generally matched these reported changes; RecA, PrtN, the negative regulator of type III secretion, PtrB (PA14\_07970), and the ribonucleotide reductase complex (NrdAB) were either significantly upregulated or uniquely identified in the treated samples (Figure 4.S4A).

As described above, early and late responses differed in our dataset. In fact, only three proteins were significantly upregulated at all times: the DNA gyrase GyrB, a direct target of ciprofloxacin; the protein chaperone HscK; and the methylisocitrate lyase PrpB. These proteins typify functional categories of proteins we found to be upregulated by ciprofloxacin challenge, namely those involved in remediating DNA damage and other stress, and proteins involved in central metabolism.

Ciprofloxacin inhibits DNA gyrase activity, causing DNA damage during replication attempts. DNA damage leads to an induction of the SOS response, characterized by depletion of the SOS repressor LexA and the resulting upregulation of genes that alleviate DNA damage stress [50]. We did not identify LexA in any experiments, but we detected upregulation of the LexA target RecA which binds to DNA lesions. Of the proteins involved in DNA damage response and repair found in our dataset, all were either significantly upregulated or uniquely identified in at least one time point (Figure 4.4A, Figure 4.S4B). This set includes both subunits of DNA gyrase



**Figure 4.3: BONCAT analysis of protein synthesis during ciprofloxacin challenge.** (A) Experimental timeline of biofilm treatment and proteome labeling. Biofilms were grown in silicone tubing for four days and then treated with ciprofloxacin (gray bars). Control biofilms were untreated. For each experimental condition, biofilms were treated with Anl at the designated time point for 1.5 h (cross hatch) and then lysed. (B) Survival of biofilm cells following treatment with 60  $\mu$ g/mL ciprofloxacin for the indicated time. (C) Visualization of Anl incorporation. (D) Spearman rank correlation coefficients for protein LFQ values, calculated among all MS runs. (E) Coordinates for each MS run in two dimensional principal component space. (F) Protein abundance fold changes for each experimental condition compared to the ciprofloxacin untreated control. Proteins that are significantly more or less abundant are shown as crosses (FDR adjusted p-value < 0.05 and fold change > 2).

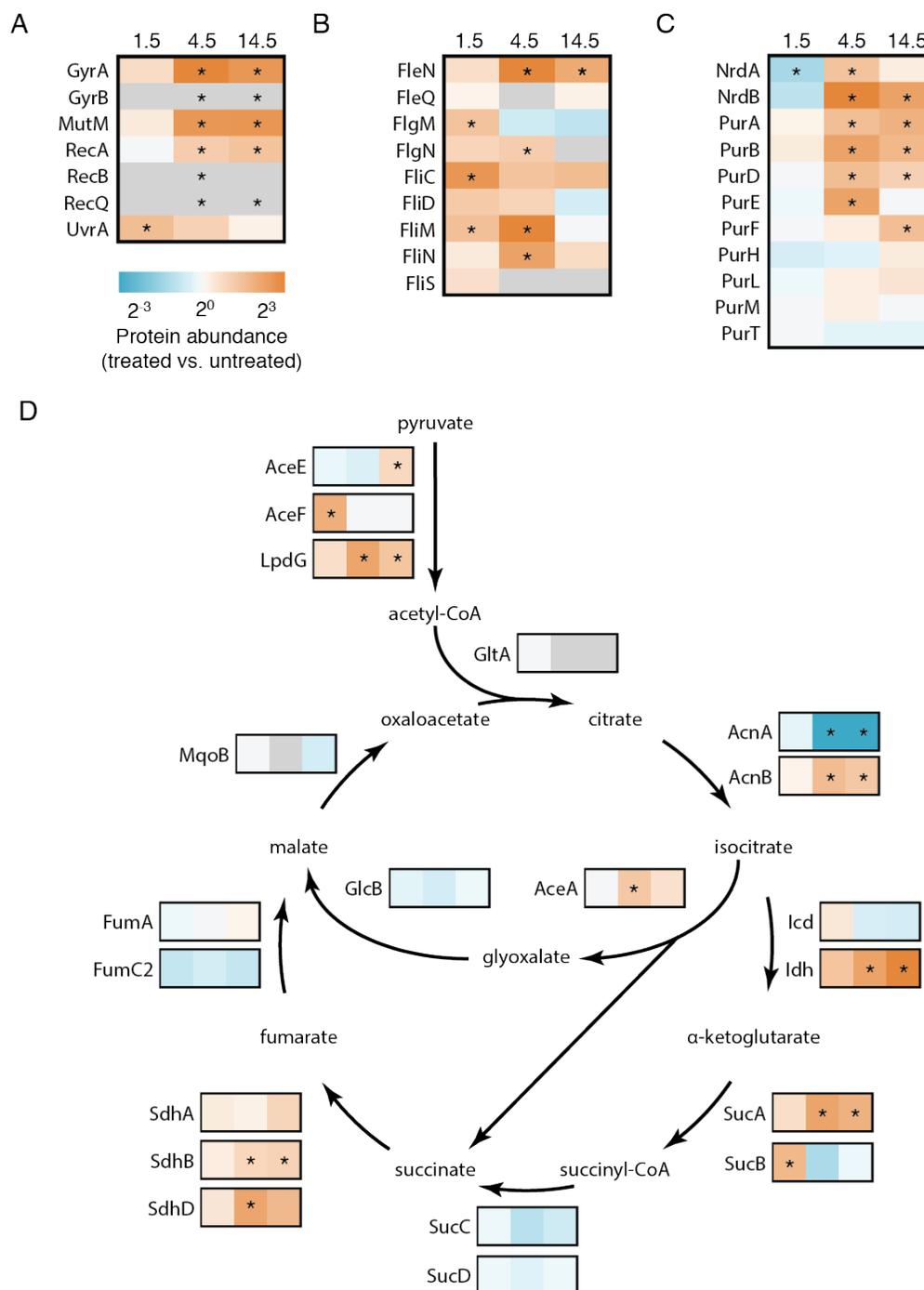
(GyrA and GyrB), the direct target of ciprofloxacin; and proteins involved in sensing (RecA, UvrA) and repairing (RecB, RecQ, MutM) various types of DNA damage.

The differences between early and late responses are best exemplified by the contrasting behavior of proteins involved in flagella synthesis and purine metabolism. Protein components of flagella and flagellar regulation were significantly upregulated throughout the course of ciprofloxacin treatment including the immediate upregulation of FliC, FliM, and FlgM (Figure 4.4B). Flagella are known to con-

tribute to the development of biofilm microstructures. Specifically, strains deficient in flagellar synthesis are unable to colonize the upper regions of flow cell biofilms [51]. At the 14.5 h time point, only FleN remains upregulated.

In contrast, many proteins involved in purine metabolism are upregulated only at the 4.5 and 14.5 h time points (Figure 4.4C). Of the proteins in the pathway for *de novo* synthesis of inosine monophosphate (IMP) that we identified (PurBDEFHLMT), four of eight were significantly upregulated in at least one of the later time points. PurA, required for IMP's conversion to the ribonucleotide adenosine monophosphate, was likewise upregulated. The ribonucleotide reductase complex (NrdA and NrdB) that generates deoxyribonucleotides from their ribonucleotide precursors was also upregulated.

Finally, some of the largest changes we observed were the up- and downregulation of proteins involved in central metabolism. These include many components of the citrate (tricarboxylic acid, TCA) (Figure 4.4D) and methylcitrate cycles (Figure 4.S4C). We found slight upregulation of the isocitrate lyase, AceE, which catalyzes the first step in the glyoxylate shunt, a metabolic pathway previously shown to be important for *Burkholderia cepacia* persistence to antibiotics *in vitro* [52] and the persistence of *Mycobacterium tuberculosis* in mice [53]. Of particular interest is the differential behavior of the aconitases and isocitrate dehydrogenases (ICD). The *P. aeruginosa* genome encodes two of each protein class (aconitases AcnA and AcnB and ICD's Icd and Idh). We found AcnA to be significantly downregulated and AcnB significantly upregulated in the later response conditions. Similarly, expression of Icd, the monomeric ICD was unchanged, while the dimeric Idh was the protein most upregulated at the 14.5 h time point.



**Figure 4.4: Dynamic cellular responses to ciprofloxacin.** Heatmaps indicating changes in abundance at each time point as compared to the untreated control for proteins involved in (A) DNA damage and repair, (B) flagella synthesis, (C) purine metabolism, and (D) the TCA cycle and glyoxalate shunt. The color scale for abundance ratios is shown under panel (A). Gray boxes represent proteins that were not quantified. Asterisks indicate abundance ratios that were significantly different from 1 (FDR adjusted  $p$ -value  $< 0.05$ ) or proteins that were identified in the treated sample but not in the untreated control (gray with an asterisk).

#### 4.4 Discussion

Here we introduce an adaptation to the BONCAT method that allows for selective proteomic analysis of a phenotypic subpopulation in genotypically identical bacterial cultures. Our approach for targeting protein labeling to a subset of cells is generally applicable toward the study of heterogeneous systems (e.g., planktonic persister cells, or *in vivo* infections), and we show that the use of endogenous regulatory elements allows precise control over the targeted phenotype. For our studies of biofilms, we found that the *rpoS* promoter allowed for targeting of biofilm interiors, but we note that the approach is general and that, in principle, any genetic regulatory element can be used to target other subpopulations within biofilms or other systems.

In the context of biofilm antibiotic tolerance, we show that targeted labeling with AnI allows for detection of protein synthesis after extended incubation with a supra-MIC of ciprofloxacin, and through enrichment and LC-MS/MS analysis of labeled proteins, we detect and differentiate the immediate response and the long-term adaptation to antibiotic stress. We find congruence with previously reported responses to ciprofloxacin by planktonic *P. aeruginosa*, including upregulation of key components of the SOS response to DNA damage and the two subunits of DNA gyrase, a direct target of the drug. Our proteomic dataset encompasses a variety of other responses, including the immediate upregulation of proteins for flagellar motility, and the delayed upregulation of purine synthesis proteins.

Some of the most striking changes were the rearrangement of proteins involved in the central carbon metabolism. The 2-methylisocitrate lyase PrpB is one of the few proteins upregulated by ciprofloxacin at all three time points, and the dimeric ICD Idh shows the highest fold upregulation following 14.5 h of treatment. We also find upregulation of various other proteins involved in the citrate and methylcitrate cycles, and interestingly, differential expression of proteins that can perform the same enzymatic activity: the ICDs Idh (upregulated) and Icd (unchanged), and the aconitases AcnA (downregulated) and AcnB (upregulated). To our knowledge, broad adjustments to central carbon metabolism in *P. aeruginosa* biofilms have not been reported as a mechanism for antibiotic tolerance or adaptation. However, changes in TCA cycle enzymes and intermediate metabolites have been discussed for their role in maintaining redox conditions in response to reactive oxygen species (ROS). Although linked to the somewhat controversial claim that ROS is the primary and general mechanism for antibiotic-induced killing, genetic experiments have shown that the deletion of some but not all TCA cycle enzymes (aconitase and ICD

genes in particular) confer increased resistance to *E. coli* treated with antibiotics, presumably through the adjustment of the intracellular NAD<sup>+</sup>/NADH balance [54].

The role of central carbon metabolism in the response to antibiotic stress has been studied more thoroughly in *Mycobacterium tuberculosis* (Mtb), another organism that chronically infects the human lung. Measurements of TCA cycle metabolites showed that treatment of Mtb with a variety of antibiotics leads to the increased abundance of pyruvate, succinate, and fumarate [55]. Though we didn't perform metabolomic measurements, the proteomic changes we observe are consistent with changes to these metabolites. Upregulation of components of the pyruvate dehydrogenase (AceEF and LpdG) and succinate dehydrogenase (SdhABC) complexes are consistent with increased flux from pyruvate, and from succinate to fumarate respectively. The role of the methylcitrate cycle in infection has also been investigated in Mtb, where it is required for growth within macrophages (an environment with high levels of ROS) [56]. From our dataset, PrpB (upregulated) catalyzes the generation of methylcitrate, and in *E. coli* AcnB (upregulated), but not AcnA (downregulated), has activity toward methylcitrate in addition to its canonical role as a citrate aconitase [57]. We caution that even core central metabolism can differ substantially among bacteria, but that there is precedent for a critical role of central carbon metabolism in general stress responses.

Our measured proteomic responses of a *P. aeruginosa* biofilm subpopulation lead to interesting questions about the roles of these proteins in ciprofloxacin tolerance. To what extent are changes to the central carbon proteome important for their effects on general redox balance vs. their effects on intermediate metabolite concentrations? What is the physiological benefit of the differential expression of redundant enzymes like the aconitases and ICDs? Dissecting the roles of these proteins and others in the data set through focused genetic experiments and metabolomics should provide better understanding of these phenomena. In addition, our temporal analysis suggests that there are important differences between immediate responses and longer term physiological adaptations. Finally, this method for selective proteomics is readily applicable for the study of subpopulation-specific responses to other stresses (e.g., antibiotics with other mechanisms of action).

#### 4.5 Experimental Procedures

**Strain construction.** All strains are listed in Table 4.1. We used standard cloning procedures. Enzymes were purchased from New England Biolabs. For chromosomal integration into the *Tn7* site, pUC18T mini-Tn7T [58] was modified with the desired expression cassette, followed by tetraparental conjugation to the PA14 host strain. Genomic DNA was prepared using the GenElute Bacterial DNA kit (Sigma-Aldrich). The 1 kb regions upstream of *rpoS* and *algP* were amplified from *P. aeruginosa* genomic DNA. GFP-expressing cassettes contain the gene for *gfpmut3b*, cloned from pBK-mini-Tn7-gfp2 [59]. The gene encoding the *E. coli* mutant methionyl-tRNA synthetase was cloned from plasmid pJTN1 [28]. A shuttle vector allowing for arabinose-inducible expression of NLL-MetRS was created by cloning the gene from pJTN1 into pBAD18 [60] and then ligation of the fragment containing *araC* and *P<sub>ara:nll</sub>* into pUCP24 [61] to generate pBADP-NLL. *P. aeruginosa* was transformed by electroporation.

**Media and growth conditions.** Planktonic cultures were grown at 37 °C with shaking. Liquid media were LB (5 g yeast extract, 10 g tryptone, 10 g NaCl per liter), or FAB with 0.05 g/L glucose (for biofilms) or 5 g/L glucose (for planktonic cultures) in place of citrate for biofilms [63]. For confocal imaging, biofilms were grown in flow cells (1x4x40 mm) (Stovall) as previously described [64], but without bubble traps. Biofilms were grown at 37 °C with a constant flow rate of 0.03 mL/min. For proteomic analyses, biofilms were grown in silicon tubing (10 mm interior diameter, 20 cm long) (McMaster-Carr) at 37 °C with a constant flow rate of 0.5 mL/min, as previously described [65]. Loosely adherent biofilm cells were extracted by collecting media within each tube and flushing with 0.9% NaCl. Tubing was cut into 1 cm pieces and vortexed in 0.9% NaCl to remove all cells.

**BONCAT labeling and enrichment.** For planktonic labeling experiments, strains were diluted from overnight cultures 1:100 into FAB medium with 5 g/L glucose. At each time point, labeling was initiated by the addition of 1 mM Anl (Iris-Biotech). The strain containing pBADP-NLL-MetRS was grown in the presence of 50 µg/mL gentamicin and treated with 1 mM Anl and 20 mM arabinose. For all, after 15 min of incubation with Anl at 37 °C with shaking, cells were pelleted at 4 °C, washed once with ice cold 0.9% NaCl and frozen at -80 °C. For biofilm experiments, flow was stopped and tubing was clamped. FAB medium with 0.05 g/L glucose and 1 mM Anl was injected by syringe and biofilms were incubated for 1.5 h at 37 °C. For

Table 4.1: **Chapter 4: Strains and plasmids.** Strains and plasmids used in this study. Plasmids are stored as *E. coli* strains carrying the plasmid, and requests should be for the *E. coli* strain.

<b><i>Pseudomonas aeruginosa</i> Strains</b>		
Name	Genotype	Source
DKN263	<i>P. aeruginosa</i> UCBPP-PA14	
BMB1	UCBPP-PA14 <i>P<sub>rpoS</sub>:gfp</i>	attTn7::mini-Tn7T- <i>Gm<sup>R</sup></i> This Study
BMB2	UCBPP-PA14 <i>P<sub>irc</sub>:gfp</i>	attTn7::mini-Tn7T- <i>Gm<sup>R</sup></i> This Study
BMB3	UCBPP-PA14 <i>P<sub>rpoS</sub>:nll-mc</i>	attTn7::mini-Tn7T- <i>Gm<sup>R</sup></i> This Study
BMB4	UCBPP-PA14 <i>P<sub>irc</sub>:nll-mc</i>	attTn7::mini-Tn7T- <i>Gm<sup>R</sup></i> This Study
BMB5	UCBPP-PA14 <i>P<sub>algP</sub>:gfp</i>	attTn7::mini-Tn7T- <i>Gm<sup>R</sup></i> This Study
<b><i>Escherichia coli</i> Strains</b>		
Name	Genotype	Source
DKN1299	SM10, pTNS1	[58]
DKN1299	HB101 pRK2013	[58]
BMB6	Mach1 pUC18T-mini-Tn7T- <i>Gm<sup>R</sup></i> <i>P<sub>rpoS</sub>:gfp</i>	This Study
BMB7	Mach1 pUC18T-mini-Tn7T- <i>Gm<sup>R</sup></i> <i>P<sub>irc</sub>:gfp</i>	This Study
BMB8	Mach1 pUC18T-mini-Tn7T- <i>Gm<sup>R</sup></i> <i>P<sub>rpoS</sub>:nll-mc</i>	This Study
BMB9	Mach1 pUC18T-mini-Tn7T- <i>Gm<sup>R</sup></i> <i>P<sub>irc</sub>:nll-mc</i>	This Study
BMB10	Mach1 pUC18T-mini-Tn7T- <i>Gm<sup>R</sup></i> <i>P<sub>algP</sub>:gfp</i>	This Study
BMB11	Mach1 pUCP18	[62]
BMB12	Mach1 pBAD18-NLL-MetRS	This Study
BMB13	Mach1 pBADP-NLL-MetRS	This Study

proteome analysis, cells were collected from tubing as described above, pelleted, and frozen at -80 °C.

All samples were lysed by resuspension in lysis buffer (100 mM Tris-HCl, pH 8, 4% SDS). Lysates were sonicated with a microtip probe for 30 s at setting 20% (Qsonica). For fluorescence detection of AnI-labeled proteins, lysates were reacted with 5 μM TAMRA-alkyne (Click Chemistry Tools), 100 μM CuSO<sub>4</sub>, 500 μM tris(3-hydroxypropyltriazolylmethyl)amine (THPTA), 5 mM aminoguanidine hydrochloride, and 5 mM sodium ascorbate for 15 min at room temperature [66]; precipitated with water, methanol, and chloroform; and washed twice with methanol. Reacted lysates were separated via SDS-PAGE and imaged on a Typhoon gel imager (GE Healthcare). Gels were stained with Colloidal Blue (Life Technologies) or Instant-

Blue (Expedeon) Coomassie stains to verify equal protein loading.

For all enrichments, cysteines were reduced by addition of 10 mM dithiothreitol (DTT) for 20 min at room temperature and alkylated by addition of 100 mM chloroacetamide for 30 min in the dark. For the comparison between *P<sub>rpoS</sub>:nll-mc* and *P<sub>irc</sub>:nll-mc* biofilms, 0.5 mg of protein lysate per sample were reacted with 12  $\mu$ M DBCO-sulfo-biotin (Click Chemistry Tools) in 0.5 mL PBS for 15 min at room temperature. Proteins were precipitated with acetone at -20 °C and resuspended in PBS, 0.3% SDS. Streptavidin UltraLink Resin (Pierce Biotechnology) was washed twice with PBS, added to biotinylated lysates, and incubated overnight at 4 °C. Resin was transferred to microfuge spin columns (Pierce Biotechnology) and washed twice with 1% SDS in PBS and once with 0.1% SDS in PBS. Proteins were eluted by incubation with 1 mM biotin at 65 °C for 20 min. Eluted proteins were separated via SDS-PAGE (4-12% Bis-Tris gradient gel, Thermo Fisher) and subjected to GeLCMS.

For the comparison between ciprofloxacin treated samples, reduced and alkylated lysates (0.5 mg per sample) in 0.5 mL PBS were reacted with 50  $\mu$ L of DBCO-agarose bead 50% slurry (Click Chemistry Tools) for 2.5 h at room temperature. Beads were washed extensively in gravity flow columns (Bio-Rad) with 40 mL each of PBS, 0.8% (w/v) SDS; 8 M urea; and 20% (v/v) acetonitrile in water. Beads were resuspended in 50 mM ammonium bicarbonate (AB) for on-bead tryptic digestion (see LC-MS/MS section for details).

**Imaging flow cell biofilms.** All treatments were applied via syringe to flow cell biofilms. For GFP imaging, flow was stopped and live biofilms were incubated with 0.05  $\mu$ M SYTO 62 (ThermoFisher) for 30 min at 37 °C. To visualize AnI incorporation, biofilms were fixed by incubation with 3.7% formaldehyde for 30 min and permeabilized by incubation with 70% ethanol for 5 min on ice. Fixed biofilms were washed with 0.9% NaCl, incubated with 100 mM chloroacetamide in the dark for 30 min, and treated with 25  $\mu$ M DBCO-TAMRA (Click Chemistry Tools) in PBS for 30 min. Biofilms were washed extensively to remove excess dye and counter stained with 0.05  $\mu$ M STYO 9 (ThermoFisher).

**LC-MS/MS** For GeLCMS, gel lanes were cut into 8 pieces each and destained by alternating washes with 50  $\mu$ L each of 50 mM ammonium bicarbonate (AB) and 1:1 50 mM AB:acetonitrile. Proteins were reduced by incubation with 6.7 mM dithiothreitol (DTT) in 50  $\mu$ L 50 mM AB at 50 °C for 30 min and alkylated by incubation with 37 mM iodoacetamide in 50  $\mu$ L 50 mM AB at room temperature for

20 min. Gel pieces were washed with 50  $\mu$ L each of 100 mM AB and acetonitrile. Proteins were digested with 300 ng endoproteinase LysC in 50  $\mu$ L 100 mM Tris-HCl at 37 °C for 18 h. Peptides were extracted by sequential washing with: 50  $\mu$ L each of 1% formic acid/2% acetonitrile, 1:1 acetonitrile:water, and 1% formic acid in acetonitrile. Peptides were desalted with C18 ZipTips (EMD Millipore).

For on-bead digestion following enrichment, agarose beads were incubated with 100 ng trypsin in 9:1 AB:acetonitrile for 18 h at 37 °C. Supernatant was collected and beads were washed twice with 20% acetonitrile to extract all peptides. Peptides were dried, passed through HiPPR spin columns (ThermoFisher) to remove any residual SDS, and desalted with C18 ZipTips.

Liquid chromatography-mass spectrometry experiments were essentially carried out as previously described [67]. The *rpoS* vs. *trc* experiments were performed on a nanoflow LC system, EASY-nLC 1000 coupled to a hybrid linear ion trap Orbitrap Classic mass spectrometer (Thermo Scientific) equipped with a nanoelectrospray ion source (Thermo Scientific) with the following modifications: For the EASY-nLC II system, solvent A consisted of 97.8% H<sub>2</sub>O, 2% ACN, and 0.2% formic acid and solvent B consisted of 19.8% H<sub>2</sub>O, 80% ACN, and 0.2% formic acid. For the LC-MS/MS experiments, digested peptides were directly loaded at a flow rate of 500 nL/min onto a 16-cm analytical HPLC column (75  $\mu$ m ID) packed in-house with ReproSil-Pur C<sub>18</sub>AQ 3  $\mu$ m resin (120 Å pore size, Dr. Maisch, Ammerbuch, Germany). The column was enclosed in a column heater operating at 30 °C. After 30 min of loading time, the peptides were separated with a 50 min gradient at a flow rate of 350 nL/min. The gradient was as follows: 0–30% B (50 min), and 100% B (10 min). The Orbitrap was operated in data-dependent acquisition mode to automatically alternate between a full scan ( $m/z$ =400–1600) in the Orbitrap and subsequent 10 CID MS/MS scans in the linear ion trap. CID was performed with helium as collision gas at a normalized collision energy of 35% and 30 ms of activation time. Ciprofloxacin experiments were performed on a hybrid ion trap-Orbitrap Elite mass spectrometer (Thermo Scientific).

Raw files were searched using MaxQuant [68] against the *P. aeruginosa* PA14 UniProt entries (5,886 sequences) and a contaminant database (246 sequences). Trypsin was specified as the digestion enzyme with up to two missed cleavages. Carbamidomethylation of cysteine was set as a fixed modification and protein N-terminal acetylation and methionine oxidation were variable modifications. Protein abundances were estimated with MaxLFQ [69], and for each experiment, peptides

were matched between runs. LFQ values were normalized and used to calculate abundance ratios between samples and to estimate variance using the limma package in R [70]. P-values were adjusted for false discovery by the Benjamini-Hochberg procedure [71].

**Software analysis and data presentation.** This section describes software packages that were not mentioned above. Data processing and statistical analysis were performed with Python version 2.7.9 with NumPy version 1.9.2, SciPy version 0.15.1, and Pandas version 0.16.1. Data were plotted with Matplotlib version 1.5.1 [72] and Seaborn version 0.7.0. Microscopy and gel images were analyzed with ImageJ 64-bit version 2.0.0 [73]. Figures were assembled in Adobe Illustrator CS5.

## 4.6 Supplementary Figures

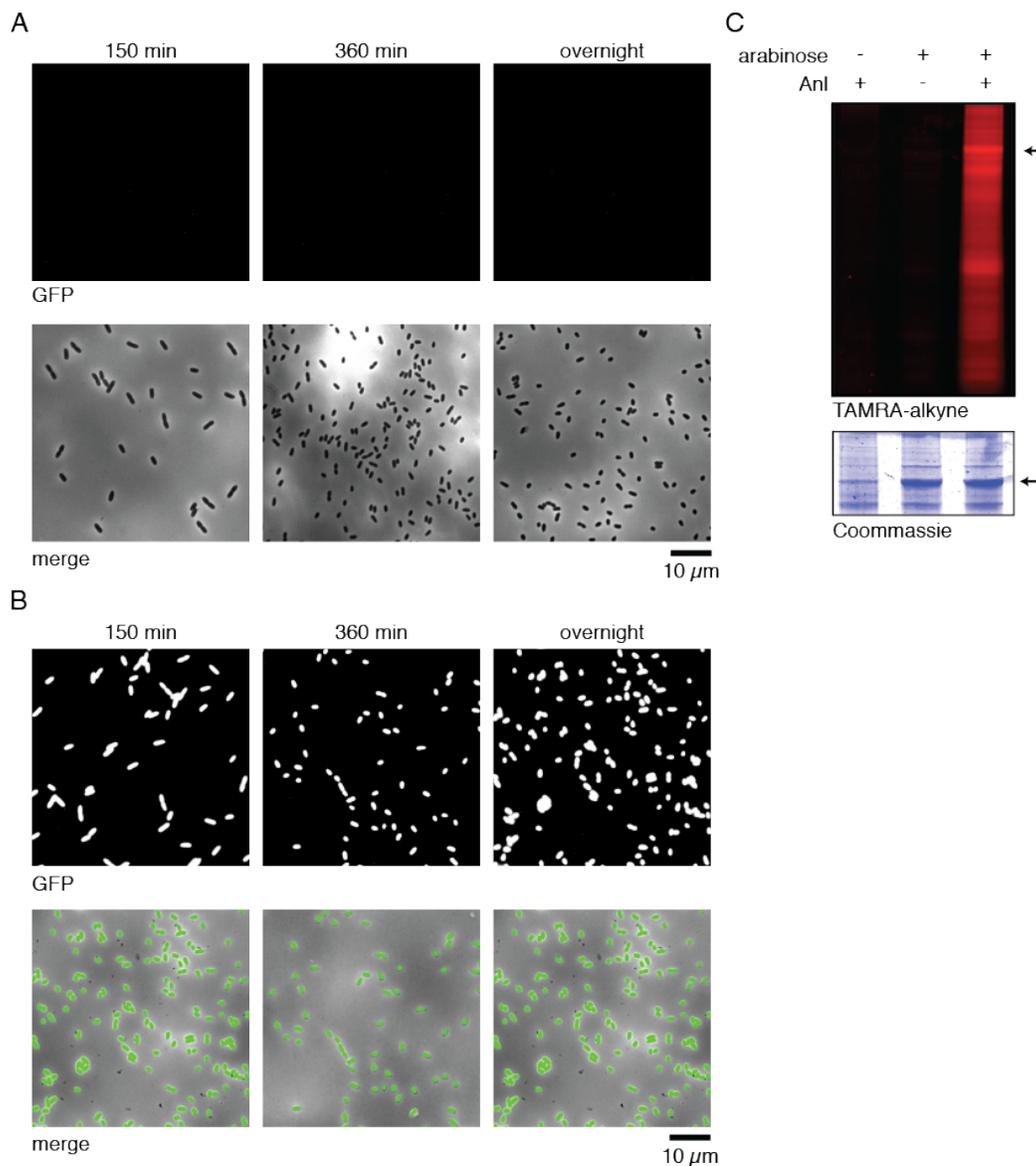


Figure 4.S1: **Promoter-controlled expression.** (A, B) Fluorescence imaging of wild type (A) and *P<sub>trc</sub>:gfp* (B). GFP fluorescence (top) and GFP-bright field merge (bottom). (C) Controlled proteome labeling with inducible expression of NLL-MetRS from the *ara* promoter. SDS-PAGE gel imaged for TAMRA fluorescence (top) and stained with Coomassie (bottom). Arrows indicate the NLL-MetRS protein.

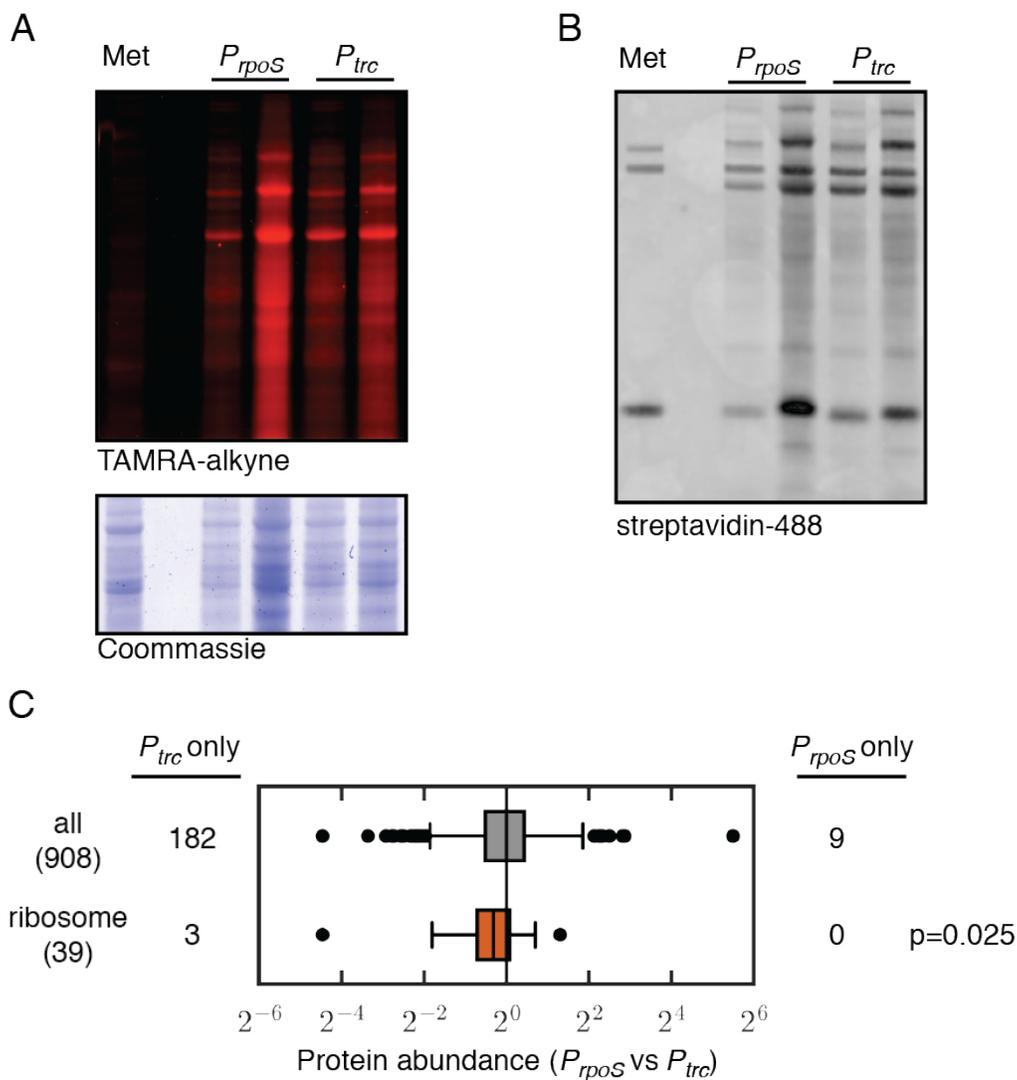


Figure 4.S2: **Enrichment from biofilms of  $P_{rpoS}$  and  $P_{trc}$  strains.** (A) SDS-PAGE gel showing AnI incorporation for each biofilm replicate. (B) Western blot with streptavidin-Alexa Fluor 488 of lysates reacted with DBCO-sulfo-biotin for affinity enrichment. Three naturally biotinylated proteins are visible in the methionine-treated negative control. (C) Distribution of protein abundance ratios for all proteins (gray) and for ribosomal proteins (orange). P-value was calculated by bootstrapped subsampling from the set of all abundance ratios ( $n = 1000$ ).

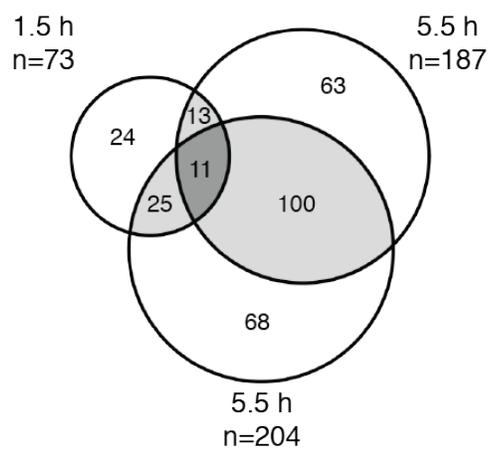


Figure 4.S3: **Shared and unique proteomic hits.** Overlap of protein hits (both down- and upregulated) at each time point throughout ciprofloxacin treatment.

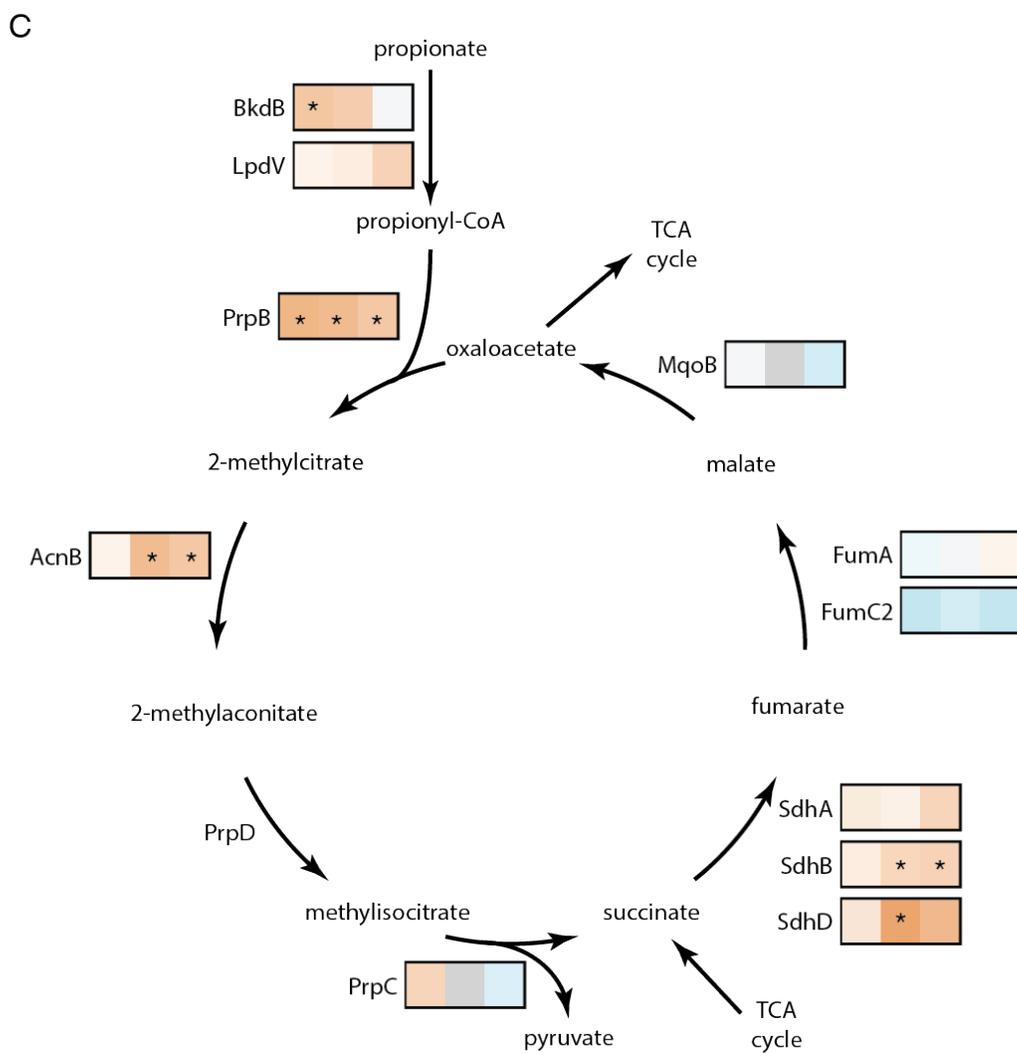
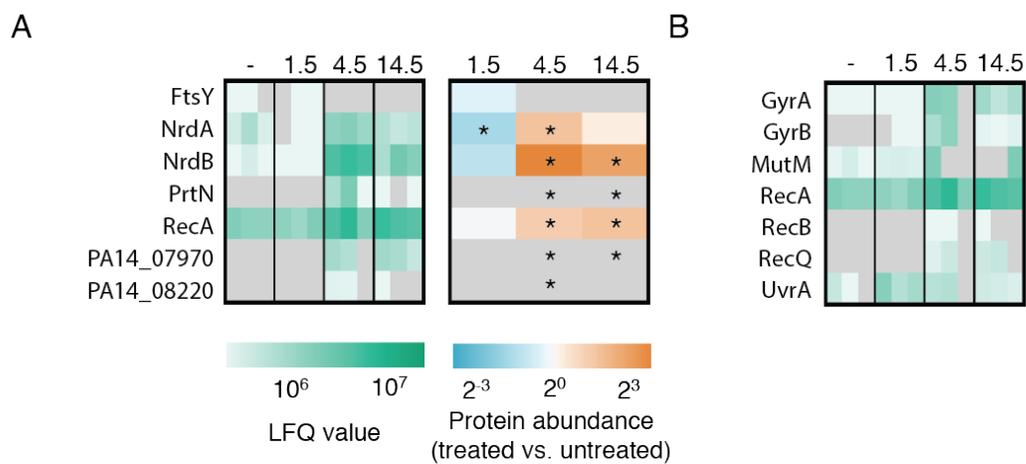


Figure 4.S4: **Other responses to ciprofloxacin.** (A) LFQ values and protein abundance ratios compared to the untreated sample for genes previously reported to be upregulated in response to ciprofloxacin treatment by Cirz, et al [10]; Brazas and Hancock [48]; and Linares, et al [49]. (B) LFQ values for proteins involved in the SOS response. (C) Abundance ratios for proteins involved in the isocitrate cycle. Asterisks indicate abundance ratios that were significantly different from 1 (FDR adjusted p-value < 0.05) or proteins that were identified in the treated sample but not in the control (gray with an asterisk).

#### 4.7 Supplementary Datasets

**Dataset 4.1. Proteomic results for *trc* and *rpoS* biofilm enrichments.** All proteins identified by LC-MS/MS from the BONCAT-enriched samples are listed. Column 1 gives the Uniprot ID. Columns 2-5 give the LFQ intensity values for each sample. Column 6 gives the  $\log_2$  abundance ratio (*rpoS* vs. *trc*). Column 7 gives the FDR-adjusted p-value. Columns 8-17 give the following gene identification information: PA14 locus tag, gene name if available, protein description, list of PseudoCAP annotations, GenBank gene accession number, GenBank GI number, gene name from the KEGG database, KEGG orthology number, KEGG enzyme number, PAO1 locus tag.

**Dataset 4.2. Proteomic results for ciprofloxacin treated biofilms.** All proteins identified by LC-MS/MS from the BONCAT-enriched ciprofloxacin experiment are listed. Column 1 gives the Uniprot ID. Columns 2-13 give the LFQ intensity values for each sample. Columns 14, 16, and 18 give the  $\log_2$  abundance ratios (treated vs. untreated) for each time point. Columns 15, 17, and 19 give the FDR-adjusted p-values for each ratio. Columns 20-29 give the following gene identification information: PA14 locus tag, gene name if available, protein description, list of PseudoCAP annotations, GenBank gene accession number, GenBank GI number, gene name from the KEGG database, KEGG orthology number, KEGG enzyme number, PAO1 locus tag.

## References

- (1) Fux, C. A.; Costerton, J. W.; Stewart, P. S.; Stoodley, P. *Trends Microbiol.* **2005**, *13*, 34–40.
- (2) Levin, B. R.; Rozen, D. E. *Nat. Rev. Microbiol.* **2006**, *4*, 556–562.
- (3) Oliver, A.; Cantón, R.; Campo, P.; Baquero, F.; Blázquez, J. *Science* **2000**, *288*, 1251–1254.
- (4) Høiby, N.; Bjarnsholt, T.; Givskov, M.; Molin, S.; Ciofu, O. *Int. J. Antimicrob. Ag.* **2010**, *35*, 322–332.
- (5) Stewart, P. S.; Franklin, M. J. *Nat. Rev. Microbiol.* **2008**, *6*, 199–210.
- (6) Pamp, S. J.; Gjermansen, M.; Johansen, H. K.; Tolker-Nielsen, T. *Mol. Microbiol.* **2008**, *68*, 223–240.
- (7) Walters, M. C.; Roe, F.; Bugnicourt, A.; Franklin, M. J.; Stewart, P. S. *Antimicrob. Agents Ch.* **2003**, *47*, 317–323.
- (8) Bagge, N.; Hentzer, M.; Andersen, J. B.; Ciofu, O.; Givskov, M.; Høiby, N. *Antimicrob. Agents Ch.* **2004**, *48*, 1168–1174.
- (9) Wu, X.; Held, K.; Zheng, C.; Staudinger, B. J.; Chavez, J. D.; Weisbrod, C. R.; Eng, J. K.; Singh, P. K.; Manoil, C.; Bruce, J. E. *Mol. Cell Proteomics* **2015**, *14*, 2126–2137.
- (10) Cirz, R. T.; O’Neill, B. M.; Hammond, J. A.; Head, S. R.; Romesberg, F. E. *J. Bacteriol.* **2006**, *188*, 7101–7110.
- (11) Park, A. J.; Krieger, J. R.; Khursigara, C. M. *FEMS Microbiol. Rev.* **2016**, DOI: 10.1093/femsre/fuv051.
- (12) Lenz, A. P.; Williamson, K. S.; Pitts, B.; Stewart, P. S.; Franklin, M. J. *Appl. Environ. Microb.* **2008**, *74*, 4463–4471.
- (13) Williamson, K. S.; Richards, L. A.; Perez-Osorio, A. C.; Pitts, B.; McInnerney, K.; Stewart, P. S.; Franklin, M. J. *J. Bacteriol.* **2012**, *194*, 2062–2073.
- (14) Khemiri, A.; Jouenne, T.; Cosette, P. *Med. Microbiol. Immunol.* **2016**, *205*, 1–19.
- (15) Chua, S. L.; Yam, J. K. H.; Hao, P.; Adav, S. S.; Salido, M. M.; Liu, Y.; Givskov, M.; Sze, S. K.; Tolker-Nielsen, T.; Yang, L. *Nat. Commun.* **2016**, *7*, 10750.
- (16) Schwanhäusser, B.; Gossen, M.; Dittmar, G.; Selbach, M. *Proteomics* **2009**, *9*, 205–209.
- (17) Dieterich, D. C.; Link, A. J.; Graumann, J.; Tirrell, D. A.; Schuman, E. M. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 9482–9487.

- (18) Dieterich, D. C.; Lee, J. J.; Link, A. J.; Graumann, J.; Tirrell, D. A.; Schuman, E. M. *Nat. Protoc.* **2007**, *2*, 532–540.
- (19) Bagert, J. D.; Xie, Y. J.; Sweredoski, M. J.; Qi, Y.; Hess, S.; Schuman, E. M.; Tirrell, D. A. *Mol. Cell Proteomics* **2014**, *13*, 1352–1358.
- (20) Howden, A. J. M.; Geoghegan, V.; Katsch, K.; Efstathiou, G.; Bhushan, B.; Boutureira, O.; Thomas, B.; Trudgian, D. C.; Kessler, B. M.; Dieterich, D. C.; Davis, B. G.; Acuto, O. *Nat. Methods* **2013**, *10*, 343–346.
- (21) Bagert, J. D.; van Kessel, J. C.; Sweredoski, M. J.; Feng, L.; Hess, S.; Bassler, B. L.; Tirrell, D. A. *Chem. Sci.* **2016**, *7*, 1797–1806.
- (22) Feng, L.; Rutherford, S. T.; Papenfort, K.; Bagert, J. D.; van Kessel, J. C.; Tirrell, D. A.; Wingreen, N. S.; Bassler, B. L. *Cell* **2015**, *160*, 228–240.
- (23) Kramer, G.; Sprenger, R. R.; Back, J.; Dekker, H. L.; Nessen, M. A.; van Maarseveen, J. H.; de Koning, L. J.; Hellingwerf, K. J.; de Jong, L.; de Koster, C. G. *Mol. Cell Proteomics* **2009**, *8*, 1599–1611.
- (24) Sinai, L.; Rosenberg, A.; Smith, Y.; Segev, E.; Ben-Yehuda, S. *Mol. Cell* **2015**, *57*, 695–707.
- (25) Tanrikulu, I. C.; Schmitt, E.; Mechulam, Y.; Goddard, W. A.; Tirrell, D. A. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 15285–15290.
- (26) Grammel, M.; Zhang, M. M.; Hang, H. C. *Angew. Chem. Int. Ed.* **2010**, *49*, 5970–5974.
- (27) Grammel, M.; Dossa, P. D.; Taylor-Salmon, E.; Hang, H. C. *Chem. Commun.* **2012**, *48*, 1473–1474.
- (28) Ngo, J. T.; Champion, J. A.; Mahdavi, A.; Tanrikulu, I. C.; Beatty, K. E.; Connor, R. E.; Yoo, T. H.; Dieterich, D. C.; Schuman, E. M.; Tirrell, D. A. *Nat. Chem. Biol.* **2009**, *5*, 715–717.
- (29) Ngo, J. T.; Babin, B. M.; Champion, J. A.; Schuman, E. M.; Tirrell, D. A. *ACS Chem. Biol.* **2012**, *7*, 1326–1330.
- (30) Xu, K. D.; Franklin, M. J.; Park, C. H.; McFeters, G. A.; Stewart, P. S. *FEMS Microbiol. Lett.* **2001**, *199*, 67–71.
- (31) Amann, E.; Ochs, B.; Abel, K. J. *Gene* **1988**, *69*, 301–315.
- (32) Winsor, G. L.; Lam, D. K. W.; Fleming, L.; Lo, R.; Whiteside, M. D.; Yu, N. Y.; Hancock, R. E. W.; Brinkman, F. S. L. *Nucleic Acids Res.* **2011**, *39*, D596–D600.
- (33) Zimmermann, A.; Reimann, C.; Galimand, M.; Haas, D. *Mol. Microbiol.* **1991**, *5*, 1483–1490.
- (34) Latifi, A.; Foglino, M.; Tanaka, K.; Williams, P.; Lazdunski, A. *Mol. Microbiol.* **1996**, *21*, 1137–1146.

- (35) Dasgupta, N.; Wolfgang, M. C.; Goodman, A. L.; Arora, S. K.; Jyot, J.; Lory, S.; Ramphal, R. *Mol. Microbiol.* **2003**, *50*, 809–824.
- (36) Suh, S.-J.; Runyen-Janecky, L. J.; Maleniak, T. C.; Hager, P.; MacGregor, C. H.; Zielinski-Mozny, N. A.; Phibbs, P. V.; West, S. E. H. *Microbiology* **2002**, *148*, 1561–1569.
- (37) Lee, S.; Hinz, A.; Bauerle, E.; Angermeyer, A.; Juhaszova, K.; Kaneko, Y.; Singh, P. K.; Manoil, C. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 14570–14575.
- (38) Werner, E.; Roe, F.; Bugnicourt, A.; Franklin, M. J.; Heydorn, A.; Molin, S.; Pitts, B.; Stewart, P. S. *Appl. Environ. Microb.* **2004**, *70*, 6188–6196.
- (39) Cavallari, J. F.; Lamers, R. P.; Scheurwater, E. M.; Matos, A. L.; Burrows, L. L. *Antimicrob. Agents Ch.* **2013**, *57*, 3078–3084.
- (40) Nair, S.; Finkel, S. E. *J. Bacteriol.* **2004**, *186*, 4192–4198.
- (41) Calhoun, L. N.; Kwon, Y. M. *J. Appl. Microbiol.* **2011**, *110*, 375–386.
- (42) Almirón, M.; Link, A. J.; Furlong, D.; Kolter, R. *Genes Dev.* **1992**, *6*, 2646–2654.
- (43) Ramsey, D. M.; Wozniak, D. J. *Mol. Microbiol.* **2005**, *56*, 309–322.
- (44) Stapper, A. P.; Narasimhan, G.; Ohman, D. E.; Barakat, J.; Hentzer, M.; Molin, S.; Kharazmi, A.; Høiby, N.; Mathee, K. *J. Med. Microbiol.* **2004**, *53*, 679–690.
- (45) Battesti, A.; Majdalani, N.; Gottesman, S. *Annu. Rev. Microbiol.* **2011**, *65*, 189–213.
- (46) Oliver, J. D. *FEMS Microbiol. Rev.* **2010**, *34*, 415–425.
- (47) Ayrapetyan, M.; Williams, T. C.; Oliver, J. D. *Trends Microbiol.* **2015**, *23*, 7–13.
- (48) Brazas, M. D.; Hancock, R. E. W. *Antimicrob. Agents Ch.* **2005**, *49*, 3222–3227.
- (49) Linares, J. F.; Gustafsson, I.; Baquero, F.; Martinez, J. L. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 19484–19489.
- (50) Schlacher, K.; Goodman, M. F. *Nat. Rev. Mol. Cell Biol.* **2007**, *8*, 587–594.
- (51) Barken, K. B.; Pamp, S. J.; Yang, L.; Gjermansen, M.; Bertrand, J. J.; Klausen, M.; Givskov, M.; Whitchurch, C. B.; Engel, J. N.; Tolker-Nielsen, T. *Environ. Microbiol.* **2008**, *10*, 2331–2343.
- (52) Van Acker, H.; Sass, A.; Bazzini, S.; De Roy, K.; Udine, C.; Messiaen, T.; Riccardi, G.; Boon, N.; Nelis, H. J.; Mahenthiralingam, E.; Coenye, T. *PLoS ONE* **2013**, *8*, e58943.

- (53) McKinney, J. D.; Höner zu Bentrup, K.; Muñoz-Elías, E. J.; Miczak, A.; Chen, B.; Chan, W. T.; Swenson, D.; Sacchettini, J. C.; Jacobs, W. R.; Russell, D. G. *Nature* **2000**, *406*, 735–738.
- (54) Kohanski, M. A.; Dwyer, D. J.; Hayete, B.; Lawrence, C. A.; Collins, J. J. *Cell* **2007**, *130*, 797–810.
- (55) Nandakumar, M.; Nathan, C.; Rhee, K. Y. *Nat. Commun.* **2014**, *5*, 4306.
- (56) Muñoz-Elías, E. J.; Upton, A. M.; Cherian, J.; McKinney, J. D. *Mol. Microbiol.* **2006**, *60*, 1109–1122.
- (57) Brock, M.; Maerker, C.; Schütz, A.; Völker, U.; Buckel, W. *EMBO J. Biochem.* **2002**, *269*, 6184–6194.
- (58) Choi, K.-H.; Schweizer, H. P. *Nat. Protoc.* **2006**, *1*, 153–161.
- (59) Koch, B.; Jensen, L. E.; Nybroe, O. *J. Microbiol. Meth.* **2001**, *45*, 187–195.
- (60) Guzman, L. M.; Belin, D.; Carson, M. J.; Beckwith, J. *J. Bacteriol.* **1995**, *177*, 4121–4130.
- (61) West, S. E.; Schweizer, H. P.; Dall, C.; Sample, A. K.; Runyen-Janecky, L. J. *Gene* **1994**, *148*, 81–86.
- (62) Schweizer, H. P. *Gene* **1991**, *97*, 109–121.
- (63) Heydorn, A.; Nielsen, A. T.; Hentzer, M.; Sternberg, C.; Givskov, M.; Ersbøll, B. K.; Molin, S. *Microbiology* **2000**, *146*, 2395–2407.
- (64) Tolker-Nielsen, T.; Sternberg, C. In *Pseudomonas Methods and Protocols*, Filloux, A., Ramos, J.-L., Eds.; Methods in Molecular Biology 1149, DOI: 10.1007/978-1-4939-0473-0\_47; Springer New York: 2014, pp 615–629.
- (65) Alhede, M.; Bjarnsholt, T.; Jensen, P. Ø.; Phipps, R. K.; Moser, C.; Christophersen, L.; Christensen, L. D.; van Gennip, M.; Parsek, M.; Høiby, N.; Rasmussen, T. B.; Givskov, M. *Microbiology* **2009**, *155*, 3500–3508.
- (66) Hong, V.; Presolski, S. I.; Ma, C.; Finn, M. G. *Angew. Chem. Int. Ed.* **2009**, *121*, 10063–10067.
- (67) Kalli, A.; Hess, S. *Proteomics* **2012**, *12*, 21–31.
- (68) Cox, J.; Mann, M. *Nat. Biotechnol.* **2008**, *26*, 1367–1372.
- (69) Cox, J.; Hein, M. Y.; Lubner, C. A.; Paron, I.; Nagaraj, N.; Mann, M. *Mol. Cell Proteomics* **2014**, *13*, 2513–2526.
- (70) Ritchie, M. E.; Phipson, B.; Wu, D.; Hu, Y.; Law, C. W.; Shi, W.; Smyth, G. K. *Nucleic Acids Res.* **2015**, gkv007.
- (71) Benjamini, Y.; Hochberg, Y. *J. Roy. Stat. Soc. B Met.* **1995**, *57*, 289–300.
- (72) Hunter, J. D. *Comput. Sci. Eng.* **2007**, *9*, 90–95.

- (73) Schneider, C. A.; Rasband, W. S.; Eliceiri, K. W. *Nat. Methods* **2012**, *9*, 671–675.

*Appendix A*

## SUPPLEMENTARY INFORMATION FOR CHAPTER 2

**A.1 Supplementary Experimental Procedures****Strain Construction.**

See Table A.1 for a full list of strains. An unmarked deletion of *sutA* (DKN1625) was generated by first cloning 1 kb of sequence upstream and downstream of this gene into the pMQ30 suicide vector [1]. This vector carries the URA3 gene from *Saccharomyces cerevisiae*, which facilitated the use of homologous recombination in yeast to stitch together the three DNA pieces. The upstream and downstream 1 kb regions were amplified from *P. aeruginosa* gDNA and cleaned up using the PCR purification kit (Qiagen). Linearized pMQ30 plasmid was transformed along with the 1 kb flanking regions into *S. cerevisiae* using standard methods and successful transformants were selected on media lacking uracil. The pMQ30 plasmid carrying the upstream and downstream sequences for *sutA* was recovered from the yeast colonies by extraction with phenol:chloroform:isoamyl alcohol and transformed into *E. coli* DH5 $\alpha$  cells. The construct was verified by sequencing and introduced into *P. aeruginosa* UCBPP-PA14 by triparental conjugation. Successful exoconjugants were selected on VBMM medium containing 100  $\mu\text{g}/\text{mL}$  gentamicin as described by Choi and Schweizer [2], and were then subjected to counterselection on LB plates lacking NaCl and containing 10% sucrose. Colonies resulting from homologous recombination to remove the wild-type copy of *sutA* and retain the clean deletion were identified by PCR.

The strain overexpressing SutA (DKN1626) was constructed by first cloning the SutA coding sequence into the multiple cloning site of the expression vector pMQ72, placing it under control of the arabinose-inducible *P<sub>ara</sub>* promoter, using yeast homologous recombination as described above. The *P<sub>ara</sub>* promoter:*sutA* coding sequence cassette was then cloned into the pUC18T-miniTn7T-GmR vector in order to direct its insertion into the attTn7 site of *P. aeruginosa* [2], using the Gibson reaction [3]. This vector was introduced into *P. aeruginosa* UCBPP-PA14 by tetraparental conjugation, and verified by PCR.

To construct the plasmid for overexpression of hemagglutinin-tagged SutA, the *sutA* gene, along with 1 kb upstream and downstream, was cloned from *P. aeruginosa*

gDNA with a 5' overhang encoding the HA epitope (MYPYDVDPYA) and inserted into pMQ30 using the Gibson reaction. The HA-*sutA* gene was then amplified and cloned into the multiple cloning site of pMQ72 between the SacI and KpnI restriction sites (DKN1640). This vector was transformed into *P. aeruginosa* by electroporation.

The GFP- and CFP-marked wild-type and  $\Delta sutA$  strains (DKN1632-1635) carry their respective fluorescent proteins under the control of the strong PA1/04/03 promoter, integrated into the *attTn7* site and marked by a gentamicin resistance cassette. The fluorescent markers were introduced into *P. aeruginosa* by tetraparental conjugation with *E. coli* strains carrying the respective fluorescent protein-encoding plasmids, which were gifts from the laboratory of Gary Schoolnik [4].

The super-folder GFP reporter strains (DKN1627-1628) were generated by first amplifying 1 kb of sequence upstream and the intergenic sequence downstream of the *sutA* and *rpsG* genes from *P. aeruginosa* gDNA. These fragments were cloned upstream and downstream of the sfGFP coding sequence [5] in the pUC18T-miniTn7T-GmR vector using the Gibson reaction, and the resulting construct was introduced into the *attTn7* site in *P. aeruginosa* by tetraparental conjugation.

### **Media and Growth Conditions.**

All cultures were grown at 37 °C with shaking unless otherwise noted. Liquid media were LB (5 g yeast extract, 10 g tryptone, 10 g NaCl per liter), 2xYT (10 g yeast extract, 16 g tryptone, and 5 g NaCl per liter), or phosphate buffered minimal medium (35.9 mM K<sub>2</sub>HPO<sub>4</sub>, 14.2 mM KH<sub>2</sub>PO<sub>4</sub>, 9.3 mM NH<sub>4</sub>Cl, 42.8 mM NaCl, 1.0 mM MgSO<sub>4</sub>, 7.5 μM FeCl<sub>2</sub> · 4 H<sub>2</sub>O, 0.8 μM CoCl<sub>2</sub> · 6 H<sub>2</sub>O, 0.5 μM MnCl<sub>2</sub> · 4 H<sub>2</sub>O, 0.5 μM ZnCl<sub>2</sub>, 0.2 μM Na<sub>2</sub>MoO<sub>4</sub> · 2 H<sub>2</sub>O, 0.1 μM NiCl<sub>2</sub> · 6 H<sub>2</sub>O, 0.1 μM H<sub>3</sub>BO<sub>3</sub>, 0.01 μM CuCl<sub>2</sub> · 2 H<sub>2</sub>O) with carbon sources added as noted. All anaerobic cultures were incubated in butyl rubber-stoppered Balch tubes in a Coy anaerobic chamber supplied with an atmosphere of 5% H<sub>2</sub>, 15% CO<sub>2</sub> and 80% N<sub>2</sub>, with trace amounts of oxygen removed by palladium-catalyzed reaction with the hydrogen gas. Anaerobic cultures were incubated without shaking.

### **BONCAT Labeling and Enrichment.**

L-azidohomoalanine (Aha) [6] and the dialkoxydiphenylsilane (DADPS) biotin-alkyne probe [7] were synthesized as previously described. *P. aeruginosa* PA14 was grown overnight in LB and diluted to *OD*<sub>500</sub> 0.02 into minimal medium containing 40 mM arginine, pH 7.2. The culture was grown to *OD*<sub>500</sub> 0.4, and split into

aerobic and anaerobic samples. To label aerobic cultures, Aha was added to a final concentration of 1 mM. After 15 min of incorporation, cells were washed once with PBS and cell pellets were frozen at -80 °C. Anaerobic samples were moved to an anaerobic chamber, washed with PBS, resuspended in minimal medium with 40 mM arginine, and sealed in Balch tubes. Anaerobic cultures were allowed to consume residual oxygen and adapt to anoxia for 24 h Aha was then added to a final concentration of 1 mM. After 16 h of incorporation, cells were pelleted, washed with PBS, and lysed immediately. For anaerobic samples, all steps up to and including lysis were performed using degassed solutions in the anaerobic chamber.

All samples were lysed by resuspension in lysis buffer (100 mM Tris-HCl, pH 8, 1% SDS). Lysates were heated to 65 °C for 5 min and clarified by addition of Benzonase Nuclease (Sigma Aldrich) for 1 h at 37 °C followed by centrifugation. For fluorescence detection of Aha-labeled proteins, lysates were reacted with 5  $\mu$ M TAMRA-alkyne (Click Chemistry Tools), 100  $\mu$ M chCuSO<sub>4</sub>, 500  $\mu$ M tris(3-hydroxypropyltriazolylmethyl)amine (THPTA), 5 mM aminoguanidine hydrochloride, and 5 mM sodium ascorbate [8] for 15 min at room temperature, precipitated with water, methanol, and chloroform, and washed twice with methanol. Reacted lysates were separated via SDS-PAGE and imaged on a Typhoon gel imager (GE Healthcare). Gels were stained with Colloidal Blue (Life Technologies) to verify equal protein loading.

For protein enrichment, 0.5 mg of each protein lysate was reacted with 100  $\mu$ M DADPS biotin-alkyne probe as above for 3.5 h at room temperature. Proteins were precipitated with acetone at -20 °C and resuspended in PBS with 0.3% SDS. Streptavidin UltraLink Resin (Pierce Biotechnology) was washed twice with PBS, added to biotinylated lysates, and incubated overnight at 4 °C. Resin was transferred to microfuge spin columns (Pierce Biotechnology) and washed twice with 1% SDS in PBS and once with 0.1% SDS in PBS. Proteins were eluted by cleavage of the DADPS linker via incubation with 5% formic acid and 0.1% SDS in PBS for 2 h at room temperature. Resin was washed with 0.1% SDS in PBS to elute all proteins. Elution fractions were combined and concentrated by centrifugation through Amicon Ultra spin columns (EMD Millipore). The entirety of the concentrated eluents were separated via SDS-PAGE and stained with Colloidal Blue.

### **Protein Digestion, Mass Spectrometry, and Data Analysis.**

For GeLCMS, gel pieces were destained by alternating washes with 50 mM ammonium bicarbonate (AB) and 1:1 50 mM AB:acetonitrile. Proteins were reduced by

incubation with 6.7 mM dithiothreitol (DTT) in 50 mM AB at 50 °C for 30 min and alkylated by incubation with 37 mM iodoacetamide in 50 mM AB at room temperature for 20 min. Gel pieces were washed with 100 mM AB and then with acetonitrile. Proteins were digested with 300 ng endoproteinase LysC in 100 mM Tris-HCl at 37 °C for 18 h. Peptides were extracted by sequential washing with: 1% formic acid/2% acetonitrile, 1:1 acetonitrile:water, and 1% formic acid in acetonitrile. Extracted peptides were dried and desalted using C18 StageTips as previously described [9].

For in-solution digestion, proteins were brought to a final concentration of 8 M urea, reduced by incubation with 3 mM tris(2-carboxyethyl) phosphine (TCEP) for 20 min at room temperature, and alkylated by incubation with 10 mM iodoacetamide for 15 min at room temperature in the dark. For immunoprecipitation, proteins were digested with 250 ng endoproteinase LysC for 18 h at room temperature. Samples were further digested by dilution with 100 mM Tris-HCl to a final urea concentration of 2 M and addition of 600 ng trypsin and 1 mM CaCl<sub>2</sub> at room temperature for 9 h. Digestion was quenched by addition of 5% formic acid. Digested peptides were desalted by HPLC using a Michrom Bioresources C18 macrotrap, (Buffer A: 0.2% formic acid in chH<sub>2</sub>O; Buffer B: 0.2% formic acid in acetonitrile) and concentrated *in vacuo*. Peptides were dimethyl labeled following established protocols [10] and mixed in a 1:1 mass ratio.

Liquid chromatography-mass spectrometry were essentially carried out as previously described [11]. Anaerobic vs. aerobic BONCAT and immunoprecipitation experiments were performed on a nanoflow LC system, EASY-nLC 1000 coupled to a hybrid linear ion trap Orbitrap Classic mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) equipped with a nanoelectrospray ion source (Thermo Fisher Scientific) with the following modifications: For the EASY-nLC II system, solvent A consisted of 97.8% chH<sub>2</sub>O, 2% ACN, and 0.2% formic acid and solvent B consisted of 19.8% chH<sub>2</sub>O, 80% ACN, and 0.2% formic acid. For the LC-MS/MS experiments, digested peptides were directly loaded at a flow rate of 500 nL/min onto a 16-cm analytical HPLC column (75 μm ID) packed in-house with ReproSil-Pur C<sub>18</sub>AQ 3 μmresin (120 Å pore size, Dr. Maisch, Ammerbuch, Germany). The column was enclosed in a column heater operating at 30 °C. After 30 min of loading time, the peptides were separated with a 50 min gradient at a flow rate of 350 nL/min. The gradient was as follows: 0–30% B (50 min), and 100% B (10 min). The Orbitrap was operated in data-dependent acquisition mode to automatically alternate between a full scan (m/z=400–1600) in the Orbitrap and subsequent 10 CID MS/MS

scans in the linear ion trap. CID was performed with helium as collision gas at a normalized collision energy of 35% and 30 ms of activation time.

For the BONCAT experiment, raw files were searched using MaxQuant [12] against the *P. aeruginosa* PA14 UniProt entries (5,886 sequences) and a contaminant database (246 sequences). Trypsin was specified as the digestion enzyme with up to two missed cleavages. Carbamidomethylation of cysteine was set as a fixed modification and protein N-terminal acetylation and methionine oxidation were variable modifications. We also included variable modifications of methionine corresponding to Aha, reduced Aha, Aha reacted to the DADPS linker, and Aha reacted to the cleaved DADPS linker. Protein ratios and their standard errors were calculated using bootstrap estimates and pooled variance estimates at the peptide level [13]. Briefly, peptide intensities were normalized to the total intensity for each run and a global estimate of measurement error was calculated using pooled variance from all peptide ratios between each sample. The protein ratio was calculated as the median of peptide ratios. The standard error of the protein ratio was calculated using a bootstrap procedure where resampling of peptide ratios is augmented by adding a random “noise” effect drawn from a normal distribution with mean zero and standard deviation equal to the previously calculated global estimate of measurement error. In total, 1000 bootstrap iterations were performed. The standard error of the protein ratio was then calculated as the standard deviation of the bootstrapped peptide ratios. Z-tests were then used to calculate p-values of overall protein ratios with respect to a 1-to-1 ratio. P-values were adjusted for false discovery by the Benjamini-Hochberg procedure.

For dimethyl labeling experiments, raw files were searched using MaxQuant as above. Trypsin was specified as the digestion enzyme with up to two missed cleavages. Carbamidomethylation of cysteine was set as a fixed modification and protein N-terminal acetylation and methionine oxidation were variable modifications. Dimethyl mass modifications (light and medium) at lysine residues and peptide N-termini were specified for quantification.

### **Colony Morphology Assay.**

Cultures were grown overnight in LB medium, diluted 1:1000, and spotted in a 10  $\mu$ L volume on solid media (1% tryptone, 1% Bacto Agar, 20  $\mu$ g/ml Coomassie blue and 40  $\mu$ g/ml Congo red) [14]. Plates were incubated at room temperature for 6 days and then imaged using a Keyence VHX-1000 digital microscope.

### **Crystal Violet Assay.**

The Crystal Violet assay was performed as previously described [15]. Cultures were grown overnight in LB and diluted 1:1000 into LB. 125  $\mu$ l of each diluted culture was transferred to 96-well round bottom polystyrene plates coated for tissue culture (Corning). Plates were sealed with parafilm and incubated for 18 h at 37 °C without shaking. Wells were washed with 0.9% NaCl and treated with 150  $\mu$ l of 0.1% Crystal Violet for 20 min at room temperature. Wells were washed three times with water and Crystal Violet was extracted from adherent cells by addition of ethanol. Ethanol containing Crystal Violet was transferred to a new well-plate and absorbance at 600 nm was measured. The average absorbance for wells containing only LB was subtracted from all measurements. Each strain was measured in two separate experiments, with four wells per experiment.

### **Phenazine Measurements.**

Phenazine concentrations in culture supernatants were measured as described previously [16]. Briefly, culture supernatants were filtered using SpinX columns with a 0.2  $\mu$ M pore size and were directly loaded onto a Beckman System Gold reverse-phase high-pressure liquid chromatography (HPLC) instrument with a UV-visible light (Vis) detector and a Waters Symmetry C<sub>18</sub> analytical column (5  $\mu$ m particle size; 4.6 by 250 mm). A gradient of water-0.1% trifluoroacetic acid (TFA; solvent A) to acetonitrile-0.1% TFA (solvent B) at a flow rate of 1 ml/min was used to elute phenazines, which can be detected based on their characteristic absorption wavelengths and retention times. Peak areas for samples were compared to peak areas from standards of purified phenazine-1-carboxylic acid and pyocyanin.

### **Competition Assay.**

Individual overnight cultures of wild-type cells carrying a *gfp* or a *cfp* marker and  $\Delta sutA$  cells carrying a *gfp* or a *cfp* marker were grown in 5 ml LB medium. Cultures were diluted 1:1000 in LB medium and mixed in equal proportions based on their ODs in the following combinations: A) wild type, *gfp* marked plus  $\Delta sutA$ , *cfp* marked; B) wild type, *cfp* marked plus  $\Delta sutA$ , *gfp* marked; and C) wild type, *gfp* marked plus wild type, *cfp* marked. The mixtures were allowed to grow to mid-exponential phase (OD approximately 0.4), with shaking at 37 °C. Small aliquots of the mixed cultures were taken for microscopy (time 0 sample) and the remainders were pelleted and transferred to an anaerobic chamber (Coy), where they were resuspended in anaerobic minimal medium with 40 mM arginine and placed in

sealed Balch tubes. Cultures were incubated anaerobically at 37 °C for 19-20 h, then were removed from the anaerobic chamber, diluted 1:100 or 1:200 into LB medium, and allowed to grow aerobically with shaking at 37 °C for 4-6 h, back to mid-exponential phase. No significant change in OD occurred during the anaerobic incubation. Once cells reached mid-exponential phase, a small aliquot of the culture was taken for microscopy (transfer 1 sample) and the remainders of the cultures were pelleted and resuspended again in the anaerobic arginine medium in the anaerobic chamber. This process was repeated for 4 transfers. At each transfer, epifluorescence microscopy using a Zeiss Axio Imager microscope was used to observe live cells placed on agarose pads. GFP was detected using the Zeiss 46HE filter cube and CFP was detected using the Zeiss 47HE filter cube. The percentage of cells carrying each marker in each mixed culture was counted. At least 500 cells were counted for each sample. A very small bias in favor of carrying CFP over GFP was detected in the wild type vs. wild type mixed culture (combination C), so at each time point, the proportion of each marker in this culture was taken to reflect the “no advantage” state, and the wild type vs.  $\Delta sutA$  proportions were adjusted by the difference observed due to carrying GFP vs. CFP. The adjusted proportions in the two marker-flipped cultures (combinations A and B) were averaged together. The entire experiment was performed three times.

#### **GFP Reporter Protein Measurement.**

For growth, transcript and reporter protein measurements, starter cultures were grown to stationary phase in LB medium, diluted 1:1000 into either LB or pyruvate minimal medium and allowed to grow into early exponential phase (approx. 4 h for LB or 18 h for pyruvate) at which point the “Time 0” measurements were made. Live cells in liquid culture were diluted between 1:250 and 1:1000 into TBS (50 mM Tris, pH 8.0, 150 mM NaCl) containing the SYTO 62 red-fluorescent, cell permeant nucleic acid stain at a concentration of 500 nM. Cells were incubated at room temperature in the dark for 15-20 min to allow for DNA staining. The BD Accuri c6 flow cytometer was used to measure both red fluorescence from the SYTO 62 dye (excitation laser: 640 nM, emission filter: 675/25 nM) and green fluorescence from GFP (excitation laser 488 nM, emission filter: 533/30 nM). Particles were gated on forward scatter vs. red fluorescence, and the mean green fluorescence for particles with red fluorescence and forward scatter values consistent with cells was measured. At each time point, mean green fluorescence in a strain lacking GFP was also measured to determine background autofluorescence and this value was

subtracted from the GFP values for that time point. GFP was measured for biological triplicates for each genotype and condition.

### **RNA Extraction.**

Total RNA was extracted from cells using the RNeasy Mini Kit (Qiagen). Briefly, approximately  $10 \times 10^9$  cells were pelleted rapidly by centrifugation at top speed in a microfuge, the supernatant was removed, and the pellet was immediately frozen in liquid nitrogen. After all samples were collected, pellets were resuspended in TE buffer (10 mM Tris pH 7.5, 1 mM ethylenediaminetetraacetic acid (EDTA)) plus 15 mg/ml lysozyme (Sigma) and 15 U per sample proteinase K (Qiagen), and incubated for 10 min at 37 °C to digest the cell wall. Samples were then processed according to the manufacturer's instructions, including on-column DNase treatment. Purified RNA was quantified by absorbance at 260 nm, and 10 µg per sample was treated with Turbo DNase Free (Ambion) according to the manufacturer's instructions. Samples were verified to be free of genomic DNA by quantitative PCR.

### **qRT-PCR.**

1 µg of DNase-treated total RNA was converted to cDNA using the iScript cDNA synthesis kit (BioRad). 1/100th of this reaction mixture (representing 10 ng total RNA) was used per qRT PCR reaction, along with 500 nM each of forward and reverse primers and the iTaq SYBR Green reaction mix (BioRad). Samples were run on the ABI platform (ABI) for 40 cycles with an annealing temperature of 60 °C. Standard curves for each primer pair were generated using serial dilutions of genomic DNA. The *oprI* gene was used to normalize against potential loading differences. See Table A.2 for primer sequences. Measurements were made on biological triplicates.

### **Co-Immunoprecipitation**

Cultures of  $\Delta sutA$  carrying pMQ72 or pMQ72-HAsutA were grown overnight in minimal medium containing 40 mM sodium pyruvate, 20 mM arabinose, and 50 µg/ml gentamicin to an OD<sub>500</sub> of approximately 1. Cells were washed once in PBS and frozen at -80 °C. Cell pellets were resuspended in IP lysis buffer (50 mM HEPES, 70 mM potassium acetate, 5 mM magnesium acetate, 0.2% n-dodecyl-β-D-maltoside, and cOmplete mini protease inhibitor, EDTA free (Roche)). Cells were gently lysed by passage through a 22G needle ten times. Lysates were clarified by incubation with Benzonase Nuclease for 1 h at 37 °C followed by centrifugation.

For IP of HA-SutA, 50  $\mu$ l agarose beads conjugated to an anti-HA antibody (Sigma-Aldrich) were washed three times in IP lysis buffer, combined with 1 ml lysate, and incubated with rotation overnight at 4 °C. For IP of RpoA, 1 ml lysate was incubated with an anti-RpoA antibody (gift of Olaf Schneewind) for 1 h at 4 °C with rotation. 50  $\mu$ l Protein A/G PLUS-agarose beads (Santa Cruz Biotechnology) were washed three times with IP lysis buffer, combined with the antibody-lysate mixture, and incubated with rotation overnight at 4 °C. For both IPs, beads were washed twice with 0.5 ml IP lysis buffer and twice with 0.5 ml 100 mM Tris-HCl, pH 8. Proteins were eluted by incubation with 64  $\mu$ l 10 M urea in 100 mM Tris-HCl. IP eluents were digested in-solution, reacted with dimethyl labels, and analyzed by LC-MS/MS, as described above.

For Western blotting, 10  $\mu$ L of each IP fraction (lysate, flow through, four washes, and elution) were separated by SDS-PAGE and transferred to a Hybond ECL membrane (GE Healthcare). Membranes were blocked with 5% milk in TBST (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 0.05% Tween 20). HA-SutA was detected by anti-HA antibody-Alexa Fluor 594 conjugate (Life Technologies). RpoA was detected by incubation with the primary anti-RpoA antibody described above, followed by incubation with a goat anti-mouse antibody-Alexa-Fluor 633 conjugate (Life Technologies). On a separate gel, the same samples were stained with Coomassie.

### **RNA Seq Library Preparation.**

For RNA-Seq experiments, starter cultures were grown to stationary phase in LB, diluted 1:1000 in pyruvate minimal medium containing 25 mM arabinose, and then allowed to grow 21 h until they reached stationary phase again ( $OD_{500}$  of approximately 1), at which point cells were collected for RNA extraction (described above). 3.8  $\mu$ g of DNase-treated total RNA was subjected to ribosomal RNA depletion using the Gram Negative Magnetic Ribo-Zero kit (Epicentre), according to the manufacturer's instructions. Following rRNA depletion, samples were cleaned up using the RNeasy MinElute kit (Qiagen) and libraries were generated for sequencing using the NEBNext mRNA Library Prep Kit for Illumina (NEB). Briefly, mRNAs were fragmented by treatment with MgCl<sub>2</sub>-containing fragmentation buffer for 1 min at 94 °C and cleaned up using the RNeasy MinElute columns. Fragmentation to an average size of approximately 200 bp was verified by running the samples on a Bioanalyzer RNA Pico chip (Agilent). The fragmented RNA was reverse-transcribed to cDNA, which was then end-repaired, dA-tailed, and ligated to adaptors. Each sample was PCR-amplified with a universal primer and a unique bar-coded primer, using 12

amplification cycles. Final libraries were verified using the High-Sensitivity DNA chip on the Bioanalyzer and quantified using the Qubit fluorimeter and dsDNA dye (Invitrogen). Sequencing was performed on biological triplicates for each genotype.

### **Chromatin Immunoprecipitation.**

Growth conditions were the same as for the RNA-seq experiments except 20 mM arabinose was used and 50 µg/ml gentamicin was added for plasmid maintenance. Stationary phase cultures of the  $\Delta sutA$  strain (DKN1625) carrying either pMQ72 or pMQ72-HA-sutA in pyruvate minimal medium were cross-linked by incubation with 1% formaldehyde at room temperature for 15 min and then crosslinking was quenched by incubation with 125 mM glycine for 10 min. Cells were pelleted and washed twice with TBS (50 mM Tris pH 7.5, 150 mM NaCl), and then pellets were frozen at -80 °C. Frozen pellets were resuspended in 1.5 mL IP buffer (50 mM Tris pH 7.5, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% deoxycholic acid, 1 mg/ml lysozyme) and incubated at 37 °C for 15 min. Samples were then chilled on ice and sonicated using a microtip sonicator for 4 min at the 4.0 setting, using a cycle of 30 s on, 30 s off. Samples were split in half; one half was subjected to immunoprecipitation by an antibody against RpoA, and the other half was subjected to immunoprecipitation by an antibody against the HA epitope, as described for protein IP above. For the RpoA immunoprecipitation, samples were pre-cleared by incubation with 1/10 volume Protein A/G PLUS-agarose beads (Santa Cruz Biotechnology) for 1 h at 4 °C, and then were incubated overnight with rotation at 4 °C with the anti-RpoA antibody. Next 50 µl of the protein A/G agarose beads were added and the mixture was incubated for an additional 1 h at 4 °C. For the HA-SutA immunoprecipitation, samples were incubated with 50 µl pre-conjugated HA bead slurry overnight with rotation at 4 °C. The beads from both immunoprecipitations were then washed 5 times for 10 min per wash. Washes 1 and 2 were with IP buffer, wash 3 was with IP buffer with 500 mM NaCl, wash 4 was with stringent buffer (10 mM Tris pH 8.0, 250 mM LiCl, 1 mM EDTA, 0.5% Nonidet P-40), and wash 5 was with TBS. DNA/protein complexes were eluted from the beads in 100 µL elution buffer (50 mM Tris pH 7.5, 10 mM EDTA, 1% SDS) by incubation for 15 min at 65 °C. The elution was repeated once and both eluates were combined, then were incubated at 65 °C overnight to reverse crosslinks. 200 µl TE buffer (10 mM Tris pH 8.0, 1 mM EDTA), 100 µg proteinase K, and 20 µg glycogen were added to each sample and they were incubated for 2 h at 37 °C to digest proteins. DNA was extracted using 25:24:1 phenol:chloroform:isoamyl alcohol and precipitated with

ethanol. The precipitated DNA was resuspended in 30  $\mu$ l TE buffer containing 10  $\mu$ g RNase A and incubated at 37  $^{\circ}$ C for 2 hr to remove RNA contamination, and then was cleaned up using a QIAquick column (Qiagen) with an elution volume of 50  $\mu$ l [17].

### **ChIP Seq library preparation.**

2-10 ng purified genomic DNA isolated by immunoprecipitation was subjected to further fragmentation by treatment with the NEB ds Fragmentase enzyme cocktail for 10 min at 37  $^{\circ}$ C. This reduced the average fragment size from approximately 500-1000 bp to approximately 200-500 bp for optimal high throughput sequencing efficiency. Fragmented DNA was cleaned up using Agencourt AMPure XP magnetic beads (Beckman Coulter). Libraries were prepared from the fragmented gDNA using the NEBNext ChIP Seq Library Prep Reagent Set for Illumina (NEB). DNA fragments were end-repaired, dA-tailed, ligated to adaptors, and PCR amplified with one universal and one bar-coded primer, using 15 amplification cycles. Final libraries sizes were verified using the Bioanalyzer, and library amounts were quantified using the Qubit fluorimeter. All ChIP-Seq was performed on biological triplicates.

### **Sequencing and Data Analysis.**

All sequencing was performed by the Millard and Muriel Jacobs Genetics and Genomics Laboratory at the California Institute of Technology using the Illumina HiSeq 2500 platform. 10-15 million reads of 50 or 75 bp each were collected for each sample. Base-calling and de-multiplexing were performed by the Illumina HiSeq Control Software (HCS, version 2.0). The resulting FASTQ files were concatenated into one file per sample and filtered and trimmed by quality score per base using the Trimmomatic software package with the following parameters: LEADING:27 TRAILING:27 SLIDINGWINDOW:4:20 MINLEN:35 [18]. Surviving reads were mapped to the *P. aeruginosa* UCBPP-PA14 genome sequence (gi|116048575|ref|NC\_008463.1) using the Bowtie package with the -n 2 and -best arguments [19]. Specifically for assessing ChIP signal at tRNA genes, Bowtie was run with the -n 2 and -m 1 arguments to require reads to be uniquely mapped in order to be reported. Mapped reads were sorted, indexed, and converted to binary format using the SAMtools package [20]. Reads per 100 bp, gene, or transcriptional unit (TU) were calculated using the easyRNASeq package from the Bioconductor project in R [21]. The .gff file describing the location of genes was generated using the bp\_genbank2gff3.pl script from the Bioperl project and the Genbank file for

the RefSeq accession NC\_008463.1. The .gff file was modified to additionally include small non-coding RNAs and novel ORFs detected by deep sequencing of the UCBPP-PA14 strain of *P. aeruginosa* [22], and to consistently name genes by their locus tags rather than a mixture of locus tags and gene names. The .gff file describing the locations of transcriptional units was derived from the table of transcriptional units published defined by Wurtzel and colleagues [22], and uses the start of the first coding sequence and the end of the last coding sequence in each operon as the operon boundaries. Average ratios and significance of differential expression or ChIP association between different genotypes or pulldowns were calculated using the Degust web server hosted by the Victorian Bioinformatics Consortium. The Degust project uses the voom and limma packages in R to perform calculations [23].

For viewing ChIP data across genomic loci, the counts per 100 bp for each sample were normalized to the size of the library by converting counts to RPKM, and then further scaled based on the values observed low- and high-signal regions. This method was adapted from one described by Mooney et al. [24]. The baseline value for each sample was defined as the average RPKM value for the 25 transcriptional units at least 1 kb in length that had the lowest signal in the RpoA pulldown from the HA-SutA strain. These transcriptional units were verified to have among the lowest RPKM values from the RNA-Seq data as well, and were assumed to be essentially not transcribed under the conditions of the experiment. The maximum value for each sample was defined as the average RPKM value for the top ten peaks associated with protein-coding genes for that type of pulldown. A peak was defined as two consecutive 100bp regions that fell among the top 100 100bp regions. While some peak regions were the same for both the HA-SutA and the RpoA pulldown, some were distinct. See Dataset A.4 for regions and values used. To scale the RPKM data, the baseline value was subtracted from each 100bp RPKM value and the result was divided by the maximum value, such that nearly all scaled values fall between 0 and 1. The biological triplicates for each pulldown were averaged. The MochiView software package [25] was used to smooth the scaled 100 bp values over a 300 bp rolling window, and then the coordinates of regions with scaled values above 0.20 for the HA-SutA pulldown and scaled values above 0.25 for either RpoA pulldown were extracted. Regions less than 100 bp apart were merged. This set of “high chip signal” regions was then filtered to include only 100 bp regions that also showed a statistically significant enrichment in the HA-SutA pulldown compared to the mock pulldown, which left a total of 2,015 100bp regions that were considered “high

ChIP.” 230 transcriptional units starting within a high chip region were identified. There were 405 genes that were contained within these transcriptional units and were considered the list of “high chip” genes that was compared with the list of up- and down-regulated genes. For the aggregate ChIP plot, transcriptional units containing ribosomal protein genes were excluded, since these had already been separately considered, and of the remaining, only transcriptional units with start sites defined by Wurtzel et al. [22] were included. See Dataset A.6 for the transcriptional unit data that was used.

Functional analysis of genes transcriptionally affected more than 2-fold was carried out using the COG category designations recorded in the *Pseudomonas* Genome Database (<http://www.pseudomonas.com>) [26]. For simplicity, several COG categories were grouped together for each bar in the bar plot. The category designated “unknown” contains COG categories R and S (“General functional prediction only” and “No functional prediction”) in addition to genes that did not have an associated COG. The category designated “maintenance and secondary metabolism” contains COG categories C, I, P, O, and Q (“Energy production and conversion”, “Lipid transport and metabolism”, “Inorganic ion transport and metabolism”, “Post-translational modification, protein turnover, and chaperones”, and “Secondary metabolites biosynthesis, transport, and catabolism”). The category designated “growth and primary metabolism” contains COG categories D, E, F, G, H, J, L, and M (“Cell cycle control, cell division, chromosome partitioning”, “Amino acid transport and metabolism”, “Nucleotide transport and metabolism”, “Carbohydrate transport and metabolism”, “Coenzyme transport and metabolism”, “Translation, ribosomal structure and biogenesis”, “Replication, recombination and repair”, and “Cell wall/membrane/envelope biogenesis”). The category designated “motility, defense, and signaling” contains COG categories N, T, U, and V (“Cell motility”, “Signal transduction mechanisms”, “Intracellular trafficking, secretion, and vesicular transport”, and “Defense mechanisms”). The category designated “transcription and nucleic acid processing” contains COG categories A, B, and K (“RNA processing and modification”, “Chromatin structure and dynamics”, and “Transcription”) (<ftp://ftp.ncbi.nih.gov/pub/COG/COG2014/static/lists/homeCOGs.html>) [27].

### **Software analysis and data presentation.**

This section describes software packages that were not mentioned above. Data processing and statistical analysis were performed with Python version 2.7.9 with

NumPy version 1.9.2, SciPy version 0.15.1, and Pandas version 0.16.1. Data were plotted with Matplotlib version 1.4.3 [28] and Seaborn version 0.5.1. Gel images were analyzed with ImageJ 64-bit version 1.45 [29]. Figures were assembled in Adobe Illustrator CS5.

## A.2 Supplementary Tables

Table A.1: **Chapter 2: Strains and plasmids.** Strains and plasmids used in this study. Plasmids are stored as *E. coli* strains carrying the plasmid, and requests should be for the *E. coli* strain.

<i>Pseudomonas aeruginosa</i> Strains		
Name	Genotype	Source
DKN263	<i>P. aeruginosa</i> UCBPP-PA14	
DKN1625	UCBPP-PA14 $\Delta sutA$	This Study
DKN1626	UCBPP-PA14 attTn7::mini-Tn7T- <i>Gm<sup>R</sup></i> <i>P<sub>ara</sub>:sutA</i>	This Study
DKN1627	UCBPP-PA14 attTn7::mini-Tn7T- <i>Gm<sup>R</sup></i> <i>P<sub>sutA</sub>:gfp</i>	This Study
DKN1628	UCBPP-PA14 attTn7::mini-Tn7T- <i>Gm<sup>R</sup></i> <i>P<sub>rpsG</sub>:gfp</i>	This Study
DKN1632	UCBPP-PA14 attTn7:: mini-Tn7T- <i>Gm<sup>R</sup></i> <i>P<sub>A10403</sub>:gfp</i>	This Study
DKN1633	UCBPP-PA14 attTn7:: mini-Tn7T- <i>Gm<sup>R</sup></i> <i>P<sub>A10403</sub>:cfp</i>	This Study
DKN1634	UCBPP-PA14 $\Delta sutA$ attTn7::mini-Tn7T- <i>Gm<sup>R</sup></i> <i>P<sub>A10403</sub>:gfp</i>	This Study
DKN1635	UCBPP-PA14 $\Delta sutA$ attTn7::mini-Tn7T- <i>Gm<sup>R</sup></i> <i>P<sub>A10403</sub>:cfp</i>	This Study
Transposon insertion mutants	UCBPP-PA14 Gene::MAR2xT7	[30]
<i>Escherichia coli</i> Strains		
Name	Genotype	Source
DKN1299	SM10, pTNS1	[2]
DKN1299	HB101 pRK2013	[2]
DKN1323	S17-1 $\lambda$ pir pMCM11(containing attTn7::mini-Tn7T- <i>Gm<sup>R</sup></i> <i>P<sub>A10403</sub>:gfp</i> )	Gary Schoolnik
DKN1325	S17-1 $\lambda$ pir pMCM11 derivative (containing attTn7::mini-Tn7T- <i>Gm<sup>R</sup></i> <i>P<sub>A10403</sub>:cfp</i> )	Gary Schoolnik
DKN1637	DH5 $\alpha$ pMQ30_ <i>sutA</i>	This Study
DKN1639	Mach1 pUC18T-mini-Tn7T- <i>Gm<sup>R</sup></i> <i>P<sub>ara</sub>:sutA</i>	This Study
DKN1640	Mach1 pMQ72_ <i>HA-sutA</i>	This Study
DKN548	DH5 $\alpha$ pMQ72	George O'Toole
DKN1641	DH10 $\beta$ pUC18T-mini-Tn7T- <i>Gm<sup>R</sup></i> <i>P<sub>sutA</sub>:gfp</i>	This Study
DKN1642	DH5 $\alpha$ pUC18T-mini-Tn7T- <i>Gm<sup>R</sup></i> <i>P<sub>rpsG</sub>:gfp</i>	This Study
<i>Saccharomyces cerevisiae</i> Strains		
Name	Genotype	Source
DKN569	InvSc1:MAT $\alpha$ /MAT $\alpha$ his3D1/his3D1 leu2/leu2 trp1-289/trp1-289 ura3-52/ura3-52	Invitrogen

Table A.2: **Chapter 2: Primers.** Primers used in strain construction and qRT-PCR experiments.

Name	Purpose	Sequence
6977del1	Generating SutA deletion construct	tgggtaacgccagggttttcccagtcacgacggtgtaaaaCTGCTCACCGGGATCTTCGC
6977del2	Generating SutA deletion construct	TGGCGGGCCTTGGGATGACGCGAAAGGTCAACCTCTCGGTGCTGCAAAAAG
6977del3	Generating SutA deletion construct	CTTTTGCAGCACCGAGAGGTTGACCTTTCGCGTCATCCCAAGGCCCGCCA
6977del4	Generating SutA deletion construct	tgtgagcggataacaatttcacacaggaaacagctatgacG TTCAGCCGGGCGGCAGCGA
Para:sutA1	Cloning SutA into pMQ72	ccataaccgtttttgggctagcgaattcgagctcAGGAGGGGTTGACCATGAGCGAAG
Para:sutA2	Cloning SutA into pMQ72	gcaaattctgtttatcagaccgcttctgcgttctgatttaaAAATCAGATGGGGCGGCT
sutA_gfp1	Generating SutA:gfp reporter construct	agtataggaacttcagagcgctttgaagctaattcgatcCTGCTCACCGGGATCTTCGC
sutA_gfp2	Generating SutA:gfp reporter construct	TGAACAGCTCTTCGCCTTTACGCATGGTCAACCTCTCGGTGCTGCAAAAAGC
sutA_gfp3	Generating SutA:gfp reporter construct	GCTTTTGCAGCACCGAGAGGTTGACCATGCGTAAAGGCCGAAGAGCTGTTCA

Continued from previous page.

Name	Purpose	Sequence
sutA_gfp4	Generating SutA:gfp reporter construct	TGGCGGGCCTTGGGATGACGCGAAATCATCATTTGTACAGTTCATCCATA
sutA_gfp5	Generating SutA:gfp reporter construct	TATGGATGAACTGTACAAATGATGATTTTCGCGTCATCCCAAGGCCCGCCA
sutA_gfp6	Generating SutA:gfp reporter construct	atagtttgaactagattcacttatctggttggcctgcaGGGATGACAACCGATGTGTC
rpsG_gfp1	Generating RpsG:gfp reporter construct	agtataggaacttcagagegctttgaagctaattcgatcATCAAAGGCGACCAGGTGGA
rpsG_gfp2	Generating RpsG:gfp reporter construct	TGAACAGCTCTTCGCCTTTACGCATTGATAAGCCCTCAAACGGTCTTCAG
rpsG_gfp3	Generating RpsG:gfp reporter construct	CTGAAGACCGTTTGAGGGCTTATCAATGCGTAAAGGCGAAGAGCTGTTCA
rpsG_gfp4	Generating RpsG:gfp reporter construct	CCTTTTCTGATGGCAGGATCAGCGATCATCATTTGTACAGTTCATCCATA
rpsG_gfp5	Generating RpsG:gfp reporter construct	TATGGATGAACTGTACAAATGATGATCGCTGATCCTGCCATCAGAAAAGG
rpsG_gfp6	Generating RpsG:gfp reporter construct	atagtttgaactagattcacttatctggttggcctgcaGACCTCAGACTCCAATTTAC
HAsutA1	Generating HA-SutA	GACCGCATGTACGCCGAAGcggggatcctctagagtcgacctgcaggca
HAsutA2	Generating HA-SutA	cagctatgacctgattacgaattc

Continued from previous page.

Name	Purpose	Sequence
HAsutA3	Generating HA-SutA	tgctgcaggtcgactctagaggatccccgCTTCGGCGTACATGCGGTC
HAsutA4	Generating HA-SutA	cagcaccgagaggttgaccATGTACCCATACGATGTTCCAGATTACGCT
HAsutA5	Generating HA-SutA	ATGTACCCATACGATGTTCCAGATTACGCTatgagcgaagaagaactggaac
HAsutA6	Generating HA-SutA	cagctatgaccatgattacgaattcACGAGATTGAACGGGGTAAC
HAsutA7	Moving HA-SutA to pMQ72	atatggtaccCTTCGGCGTACATGCGGTC
HAsutA8	Moving HA-SutA to pMQ72	atatgagctcACGAGATTGAACGGGGTAAC
Sfgfp_f	QPCR	TGGTGTTTCAGTGCTTTGCTC
Sfgfp_r	QPCR	TGTACGTGCCGTCATCCTTA
oprI_f	QPCR	AGCAGCCACTCCAAAGAAAC
oprI_r	QPCR	CAGAGCTTCGTCAGCCTTG
Intergenic_f	QPCR	GGGGTGGGGGTAGTTAAAGA
Intergenic_r	QPCR	GCAAAACAAGCCCCTACAAA
16Sleader_f	QPCR	ACGAAAGCCTTGACCAACTG
16Sleader_r	QPCR	TTGCGCTGCTGATAATCTTG

### A.3 Supplementary Datasets

**Dataset A.1: Proteins more abundant or uniquely identified in the anaerobic sample.** All proteins identified by LC-MS/MS from the BONCAT-enriched samples are listed. Columns 1-3 give the locus ID in both the PA14 and PAO1 strains as well as the gene name if available. Columns 4-8 give LC-MS/MS measurements for each protein or protein group: the number of unique peptides identified, number of evidences in the anaerobic and aerobic samples respectively, and the total peak intensities in the anaerobic and aerobic samples respectively. Columns 9 and 10 give the  $\log_2$ -transformed median of all intensity ratios for peptides shared between the two samples, and the probability that the ratio is not different from zero, with an adjustment for the pooled variance of the experiment. Sheet1 lists proteins identified in both samples, Sheet2 lists proteins identified only in the anaerobic sample, and Sheet3 lists proteins only identified in the aerobic sample. NA: not available. NQ indicates that there was insufficient information for that protein to quantify a ratio between the anaerobic and aerobic samples.

**Dataset A.2: Proteomic results from co-immunoprecipitations.** All co-precipitated proteins identified by LC-MS/MS following pull-down of either SutA (Sheet1) or RpoA (Sheet 2) are listed. For the SutA immunoprecipitation, two independent experiments were performed, and the results are listed separately. Columns 1-3 list the locus IDs for both the PA14 and PAO1 strains, and the gene name if available. Column 4 lists the  $\log_2$ -transformed ratios between protein abundance in the HA-SutA sample and the untagged control sample, as quantified by dimethyl labeling, and columns 5-7 give the number of evidences and total peak intensities for the differentially labeled peaks. Columns 8-11 give this information for the second IP experiment. For the RpoA pulldown, the gene identification information is the same as for the HA-SutA pulldown, and number of evidences and total peak intensities are given, ordered by peak intensity.

**Dataset A.3: ChIP-Seq and RNA-Seq data per gene.** The first six columns give the locus ID for both the PA14 and PAO1 strains, the gene name if available, and genomic locus of the gene. Columns seven and eight give the  $\log_2$ -transformed ratio between normalized ChIP-Seq counts for the HA-SutA immunoprecipitation versus the mock immunoprecipitation from the strain lacking HA-SutA, and the empirical Bayes F-test corrected p-value (FDR) indicating the probability that this ratio is not different from zero. Columns eight and nine give the  $\log_2$ -transformed

ratio between normalized ChIP-Seq counts for the RpoA immunoprecipitation from the strain containing HA-SutA versus the RpoA immunoprecipitation from the strain lacking HA-SutA, and the empirical Bayes F-test corrected p-value (FDR) indicating the probability that this ratio is not different from zero. Columns ten through thirteen give the average normalized (RPKM) ChIP-Seq counts per gene for each of the four immunoprecipitation samples. Columns fourteen and fifteen give the  $\log_2$ -transformed ratio between the normalized RNA-Seq counts in the *Para:sutA* strain versus the  $\Delta sutA$  strain, and the empirical Bayes F-test corrected p-value (FDR) indicating the probability that this ratio is not different from zero. Columns sixteen through nineteen give the RNA-Seq count ratios between the  $\Delta sutA$  strain and the wild-type strain and between the *Para:sutA* strain and the wild-type strain, and the empirical Bayes F-test corrected p-value (FDR) indicating the probability that there is no differential expression among the three strains. Columns twenty through twenty-two give the average RNA-Seq RPKM values for each of the three strains. Full raw data, and processed data for individual replicate samples, are available through the NCBI GEO repository (<http://www.ncbi.nlm.nih.gov/geo/>) accession GSE66181.

**Dataset A.4: Values for baseline and maximum regions used to scale ChIP data.** The RPKM values for the 25 transcriptional units that had the lowest RPKM values in the RpoA immunoprecipitation in the HA-SutA containing strain were used to estimate a baseline level for each ChIP sample. Additionally, the values from the top ten peak regions associated with protein coding genes (where a peak is defined as two consecutive 100bp tiles that fall in the top 100 100bp tiles in the genome) for each type of pulldown were used to estimate the maximum level for each ChIP sample. The baseline and peak regions that were chosen for this analysis are shown in this table. The dynamic range for the HA pulldown in the  $\Delta sutA$ /pMQ72 empty vector strain was significantly lower than those for the other pulldowns, as expected for a control pulldown in which no specific association occurs. For the purpose of comparing association patterns in this strain to association patterns in the  $\Delta sutA$ /pMQ72-HA-SutA strain, reads per 100 bp in the empty vector strain were scaled to the baseline and maximum values observed in the HA-SutA ChIP samples.

**Dataset A.5: ChIP-Seq data per 100bp region.** This table summarizes ChIP-Seq results by 100bp region. The first column gives the region number for the 100bp region. 100bp regions were numbered in order throughout the genome. Columns

two and three give the  $\log_2$ -transformed ratio between normalized ChIP-Seq counts for the HA-SutA immunoprecipitation versus the mock immunoprecipitation from the strain lacking HA-SutA, and the empirical Bayes F-test corrected p-value (FDR) indicating the probability that this ratio is not different from zero. Columns four and five give the  $\log_2$ -transformed ratio between normalized ChIP-Seq counts for the RpoA immunoprecipitation from the strain containing HA-SutA versus the RpoA immunoprecipitation from the strain lacking HA-SutA, and the empirical Bayes F-test corrected p-value (FDR) indicating the probability that this ratio is not different from zero. Columns six through nine give the average normalized (RPKM) ChIP-Seq counts per gene for each of the four immunoprecipitation samples. Columns ten through seventeen give the average scaled values for each immunoprecipitation following linear scaling to the baseline and maximum values described in Dataset A.4, and the standard deviations for the three biological replicates for each immunoprecipitation. Columns eighteen and nineteen give the mean difference between the scaled value for the RpoA pulldown in the strain lacking HA-SutA and the strain containing HA-SutA, and the uncorrected p-value indicating the probability that this difference is not zero. Column twenty indicates whether the 100bp region was included in our “high chip” subset, which satisfied criteria of having scaled ChIP values above a threshold of 0.20 for the HA-SutA ChIP and above 0.25 for the RpoA ChIP in either strain, plus having a statistically significant enrichment in the HA-SutA ChIP compared to the mock control. Full raw data, and processed data for individual replicate samples, are available through the NCBI GEO repository (<http://www.ncbi.nlm.nih.gov/geo/>) accession GSE66181.

**Dataset A.6: Transcription unit data.** Transcription unit (operon) predictions made by Wurtzel and colleagues were used in this study and are presented here for convenience. Additional information on transcription unit sizes, distances between transcription units, and transcription unit orientation compared to neighboring transcription units is also collected here.

## References

- (1) Shanks, R. M. Q.; Caiazza, N. C.; Hinsa, S. M.; Toutain, C. M.; O'Toole, G. A. *Appl. Environ. Microb.* **2006**, *72*, 5027–5036.
- (2) Choi, K.-H.; Schweizer, H. P. *Nat. Protoc.* **2006**, *1*, 153–161.
- (3) Gibson, D. G. In *Methods in Enzymology*; Elsevier: 2011; Vol. 498, pp 349–361.
- (4) Fong, J. C. N.; Karplus, K.; Schoolnik, G. K.; Yildiz, F. H. *J. Bacteriol.* **2006**, *188*, 1049–1059.
- (5) Pédelacq, J.-D.; Cabantous, S.; Tran, T.; Terwilliger, T. C.; Waldo, G. S. *Nat. Biotechnol.* **2006**, *24*, 79–88.
- (6) Link, A. J.; Vink, M. K. S.; Tirrell, D. A. *Nat. Protoc.* **2007**, *2*, 1879–1883.
- (7) Szychowski, J.; Mahdavi, A.; Hodas, J. J. L.; Bagert, J. D.; Ngo, J. T.; Landgraf, P.; Dieterich, D. C.; Schuman, E. M.; Tirrell, D. A. *J. Am. Chem. Soc.* **2010**, *132*, 18351–18360.
- (8) Hong, V.; Presolski, S. I.; Ma, C.; Finn, M. G. *Angew. Chem. Int. Ed.* **2009**, *121*, 10063–10067.
- (9) Rappsilber, J.; Mann, M.; Ishihama, Y. *Nat. Protoc.* **2007**, *2*, 1896–1906.
- (10) Boersema, P. J.; Raijmakers, R.; Lemeer, S.; Mohammed, S.; Heck, A. J. R. *Nat. Protoc.* **2009**, *4*, 484–494.
- (11) Kalli, A.; Hess, S. *Proteomics* **2012**, *12*, 21–31.
- (12) Cox, J.; Mann, M. *Nat. Biotechnol.* **2008**, *26*, 1367–1372.
- (13) Bagert, J. D.; Xie, Y. J.; Sweredoski, M. J.; Qi, Y.; Hess, S.; Schuman, E. M.; Tirrell, D. A. *Mol. Cell Proteomics* **2014**, *13*, 1352–1358.
- (14) Dietrich, L. E. P.; Okegbe, C.; Price-Whelan, A.; Sakhtah, H.; Hunter, R. C.; Newman, D. K. *J. Bacteriol.* **2013**, *195*, 1371–1380.
- (15) Müsken, M.; Di Fiore, S.; Dötsch, A.; Fischer, R.; Häussler, S. *Microbiology* **2010**, *156*, 431–441.
- (16) Dietrich, L. E. P.; Price-Whelan, A.; Petersen, A.; Whiteley, M.; Newman, D. K. *Mol. Microbiol.* **2006**, *61*, 1308–1321.
- (17) Gilbert, K. B.; Kim, T. H.; Gupta, R.; Greenberg, E. P.; Schuster, M. *Mol. Microbiol.* **2009**, *73*, 1072–1085.
- (18) Bolger, A. M.; Lohse, M.; Usadel, B. *Bioinformatics* **2014**, *30*, 2114–2120.
- (19) Langmead, B.; Trapnell, C.; Pop, M.; Salzberg, S. L. *Genome Biol.* **2009**, *10*, R25.
- (20) Fröhlich, K. S.; Vogel, J. *Curr. Opin. Microbiol.* **2009**, *12*, 674–682.

- (21) Delhomme, N.; Padioleau, I.; Furlong, E. E.; Steinmetz, L. M. *Bioinformatics* **2012**, *28*, 2532–2533.
- (22) Wurtzel, O.; Yoder-Himes, D. R.; Han, K.; Dandekar, A. A.; Edelheit, S.; Greenberg, E. P.; Sorek, R.; Lory, S. *PLoS Pathog.* **2012**, *8*, e1002945.
- (23) Law, C. W.; Chen, Y.; Shi, W.; Smyth, G. K. *Genome Biol.* **2014**, *15*, R29.
- (24) Mooney, R. A.; Davis, S. E.; Peters, J. M.; Rowland, J. L.; Ansari, A. Z.; Landick, R. *Mol. Cell* **2009**, *33*, 97–108.
- (25) Homann, O. R.; Johnson, A. D. *BMC Biology* **2010**, *8*, 49.
- (26) Winsor, G. L.; Lam, D. K. W.; Fleming, L.; Lo, R.; Whiteside, M. D.; Yu, N. Y.; Hancock, R. E. W.; Brinkman, F. S. L. *Nucleic Acids Res.* **2011**, *39*, D596–D600.
- (27) Tatusov, R. L. et al. *BMC Bioinformatics* **2003**, *4*, 41.
- (28) Hunter, J. D. *Comput. Sci. Eng.* **2007**, *9*, 90–95.
- (29) Schneider, C. A.; Rasband, W. S.; Eliceiri, K. W. *Nat. Methods* **2012**, *9*, 671–675.
- (30) Liberati, N. T.; Urbach, J. M.; Miyata, S.; Lee, D. G.; Drenkard, E.; Wu, G.; Villanueva, J.; Wei, T.; Ausubel, F. M. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 2833–2838.

*Appendix B*

## OTHER CONTRIBUTIONS

**B.1 Contributions to Ngo, et al.**

Published as:

- (1) Ngo, J. T.; Babin, B. M.; Champion, J. A.; Schuman, E. M.; Tirrell, D. A. *ACS Chem. Biol.* **2012**, *7*, 1326–1330.

**Abstract**

Transcriptional activity from a specified promoter can provide a useful marker for the physiological state of a cell. Here we introduce a method for selective tagging of proteins made in cells in which specified promoters are active. Tagged proteins can be modified with affinity reagents for enrichment or with fluorescent dyes for visualization. The method allows state-selective analysis of the proteome, whereby proteins synthesized in predetermined physiological states can be identified. The approach is demonstrated by proteome-wide labeling of bacterial proteins upon activation of the  $P_{BAD}$  promoter and the SoxRS regulon and provides a basis for analysis of more complex systems including spatially heterogeneous microbial cultures and biofilms.

**Contributions**

To evaluate our ability to label cells in response to oxidative stress, I treated *E. coli* engineered to express NLL-MetRS under control of the *soxS* promoter with paraquat and Anl. I reacted labeled cells with akyne-TAMRA and quantified the response by measuring Anl incorporation via fluorescent imaging. The figures below correspond to Figure 3D, Supplementary Figure 2, and Supplementary Figure 3 in the publication. I also contributed to writing the manuscript.

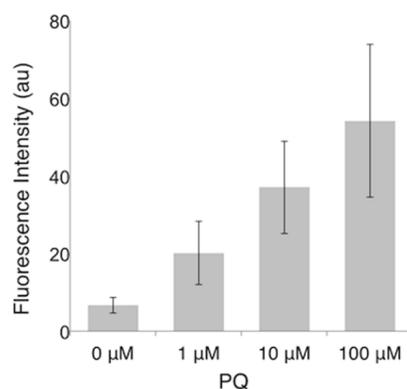


Figure B.1: **Proteomic labeling with Anl under conditions of oxidative stress.** The degree of tagging is sensitive to the level of oxidative stress induction by paraquat. Cells were treated with various concentrations of PQ and pulsed with a fixed concentration of Anl (125 μM). The extent of tagging was assessed by conjugation to alkyne-TAMRA and measurement of individual cell intensities by fluorescence microscopy. Error bars represent the standard deviation of each population examined.

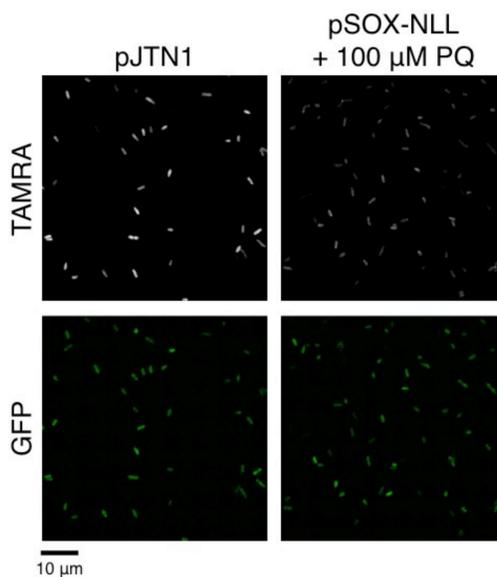
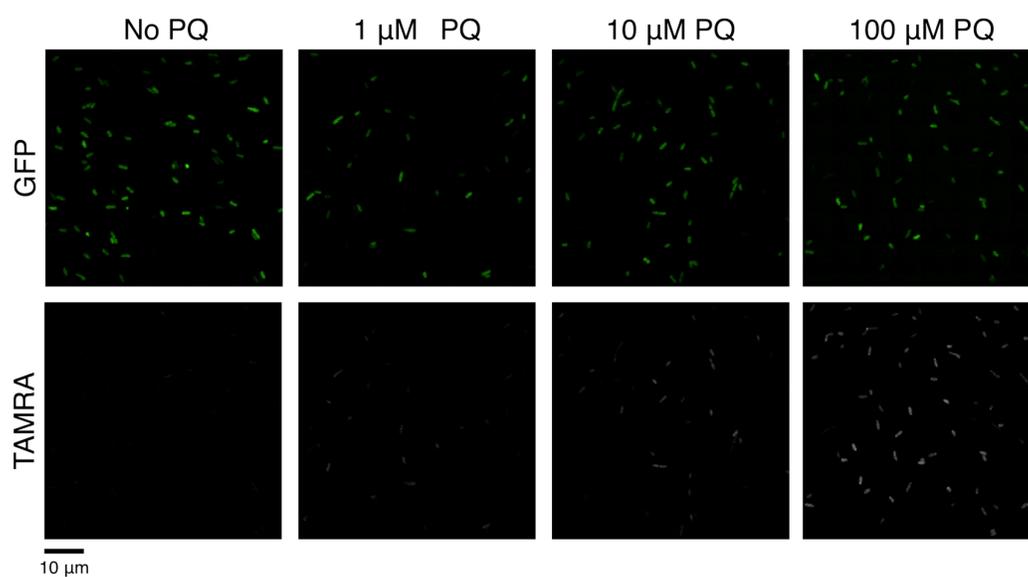


Figure B.2: **Tagging rate in the SoxRS system is less than 10%** *E. coli* cells harboring the pJTN1 plasmid pulsed with Anl under conditions that yield a 10% substitution rate are compared with *E. coli* harboring pSOX-NLL induced with paraquat and pulsed with Anl. Incorporation of Anl is assessed by conjugation with alkyne-TAMRA and subsequent detection by fluorescence microscopy. SoxRS-directed labeling with Anl yields a substitution rate of less than 10%, as fluorescence emission from pJTN1 cells is more intense than that observed with SoxRS-directed labeling. Cells constitutively expressed GFP, which was separately detected to confirm the presence of cells.



**Figure B.3: Tagging rate in the SoxRS system is dependent on the degree of induction of NLL-MetRS expression.** The NLL-MetRS is under control of the *soxS* promoter and is activated by addition of paraquat (PQ) to the culture medium. As the degree of transcription from the *soxS* promoter is dependent on the concentration of PQ used, so is the level of NLL-MetRS induction. Cells induced with PQ and pulsed with 125  $\mu$ M AnI exhibit increasing levels of AnI incorporation as more PQ is added. Incorporation of AnI is assessed by conjugation to alkyne-TAMRA and subsequent detection by fluorescence microscopy. Cells constitutively expressed GFP, which was separately detected to confirm the presence of cells.

## B.2 Contributions to Hatzenpichler, et al.

Published as:

- (1) Hatzenpichler, R.; Scheller, S.; Tavormina, P. L.; Babin, B. M.; Tirrell, D. A.; Orphan, V. J. *Environ. Microbiol.* **2014**, *16*, 2568–2590.

### Abstract

Here we describe the application of a new click chemistry method for fluorescent tracking of protein synthesis in individual microorganisms within environmental samples. This technique, termed bioorthogonal non-canonical amino acid tagging (BONCAT), is based on the *in vivo* incorporation of the non-canonical amino acid L-azidohomoalanine (AHA), a surrogate for L-methionine, followed by fluorescent labelling of AHA-containing cellular proteins by azide-alkyne click chemistry. BONCAT was evaluated with a range of phylogenetically and physiologically diverse archaeal and bacterial pure cultures and enrichments, and used to visualize translationally active cells within complex environmental samples including an oral biofilm, freshwater and anoxic sediment. We also developed combined assays that couple BONCAT with ribosomal RNA (rRNA)-targeted fluorescence in situ hybridization (FISH), enabling a direct link between taxonomic identity and translational activity. Using a methanotrophic enrichment culture incubated under different conditions, we demonstrate the potential of BONCAT-FISH to study microbial physiology *in situ*. A direct comparison of anabolic activity using BONCAT and stable isotope labelling by nano-scale secondary ion mass spectrometry ( $^{15}\text{NH}_3$  assimilation) for individual cells within a sediment-sourced enrichment culture showed concordance between AHA-positive cells and  $^{15}\text{N}$  enrichment. BONCAT-FISH offers a fast, inexpensive and straightforward fluorescence microscopy method for studying the *in situ* activity of environmental microbes on a single-cell level.

### Contributions

I contributed to the chemical synthesis of Aha; provided advice on Aha labeling experiments, fluorescence gel measurements, and cell imaging; and contributed to writing the manuscript.