Cell-Selective Proteomic Profiling in Complex Biological Systems

by

Xinran Liu

In Partial Fulfillment of the Requirements for

the Degree of

Doctor of Philosophy in Chemical Engineering

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2020

(Defended May 18th, 2020)

Xinran Liu

ORCID: 0000-0003-0045-8823

ACKNOWLEDGEMENTS

Frist, I would like to thank my advisor, Prof. David Tirrell, for his continuous support and guidance throughout my whole doctoral journey. Dave's immense knowledge, patience, and encouragement, as well as the intellectual freedom in his lab, allow me to learn from initial project conception all the way to project completion and foster me to grow as an independent research scientist.

I would also like to thank my committee members, Prof. Julia Kornfield, Prof. Dianne Newman, and Prof. Rustem Ismagilov, for their insightful advice and help, and also for their hard questions which challenged and incented me to widen my research.

Many thanks to Dr. Brett Babin, Prof. Megan Bergkessel, and Prof. William DePas for teaching me everything important in microbiology. I could not have accomplished my projects without their help.

I greatly enjoyed working with Yue Hui, my talented office mate and project mate. Yue and I worked closely in the project discussed in Chapter 3. Her creativity and technical expertise made the collaboration very pleasant. Thank you, Yue.

Thanks to the team of Biological Imaging Facility, for training and helping me with confocal microscopy imaging. Many thanks to the talented and wonderful team of the Proteome

Exploration Laboratory: Dr. Annie Moradian, Dr. Michael Sweredoski, Dr. Roxana Eggleston-Rangel, and Dr. Brett Lomenick. I enjoyed all the enlightening discussions I had with Annie and Mike, both about science and life. Thanks to Roxana and Brett, for their professional assistance with mass spectrometry.

Thanks to all my colleagues and my friends, for making my time at Caltech fun and memorable.

Lastly, thank you to my family, for their love and support.

ABSTRACT

Cells within biological systems are constantly adjusting their protein synthesis in response to various environmental changes. To study the rapid cellular regulations in complex biological systems, global proteomic profiling provides important information on system-level regulations, yet physiological properties characteristic of individual cellular subpopulations could be hidden under the characterization. Instead, cell-selective proteomic profiling allows researchers to reveal the heterogeneities in biological systems with phenotypically and even genetically distinct subpopulations under different microenvironments.

Chapter 1 describes the development of bioorthogonal noncanonical amino acid tagging (BONCAT) for proteomic profiling with resolution in both space and time: its initial role is protein labeling with temporal resolution via pulse-addition of noncanonical amino acid, which could be recognized by endogenous aminoacyl tRNA-synthetase (aaRS), into systems of interest; later on, mutant aaRSs are identified through mutant synthetase library screening, which allows for efficient incorporation of various types of noncanonical amino acids that could hardly be activated by endogenous machineries. The identification and exploitation of mutant aaRSs allow sensitive cellular selectivity during protein labeling. With unprecedented spatiotemporal resolution of BONCAT, and the advancement in high-resolution mass spectrometry and computational algorithms, BONCAT is a powerful technique for selective proteomic profiling to study physiological regulations in a wide range of complex biological systems.

Chapter 2 describes the application of the BONCAT method in cell-selective proteomic profiling in *Pseudomonas aeruginosa* biofilms. In this work, we targeted an iron-starved subpopulation in biofilms and compared its proteomic profile with that of the entire system. Key gene and pathway regulations in the subpopulation are found through the analysis of the proteomic data, which suggest that iron-starved cells shift their priority towards housing keeping pathways, adapt an energy- and resources-saving mode to cope with their harsh local environmental conditions, and get prepared to disperse for better survival. Analysis of poorly studied proteins highly upregulated in the subpopulation led to the discovery of a previously uncharacterized protein (PA14_52000) that is potentially related to iron acquisition. The transposon insertion mutant *PA14_52000::tn* showed significantly enhanced pyoverdine production in rich medium and reduced biofilm formation.

Chapter 3 describes the study of physiological regulations in *Bacillus subtilis* K-state subpopulation via BONCAT. A subset of *B. subtilis* cells, typically 10% - 20% of the entire population, enter K-state in a stochastic manner. With the low level of K-state entry rate and high randomness, we challenged BONCAT to specifically capture gene and pathway regulations in K-state cells and compared the proteomic profiling with that of the entire population. Regardless of the difficulties of selective protein labeling inherent in the system, our results indicate that BONCAT has high specificity and resolution in proteomic profiling for minor subpopulations and proteins with low overall absolute abundance. We found multiple pathways and genes characteristic of K-state regulated differentially from the entire population, either significantly up- or down-regulated. Proteins that are uncharacterized or previously

known for functions irrelevant of K-state are highly abundant in the subpopulation, providing new insight toward their alternative functions critical for K-state cells and future investigation directions of K-state study.

TABLE OF CONTENTS

Acknowledgements iii
Abstract v
Table of Contents viii
List of Figures xi
List of Tables xiii
Nomenclature xiv

Chapter 1: Cell-Selective Proteomic Profiling 1
1.1 Introduction 1
1.2 Proteomic Profiling with Spatiotemporal Resolution 2
1.2.1 Bioorthogonal Noncanonical Amino Acid Tagging 2
1.2.2 Spatiotemporal Resolution Achieved via BONCAT 3
1.3 Conclusions 5
1.4 References 6
1.5 Supplementary Figures 8

Chapter 2: Selective Proteomic Analysis of Iron-Starved Cellular Subpopulations in *Pseudomonas aeruginosa* Biofilms 9
2.1 Abstract 9
2.2 Introduction 10
2.3 Results 13
2.3.1 The endogenous *pvdF* promoter controls cell-state-selective labeling under iron-limited conditions 13
2.3.2 Visualization of labeled biofilms in flow cells 17
2.3.3 Selective proteomic profiling of iron-starved biofilm subpopulations via BONCAT 17
2.3.4 Gene and pathway regulation in the iron-starved subpopulation 21
2.3.5 Altered expression of proteins in the iron-starved subpopulation 26
2.3.6 The iron-starved subpopulation of *P. aeruginosa* biofilm is under a unique microenvironment 31

2.3.7 Uncharacterized protein PA14_52000 enriched in the iron-starved subpopulation is iron-related ····· 34
2.4 Discussion ····· 36
2.5 Materials and Methods ····· 37
2.5.1 Bacterial strains and growth conditions ····· 37
2.5.2 Biofilm imaging ····· 39
2.5.3 BONCAT labeling and enrichment ····· 40
2.5.4 LC-MS/MS ····· 42
2.5.5 Proteomic data analysis ····· 43
2.5.6 Software and database used ····· 44
2.6 Acknowledgements ····· 45
2.7 References ····· 46
2.8 Supplementary Figures ····· 54
2.9 Supplementary Tables ····· 64

Chapter 3: Bioorthogonal Noncanonical Amino Acid Tagging (BONCAT) Enables Cell-Selective Proteomic Labeling and Analysis of *Bacillus subtilis* K-State Subpopulation ····· 66
3.1 Summary of Contributions ····· 66
3.2 Abstract ····· 67
3.3 Introduction ····· 68
3.4 Results ····· 69
3.4.1 P_{comF}:*nll-metRS* enables K-state-selective labeling in a heterogeneous population ····· 69
3.4.2 Visualization of K-state labeling ····· 71
3.4.3 Proteomic profiling in *B. subtilis* K-state subpopulation via BONCAT ····· 73
3.4.4 Pathway regulations in *B. subtilis* subpopulation ····· 75
3.4.5 Altered expression in genes and gene clusters with shared physiological functions ····· 81
3.5 Discussion ····· 84
3.6 Materials and Methods ····· 87
3.6.1 Strain construction ····· 87
3.6.2 K-state induction in *B. subtilis* ····· 88

3.6.3 Transformation assay ····· 88
3.6.4 Confocal imaging ····· 89
3.6.5 BONCAT labeling and enrichment ····· 90
3.6.6 LC-MS/MS ····· 92
3.6.7 Proteomic data analysis ····· 93
3.6.8 Software and database used ····· 94
3.7 Acknowledgements ····· 94
3.8 References ····· 95
3.9 Supplementary Figures ····· 102
3.10 Supplementary Tables ····· 113

Appendix A: Supplementary Information for Chapter 2 ····· 114
Supplementary Datasets ····· 114

Appendix B: Supplementary Information for Chapter 3 ····· 118
Supplementary Datasets ····· 118

LIST OF FIGURES

1.1 Mechanism of bioorthogonal noncanonical amino tagging ····· 4
2.1 Cell-state-selective labeling of planktonic *P. aeruginosa* cells ····· 15
2.2 Visualization of biofilm labeling in flow cells ····· 18
2.3 Cell-state-selective labeling of *P. aeruginosa* biofilms ····· 20
2.4 Regulation of cellular processes in *P. aeruginosa* biofilms ····· 22
2.5 Differentially regulated pathways in the biofilm subpopulation ····· 32
3.1 K-state-selective proteomic labeling of P_{comF}*:nll-metRS* strain ····· 72
3.2 Selective proteomic labeling of *B. subtilis* K-state subpopulations ····· 75
3.3 Pathway regulations in K-state subpopulations ····· 77

1.S1 Structures of amino acids discussed in the chapter ····· 8
2.S1 Structures of reagents used in the study ····· 54
2.S2 Labeling of planktonic P_{pvdF}*:nll-metRS* cells ····· 55
2.S3 Labeling of planktonic P_{trc}*:nll-metRS* cells ····· 56
2.S4 Growth assays of the wild type, P_{trc}*:nll-metRS*, and P_{pvdF}*:nll-metRS* strains ····· 57
2.S5 Biofilm formation assays of the wild type, P_{trc}*:nll-metRS*, and P_{pvdF}*:nll-metRS* strains ····· 58
2.S6 Individual and merged images of labeled P_{pvdF}*:nll-metRS* biofilms ····· 59
2.S7 Individual and merged images of labeled P_{trc}*:nll-metRS* biofilms ····· 60
2.S8 Individual and merged images of labeled wild-type biofilms ····· 61
2.S9 Comparison of proteomic profiles obtained in replicates of the biofilm labeling experiment ····· 62
2.S10 Characterization of the transposon insertion mutant *PA14_52000::tn* ····· 63
3.S1 The W168 strain incorporates Anl into protein synthesis using NLL-MetRS ····· 102
3.S2 Optimization of K-state inducing conditions ····· 103
3.S3 Growth patterns of P_{comF}*:nll-metRS* strain ····· 104
3.S4 Selective proteomic labeling of *B. subtilis* K-state subpopulation with the strain P_{comF}*: nll-metRS* ····· 105
3.S5 Proteomic labeling of *B. subtilis* whole population with P_{veg}*: nll-metRS* before and after t_0 ····· 106

3.S6 Growth curves of W168, P_{comF}*:nll-metRS*, and P_{veg}*:nll-metRS* ····· 107
3.S7 Characterization of the effect of NLL-MetRS expression on ComK expression and competence induction ····· 108
3.S8 Visualization of P_{comF} activity over time ····· 109
3.S9 Visualization of ComK-mWasabi expression and Anl-labeling in W168 *comK-mWasabi* and P_{veg}*:nll-metRS_comK-mWasabi* strains ····· 110
3.S10 Structure of TAMRA-DBCO ····· 111
3.S11 Amino acid metabolism and degradation regulations in K-state Subpopulation ····· 112

LIST OF TABLES

2.1 Gene Ontology enrichment analysis ····· 16
2.2 Gene Ontology enrichment analysis of proteins upregulated or found exclusively in P_{pvdF}:*nll-metRS* samples ····· 28
2.3 Proteins significantly upregulated in both the iron-Starved subpopulation and the RpoS-elevated subpopulation ····· 33
2.4 Proteins significantly upregulated in both the iron-Starved subpopulation and the entire *P. aeruginosa* biofilm under anaerobic growth condition ····· 34
3.1 Summary of genes with shared cellular functions that are differentially regulated in K-state subpopulation ····· 82

2.S1 Physico-chemical properties of PA14_52000 calculated by ExPASy ProtParam ····· 64
2.S2 Strains and primers used in the study ····· 65
3.S1 Strains and primers used in this study ····· 113

NOMENCLATURE

aaRS. Aminoacyl-tRNA synthetase

ACN. Acetonitrile

Aha. Azidohomoalanine

AmmBic. Ammonium bicarbonate

Anl. Azidonorleucine

Aoa. 2-aminooctynoic acid

*att*Tn7. Tn7 attachment site

Azf. *p*-azidophenylalanine

BCAA. Branched-chain amino acid

BCFA. Branched-chain fatty acid

BONCAT. Bioorthogonal noncanonical amino acid tagging

CDS. Coding sequence

Cm. Chloramphenicol

Cm^R. Chloramphenicol resistance gene

CuAAC. Copper catalyzed alkyne-azide cycloaddition

DBCO. Aza-dibenzocyclooctyne

DMSO. Dimethyl sulfoxide

DTT. Dithiothreitol

Erm. Erythromycin

Erm^R. Erythromycin resistance gene

FDR. False discovery rate

Gm^R. Gentamicin resistance gene

GO. Gene Ontology

Hpg. Homopropargylglycine

HPLC. High performance liquid chromatography

iBAQ. Intensity based absolute quantification

KEGG. Kyoto Encyclopedia of Genes and Genomes

LB. Luria-Bertani broth

LC. Liquid chromatography

LCM. Laser capture microdissection

LFQ. Label-free quantitation

Limma. Linear models for microarray data

LTQ. Linear trap quadrupole

MetRS. Methionyl-tRNA synthetase

MS. Mass spectrometry

ncAA. Noncanonical amino acid

PBS. Phosphate-buffered saline

PCA. Principal coordinate analysis

PCR. Polymerase chain reaction

Phe. Phenylalanine

PheRS. Phenylalanyl-tRNA synthetase

Pra. Propargylglycine

SDS. Sodium dodecyl sulfate

SDS-PAGE. Sodium dodecyl sulfate–polyacrylamide gel electrophoresis

SM1. Starvation medium 1

SM2. Starvation medium 2

TA. Teichoic acid

TAMRA. Tetramethylrhodamine

THPTA. Tris(3-hydroxypropyltriazolylmethyl)amine

TSB. Tryptic soy broth

Chapter 1

CELL-SELECTIVE PROTEOMIC PROFILING

1.1 Introduction

To investigate physiological regulations of complex biological systems, mRNA profiling (transcriptomics) is often employed, which mainly relies on microarray and RNA-seq techniques (1-3). System level transcriptome profiling provides exceptional new insights into gene expression of a wide range of biological systems. However, because limited information is revealed on the behavior of individual cellular subpopulations in global profiling studies, various cell-selective transcriptome profiling techniques are introduced, leading to much higher spatial resolution in the analysis of gene expression (4).

Proteomic profiling, on the other hand, was not as popular in studies of physiological regulations until recently, with advances in mass spectrometry, computational algorithms, and data analysis. Because variance between mRNA and protein levels in biological systems is relatively high, which can be largely explained by post-translational regulations (5, 6), high-resolution proteomic profiling of complex biological samples has drawn increasing attention from investigators, providing a new, promising way to study physiological regulations of complex samples via mass protein identification and quantification.

In this chapter, we introduce a recently developed proteomic profiling technique that could be adapted to achieve spatiotemporal resolution.

1.2 Proteomic profiling with spatiotemporal resolution

1.2.1 Bioorthogonal noncanonical amino acid tagging

Bioorthogonal noncanonical amino acid tagging (BONCAT), introduced by Dieterich and coworkers, is initially employed to label proteins in systems of interest via incorporation of noncanonical amino acids (ncAAs) into proteins by endogenous aminoacyl-tRNA synthetases (aaRSs) and the cognate tRNAs (7, 8) (Figure 1.1A). Because the ncAAs bear a bioorthogonal functional group that can be conjugated to an affinity tag or solid support, proteins tagged by the ncAA can be separated from their unlabeled counterparts, and subsequently identified and quantified by liquid chromatography-tandem mass spectrometry (LC-MS/MS) (Figure 1.1B). The methionine surrogates azidohomoalanine (Aha, Figure 1.S1A) and homopropargylglycine (Hpg, Figure 1.S1B) have been successfully used to explore the proteomic profiles in various biological systems (9-12). Because the sensitivity of LC-MS/MS is limited by sample complexity, BONCAT enrichment can facilitate the detection of proteins of interest (13).

1.2.2 Spatiotemporal resolution achieved via BONCAT

Cell selectivity of BONCAT could be achieved by using ncAA probes that are activated slowly or not at all by any of the endogenous aaRSs. In 2009, Ngo and coworkers identified a mutant methionyl-tRNA synthetase (NLL-MetRS) adapted from *Escherichia coli* that efficiently

charges a methionine (Met, Figure 1.S1C) surrogate azidonorleucine (Anl, Figure 1.S1D) to the cognate tRNA (Figure 1.1A) (14). Because wild-type cells without NLL-MetRS could not use Anl, protein labeling is restricted to those engineered to express the mutant synthetase. In addition to Anl, several other mutant aaRSs have been developed for the methionine surrogates 2-aminooctynoic acid (Aoa, Figure 1.S1E) (15) and propargylglycine (Pra, Figure 1.S1F) (16), and for the phenylalanine (Phe, Figure 1.S1G) surrogate *p*-azidophenylalanine (Azf, Figure 1.S1H) (17). With the mutant aaRSs and ncAAs, cell-selective BONCAT has been successfully used in many co-culture systems (14, 15, 18). To achieve even stricter cell-selectivity to label cellular subpopulations within a single strain, cell type- or cell state-specific promoters should be carefully selected with negligible leaky expression levels to control the expression of mutant aaRSs. In 2015, Kai and coworkers used the *myo-2* promoter of *Caenorhabditis elegans* to control a mutant phenylalanyl-tRNA synthetase PheRS, and achieved selective protein labeling restricted to the 20 pharyngeal muscle cells of *C. elegans* (17). BONCAT also provides excellent temporal resolution in proteomic profiling through control of the conditions used to deliver the ncAA label. Labeling times as short as 10 min have been used to interrogate bacterial quorum sensing (11, 12), which can be extended to hours for cells with low rates of protein synthesis (19).

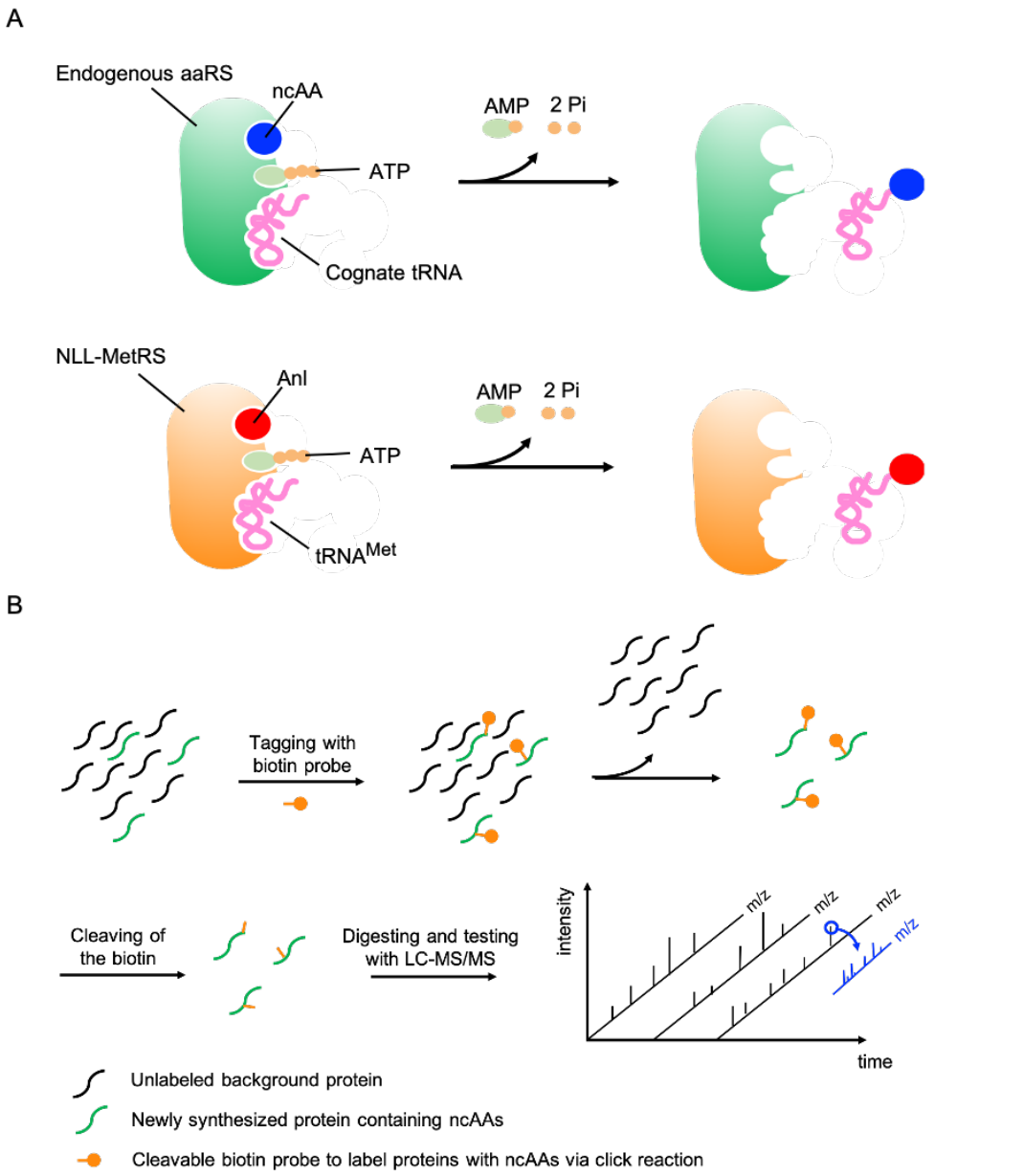

Figure 1.1 Mechanism of bioorthogonal noncanonical amino acid tagging. (A) Schematic diagram for the charge of ncAA to its cognate tRNA by endogenous aaRS and the charge of Anl to the cognate tRNA for methionine ($tRNA^{Met}$) by NLL-MetRS. (B) Schematic diagram for the enrichment and purification of labeled proteins via biotin probe, and identification and analysis with LC-MS/MS.

1.3 Conclusions

Recently, with the advancements in high-resolution LC-MS/MS techniques, selective proteomic analysis via metabolic labeling, such as BONCAT, is gradually becoming a key new study strategy to examine physiological regulations in complex biological systems with spatiotemporal resolution. It helps reveal new mechanisms governing and special roles within physiologically differentiated cells and subpopulations to promote integrity and overall health for entire systems in response to a wide range of environmental conditions.

1.4 References

1. Gresham D, Dunham MJ, Botstein D (2008) Comparing whole genomes using DNA microarrays. *Nat Rev Genet 9*(4):291-302.
2. Heller MJ (2002) DNA microarray technology: devices, systems, and applications. *Annu Rev Biomed Eng* 4(1):129-153.
3. Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 10(1):57-63.
4. Stark R, Grzelak M, Hadfield J (2019) RNA sequencing: the teenage years. *Nat Revs Genet* 20(11):631-656.
5. Vogel C, Marcotte EM (2012) Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet* 13(4):227-232.
6. de Sousa Abreu R, Penalva LO, Marcotte EM, Vogel C (2009) Global signatures of protein and mRNA expression levels. *Mol BioSyst 5*(12):1512-1526.
7. Dieterich DC, Link AJ, Graumann J, Tirrell DA, Schuman EM (2006) Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT). *Proc Natl Acad Sci USA* 103(25):9482-9487.
8. Dieterich DC, Lee JJ, Link AJ, Graumann J, Tirrell DA, Schuman EM (2007) Labeling, detection and identification of newly synthesized proteomes with bioorthogonal non-canonical amino-acid tagging. *Nat Protoc*, *2*(3), 532.
9. Kramer G, et al. (2009) Identification and quantitation of newly synthesized proteins in Escherichia coli by enrichment of azidohomoalanine-labeled peptides with diagonal chromatography. *Mol Cell Proteomics* 8(7):1599-1611.
10. Howden AJ, et al. (2013). QuaNCAT: quantitating proteome dynamics in primary cells. *Nat Methods* 10(4):343-346.
11. Feng L, et al. (2015) A Qrr noncoding RNA deploys four different regulatory mechanisms to optimize quorum-sensing dynamics. *Cell* 160(1-2):228–240.
12. Bagert JD, et al. (2016) Time-resolved proteomic analysis of quorum sensing in Vibrio harveyi. *Chem Sci* 7(3):1797–1806.

13. Delahunty C, Yates III JR (2005) Protein identification using 2D-LCMS/MS. *Methods* 35(3):248–255.
14. Ngo JT, et al. (2009). Cell-selective metabolic labeling of proteins. *Nat Chem Biol 5*(10):715-717.
15. Grammel M, Zhang MZM, Hang HC (2010) Orthogonal Alkynyl Amino Acid Reporter for Selective Labeling of Bacterial Proteomes during Infection. *Angew Chem* 49:5970-5974.
16. Truong F, Yoo TH, Lampo TJ, Tirrell DA (2012) Two-strain, cell-selective protein labeling in mixed bacterial cultures. *J Am Chem Soc* 134(20):8551–8556.
17. Yuet KP, et al. (2015). Cell-specific proteomic analysis in Caenorhabditis elegans. *Proc Natl Acad Sci USA* 112(9):2705-2710.
18. Mahdavi A, et al. (2014) Identification of secreted bacterial proteins by noncanonical amino acid tagging. *Proc Natl Acad Sci USA* 111:433-438.
19. Babin BM, et al. (2017) Selective proteomic analysis of antibiotic-tolerant cellular subpopulations in Pseudomonas aeruginosa biofilms. *mBio* 8(5):e01593-17.

1.5 Supplementary Figures

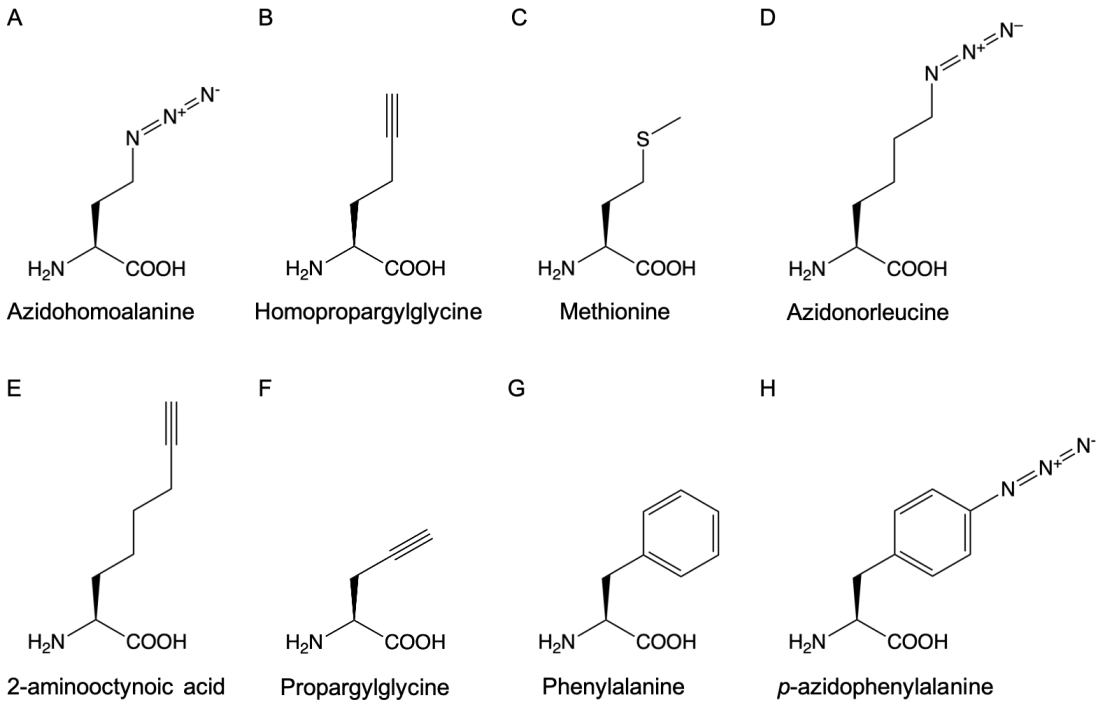

Figure 1.S1 Structures of amino acids discussed in the chapter.

(A) Azidohomoalanine (Aha). (B) Homopropargylglycine (Hpg). (C) Methionine (Met). (D) Azidonorleucine (Anl). (E) 2-aminooctynoic acid (Aoa). (F) Propargylglycine (Pra). (G) Phenylalanine (Phe). (H) *p*-azidophenylalanine (Azf).

Chapter 2

SELECTIVE PROTEOMIC ANALYSIS OF IRON-STARVED CELLULAR SUBPOPULATIONS IN *PSEUDOMONAS AERUGINOSA* BIOFILMS

2.1 Abstract

Bacterial biofilms are complex systems with phenotypically and even genetically distinct cells localized in chemically heterogeneous environments. The complexity of such systems impedes the study of genetic regulation and protein synthesis in cellular subpopulations in specific physiological states. Here we introduce bioorthogonal noncanonical amino acid tagging (BONCAT) as a method for the study of protein synthesis in iron-starved subpopulations in *Pseudomonas aeruginosa* biofilms. We engineered a strain of *P. aeruginosa* (designated P_{pvdF}:*nll-metRS*) to express a mutant methionyl-tRNA synthetase (NLL-MetRS) under control of an endogenous iron-responsive promoter element of the gene pyoverdine synthetase F (*pvdF*) so that newly synthesized proteins in iron-starved cells could be labeled upon addition of the methionine surrogate azidonorleucine (Anl). We labeled four-day old P_{pvdF}:*nll-metRS* biofilms with Anl, and enriched and quantified the labeled proteins. In comparison with entire biofilm, the subpopulation is characterized by an adaptive strategy that shifts metabolic priority toward housekeeping pathways, including ribosomal protein biosynthesis and assembly, oxidative phosphorylation, and nitrogen metabolism. We also found evidence for upregulation of

transporters for ion exchange and nutrient uptake, and for downregulation of xenobiotic metabolism and biofilm formation pathways. Finally, we found a previously uncharacterized gene (*PA14_52000*) to be upregulated in P_{pvdF}*:nll-metRS* biofilms and potentially related to iron acquisition. The transposon insertion mutant *PA14_52000::tn* showed significantly enhanced pyoverdine production in rich medium and reduced biofilm formation.

2.2 Introduction

The gram-negative bacterium *Pseudomonas aeruginosa* is an opportunistic pathogen that can cause severe chronic infections, especially in patients with cystic fibrosis or traumatic burns (1). As one of the most common pathogens found in nosocomial infections, *P. aeruginosa* presents challenges of vital clinical importance (2, 3). Complete eradication of *P. aeruginosa* is thwarted by its antibiotic resistance and its capacity to form complex biofilm structures, in which bacterial cells attach to surfaces and adhere to one another within a matrix of secreted extracellular polymeric substances (EPS) (1, 4). Cells within biofilms often exhibit physiological properties distinct from those of their planktonic counterparts (5). Moreover, biofilms are characterized by substantial phenotypic heterogeneity, mainly a consequence of concentration gradients of nutrients and metabolites, including signaling compounds and waste products (4, 5). As cells respond to their local environments, gene expression varies and physiologically distinct cells emerge, contributing to increased metabolic complexity and survival (4, 5).

Substantial effort has been devoted to characterization of the physiological properties of *P. aeruginosa* biofilms. Many studies have been done to analyze the global expression profile of the biofilms through the identification and quantification of transcriptomes using microarrays and RNAseq (6-10). However, global expression profiles provide only averaged information, while the cells that constitute the film are in fact in many distinct physiological states and experience different local environments. To address this issue, investigators have used microfluidics and advanced imaging technologies to explore the role of spatial heterogeneity in biofilm development (11-13). To study the transcriptomes of localized subpopulations, laser capture microdissection (LCM) has been employed to isolate biofilm cells and transcriptomic analysis via RNA extraction and quantitative real-time reverse transcription-PCR (qRT-PCR) or via microarray is often accompanied for gene expression studies (14, 15). This approach has been employed to show that the relative abundance of RNA transcripts, such as mRNAs for genes related to housekeeping, quorum sensing, and dividing varied between cells at the top and those at the bottom of *P. aeruginosa* biofilms (14, 15). However, cell-selective response to a specific stress such as a nutrient limitation or an antibiotic treatment has not been explored using LCM.

Iron is essential for bacterial life (16). In *P. aeruginosa*, iron fosters biofilm formation, and iron limitation stimulates the twitching motility of cells on surfaces, compromising their ability to form biofilms (17, 18). Iron deprivation, achieved either by sequestering iron via chelators or by blocking the major iron acquisition systems, causes formation of thin, flat biofilms rather than the more typical mushroom-like biofilm architectures (19). But local iron concentrations

in biofilms can be reduced by host defense mechanisms that employ iron sequestration, and by slow diffusion of iron in the biofilm matrix (16, 19, 20). *P. aeruginosa* has developed multiple iron uptake systems to acquire the iron needed for growth and infection, including the high-affinity iron acquisition system that transports extracellular iron into cells in complex with the siderophore pyoverdine (19, 21). At low iron concentrations, ferric uptake regulator (Fur), a major repressor of iron-responsive genes in *P. aeruginosa*, is released from the promoter region of an iron starvation sigma factor gene *pvdS*, which in turn initiates the transcription of multiple genes required for pyoverdine synthesis (22, 23).

Here we sought to exploit both the temporal resolution and the cell-state selectivity of the BONCAT method to interrogate iron-starved subpopulations of cells in *P. aeruginosa* biofilms. We engineered cells to express a mutant methionyl-tRNA synthetase (NLL-MetRS), derived from *Escherichia coli* by library screening, which can efficiently attach Anl to the cognate tRNA (24, 25). We chose the promoter of the gene pyoverdine synthetase F (*pvdF*) for expression control, restricting protein labeling to the iron-starved biofilm cells. We analyzed the proteomic profile of the subpopulation and compared it with the proteomic profile of the entire biofilm population.

2.3 Results

2.3.1 The endogenous pvdF promoter controls cell-state-selective labeling under iron-limited conditions

To selectively label the iron-starved subpopulation of cells in *P. aeruginosa* biofilms, we sought to control expression of the NLL-MetRS with a promoter that is active only under low-iron conditions. Because a single NLL-MetRS synthetase could lead to multiple Anl-labeled proteins, cell state-specific promoters should be carefully selected with negligible basal expression levels to achieve strict cell selectivity. After thorough screening of iron- and PvdS-regulated genes, we chose the promoter of the gene pyoverdine synthetase F (*pvdF*) which catalyzes the formylation of N^5-hydroxyornithine residues required for pyoverdine biosynthesis in *P. aeruginosa*, to regulate *nll-metRS* (26-28). The promoter region of *pvdF* contains an iron starvation box for binding of PvdS and expected to be activated only under conditions of iron limitation (27).

To verify that P_{pvdF} could be used to control expression of the NLL-MetRS in a state-selective manner, we first tested its activity in planktonic cells. We fused the 1.5 kb region upstream of the endogenous *pvdF* coding sequence (CDS) to the 5'-end of *nll-metRS*, and a hexahistidine tag (His-tag) to the 3'-end to facilitate Western blotting. This gene cassette was transposed into the wild-type *P. aeruginosa* PA14 genome at the Tn7 attachment (*att*Tn7) site through four-parental mating conjugation to generate the strain P_{pvdF}*:nll-metRS* (Figure 2.1A) (29). P_{pvdF}*:nll-metRS* cells were grown in 1% TSB plus 1 mM $MgCl_2$ in the presence or absence of 50 μM

$FeCl_3$ to late-exponential phase and treated with 1 mM Anl for 1 h. We used the copper-catalyzed alkyne-azide cycloaddition (CuAAC) to detect incorporation of Anl into cellular proteins. Cell lysates were treated with TAMRA-alkyne (Figure 2.S1A), separated by SDS-PAGE, and imaged by in-gel fluorescence (Figures 2.1B and 2.S2A). Western blotting for the His-tag was used to verify the iron-sensitive expression of NLL-MetRS (Figure 2.1B). Wild-type PA14 cell lysates were used as a control. The TAMRA signal was strong for P_{pvdF}*:nll-metRS* cells subjected to iron starvation, while minimal fluorescence was observed for P_{pvdF}*:nll-metRS* cells grown with iron and for wild-type cells grown either with or without iron (Figure 2.S2B). The weak TAMRA signal for the wild-type cells was likely due to a low level of nonspecific association of the dye with cellular proteins. To create a positive control for BONCAT labeling, a second gene cassette was inserted into the wild-type PA14 genome at the same *att*Tn7 site to obtain the strain P_{trc}*:nll-metRS* (Figure 2.S3A), in which the constitutive promoter P_{trc} drives expression of the NLL-MetRS regardless of cell state. As expected, the TAMRA signals derived from P_{trc}*:nll-metRS* samples were strong and independent of growth conditions (Figures 2.S3B and 2.S3C). P_{pvdF}*:nll-metRS*, P_{trc}*:nll-metRS*, and wild-type cells exhibited essentially identical growth rates in planktonic culture and indistinguishable capacities to form biofilms as reported by a crystal violet assay (Figures 2.S4 and 2.S5) (30).

Cell lysates from Anl-treated wild-type and P_{pvdF}*:nll-metRS* samples were used to assess the extent to which BONCAT labeling and LC-MS/MS quantification could provide selective enrichment and analysis of proteins from iron-starved cells. Samples were treated with DBCO-agarose beads (Figure 2.S1B). After extensive washing, bound proteins were released and digested with Lys-C/trypsin and analyzed by LC-MS/MS (Figure 2.1C) (31). Samples of

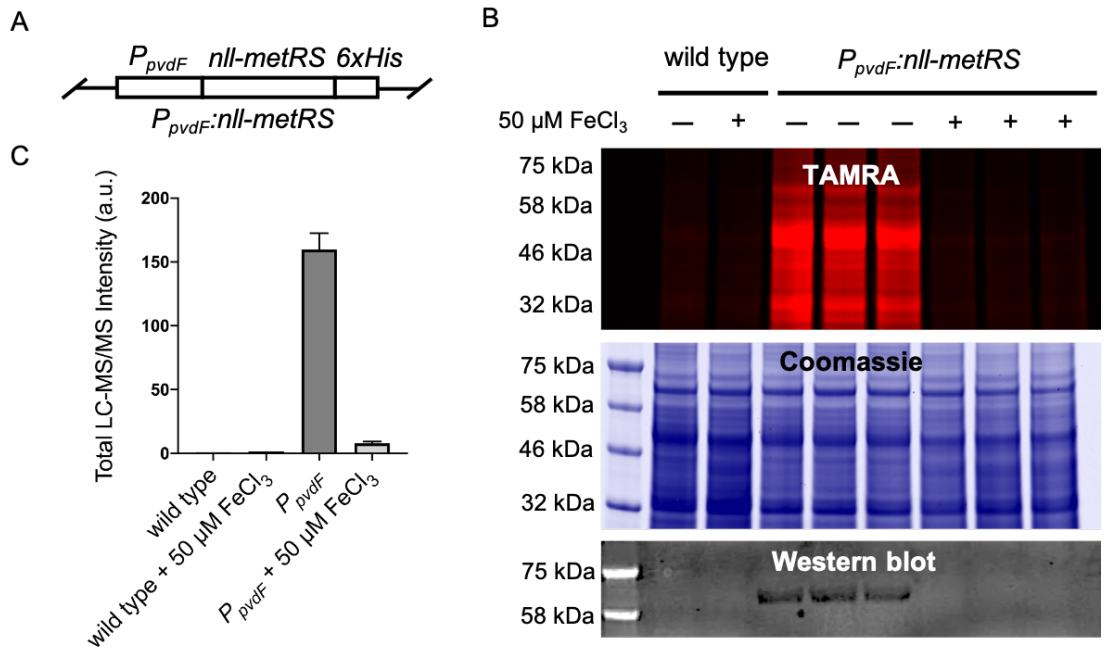

Figure 2.1 Cell-state-selective labeling of planktonic *P. aeruginosa* cells.

(A) The PA14 strain of *P. aeruginosa* was engineered to express the *E. coli* NLL-MetRS under control of the P_{pvdF} promoter. The expression cassette was transposed to the *att*Tn7 site in the PA14 genome. (B) The wild-type and P_{pvdF}*:nll-metRS* strains were cultured either with or without supplementation with 50 μ15M $FeCl_3$. Cell cultures were incubated with 1 mM Anl for 1 h. Lysates were treated with TAMRA-alkyne and separated by SDS-PAGE, and the TAMRA fluorescence of the gel was imaged. Coomassie staining of the same gel verified equal protein loading in each lane. Western blotting was applied to each lysate to probe expression of NLL-MetRS. (C) The same cell lysates were used for BONCAT enrichment and LC-MS/MS analysis. The total LC-MS/MS intensity for each sample was used to quantify the extent of protein enrichment. (Error bars: standard deviation, N = 3.)

P_{pvdF}*:nll-metRS* grown in the absence of iron yielded total ion intensities approximately 20-fold higher than those grown with 50 μM $FeCl_3$, and intensities more than 100-fold higher than those observed for wild-type samples. The slightly higher background of P_{pvdF}*:nll-metRS*

samples grown with iron compared to the two wild-type samples was likely due to low levels of leaky expression of the NLL-MetRS.

We used Gene Ontology (GO) enrichment analysis to categorize the 50 most highly expressed proteins from the total of 1834 proteins identified among three replicates of P_{pvdF}:*nll-metRS* cultures grown without iron (32). GO analysis showed that proteins involved in the pyoverdine biosynthetic process were highly over-represented in these samples (Table 2.1). In addition, three proteins involved in pyochelin biosynthesis (PchE, PchF) and transport (FptA), which

Table 2.1 Gene Ontology enrichment analysis. The 50 most highly expressed proteins found in 3 biological replicates of planktonic P_{pvdF}:*nll-metRS* samples were analyzed in terms of the GO biological process database. Proteins involved in pyoverdine biosynthesis were found to be highly enriched. Fold change indicates the ratio of the number of genes with the annotation observed in the list to the number of genes expected. A fold change larger than 1 suggests over-representation of the annotation in the list, while a fold change smaller than 1 indicates under-representation.

GO biological process	GO ID	# proteins	Fold change	FDR adjusted p-value
Cellular catabolic process	GO: 0044248	11	3.74	1.24E-02
Organic substance catabolic process	GO: 1901575	10	3.30	4.68E-02
Regulation of biological quality	GO: 0065008	7	4.78	4.35E-02
Pyoverdine biosynthetic process	***GO: 0002049***	***6***	***30.34***	***3.09E-04***
Tricarboxylic acid cycle	GO: 0006099	5	24.19	7.13E-04
Purine ribonucleoside triphosphate biosynthetic process	GO: 0009206	4	14.83	1.62E-02
Translational elongation	GO: 0006414	3	37.09	1.24E-02
Protein autophosphorylation	GO: 0046777	2	> 100	3.23E-02

are reported to be positively regulated upon iron starvation (through Fur and the transcriptional activator PchR), were found in the list (33-36), and PvdF itself was identified among the 100 most highly expressed proteins (Data Set A.1). Taken together, our analyses of planktonic cultures suggested that P_{pvdF} could be used to target BONCAT labeling to iron-starved cells. Full proteomic results are provided in Data Set A.1.

2.3.2 Visualization of labeled biofilms in flow cells

To confirm selective BONCAT labeling in *P. aeruginosa* biofilms, we grew films on cover glasses in flow cells under constant flow (0.06 mL/min) of 1% TSB with 1 mM $MgCl_2$ and 100 μM $FeCl_3$ at 37°C. After four days, films were treated with 1 mM Anl for 1.5 h. To determine the spatial patterning of Anl incorporation, biofilms were fixed and permeabilized, treated with TAMRA-alkyne, counterstained with SYTO 9, and examined by confocal microscopy. Only the interior regions of P_{pvdF}*:nll-metRS* biofilms exhibited TAMRA fluorescence, suggesting that cells in the interior were depleted of iron while those on the periphery were not (Figures 2.2A and 2.S6). In contrast, TAMRA fluorescence was strong throughout P_{trc}*:nll-metRS* biofilms, and hardly detected in the wild-type samples (Figures 2.2B, 2.2C, 2.S7 and 2.S8).

2.3.3 Selective proteomic profiling of iron-starved biofilm subpopulations via BONCAT

To prepare samples for selective proteomic profiling of iron-starved cellular subpopulations, we grew biofilms of P_{pvdF}*:nll-metRS* and P_{trc}*:nll-metRS* (each in triplicate) at 37°C for four days in silicone rubber tubing with 1% TSB plus 1 mM $MgCl_2$ and 100 μM $FeCl_3$. The medium flow rate was kept at 0.25 mL/min. After treatment with 1 mM Anl for 3 h, cells were collected

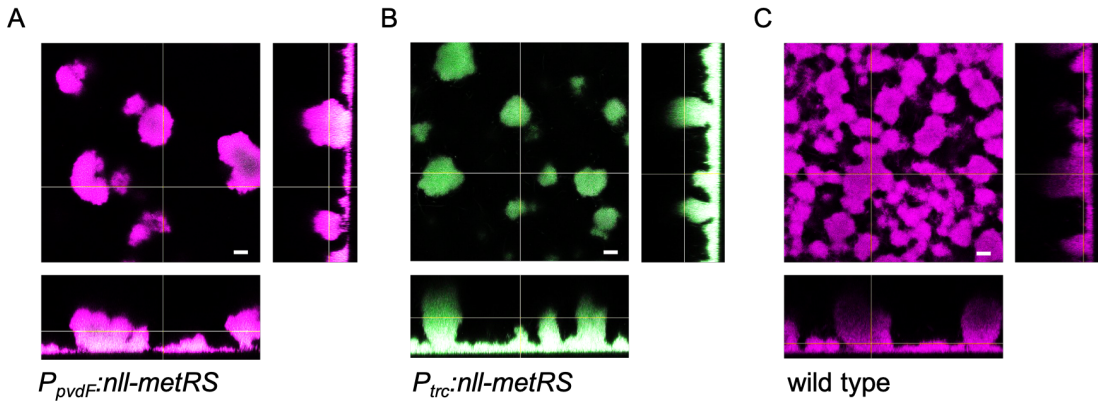

Figure 2.2 Visualization of biofilm labeling in flow cells. Four-day-old biofilms were treated with 1 mM Anl for 1.5 h, fixed and permeabilized, treated with TAMRA-alkyne (false-colored green), counterstained with SYTO 9 (false-colored magenta), and imaged by confocal microscopy. Colocalization of the TAMRA and SYTO 9 signals is shown in white. The cross-sectional images were reconstructed from confocal image stacks using ImageJ. (A) Imaging of the *P_{pvdF}:nll-metRS* biofilm revealed Anl labeling only in the interior of the film. (B) Imaging of the *P_{trc}:nll-metRS* biofilm showed Anl labeling throughout the film. (C) Imaging of the wild-type biofilm showed minimal TAMRA signal. (Scale bars: 10 μm.) Images recorded in individual TAMRA and SYTO 9 channels are shown in Figures S6-S8.

and lysed. Anl incorporation was verified by treating cell lysates with TAMRA-alkyne and imaging in-gel fluorescence (Figure 2.3A). As expected, the fluorescence of *P_{pvdF}:nll-metRS* samples was weaker than that of *P_{trc}:nll-metRS* samples, reflecting labeling of only a subpopulation of cells. Anl-labeled proteins were enriched by treatment with DBCO-agarose beads and analyzed by LC-MS/MS. Intensity Based Absolute Quantification (iBAQ) values, which are proportional to the absolute molar quantities of the identified proteins within each sample, were calculated (37), and label-free quantification (LFQ) with normalization was used

to quantify relative amounts of each identified protein across all samples (38). Among the six LC-MS/MS runs, 2287 proteins were identified, of which 107 were found only in P_{pvdF}*:nll-metRS* samples and 613 were found only in P_{trc}*:nll-metRS* samples (Figure 2.3B). Spearman's rank correlation coefficients for iBAQ values confirmed strong correlation of protein abundances among replicates for each strain (Figure 2.3C). We performed principal component analysis (PCA) using the LFQ values to visualize the variance among the samples (Figure 2.3D). Biological replicates for each strain clustered in the two-dimensional PCA plot, with P_{trc}*:nll-metRS* clustered more tightly than P_{pvdF}*:nll-metRS*. Separation between the two strains was substantially larger, suggesting that the program of protein synthesis in the iron-depleted subpopulation of cells is distinct from that of the biofilm as a whole. A second set of three biological replicates yielded similar results (Figure 2.S9).

For each of the proteins found in both strains, we calculated relative abundance (P_{pvdF}*:nll-metRS*/P_{trc}*:nll-metRS*) using normalized LFQ values, and determined the corresponding FDR adjusted *p*-values (Figure 2.3E). Normalization of LFQ values was such that the median relative abundance was approximately 1:1 (38). This step accounts for the observation that cells in biofilm interiors are often characterized by reduced metabolic rates (39-41), making direct comparison of absolute protein abundances problematic. Use of normalized LFQ values illustrates more clearly the shifts in protein synthesis that cells use to adapt to local environmental conditions in different regions of the film. Full proteomic results are provided in Data Sets A.2 and A.3.

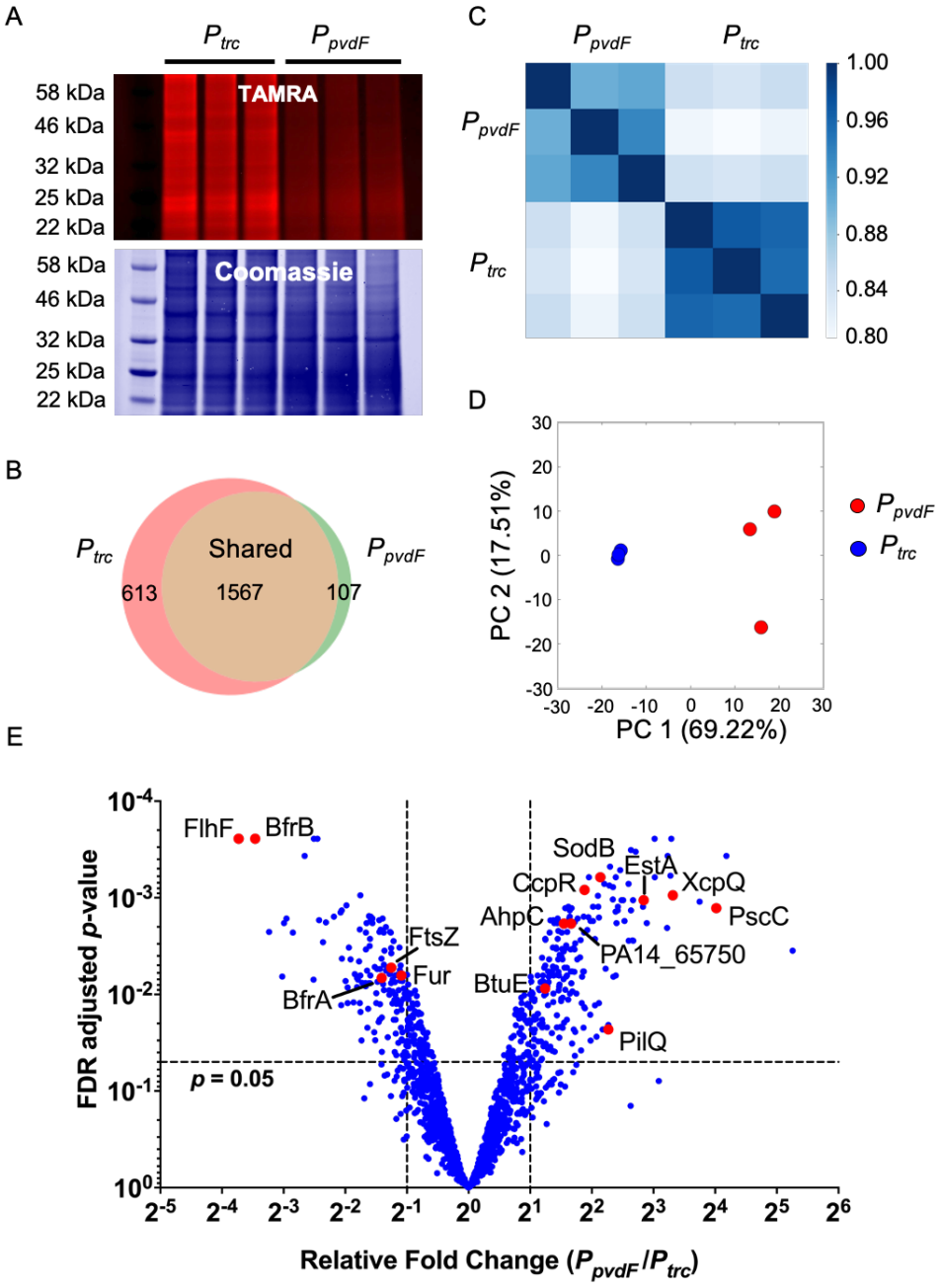

Figure 2.3 Cell-state-selective labeling of *P. aeruginosa* biofilms. Biofilms of *P_{pvdF}:nll-metRS* and *P_{trc}:nll-metRS* strains (each in triplicate) were grown for 4 days in silicone rubber tubing and treated with 1 mM Anl for 3 h. Cells were collected and lysed. (A) Anl-labeling was visualized by treating cell lysates with TAMRA-alkyne. Coomassie staining was used to verify equal protein loading among lanes. (B) Venn diagram of proteins quantified in the two types of biofilms. (C) Spearman's rank correlation coefficients of iBAQ values for each sample

showed strong correlation of protein abundances among replicates of each strain. (D) Principal component analysis (PCA) using normalized LFQ values to visualize variance among the samples; 1162 proteins were included in the analysis. (E) Quantification of relative fold change and the corresponding FDR adjusted *p*-value for each protein found both in P_{pvdF}*:nll-metRS* and in P_{trc}*:nll-metRS* samples.

2.3.4 Pathway regulation in the iron-starved subpopulation

We used the Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathway database to identify metabolic pathways that show evidence of differential regulation in iron-depleted and whole-biofilm samples (42). The relative abundances of proteins associated with each KEGG pathway were compared to a control in which all proteins shared by the two strains were included (Figure 2.4). Because the median relative abundance was set to be 1 as previously described, the median log2 fold change of the control group was close to 0. As determined by a Mann-Whitney *U* test, proteins from the oxidative phosphorylation, nitrogen metabolism, and ribosome pathways were found to be upregulated in the iron-depleted subpopulation, while the xenobiotic biodegradation and metabolism, and biofilm formation pathways showed evidence of reduced expression (43). Full results of the KEGG Pathway analysis are provided in Data Set A.4.

Nitrogen metabolism facilitates the growth of *P. aeruginosa* under anaerobic conditions through a denitrification process that reduces nitrate (NO_3^-) sequentially to nitrite (NO_2^-), nitric oxide (NO), nitrous oxide (N_2O), and dinitrogen (N_2) (Figure 2.5A) (44). It has been reported that the oxygen concentration at the biofilm-fluid interface can be as low as 40% of the

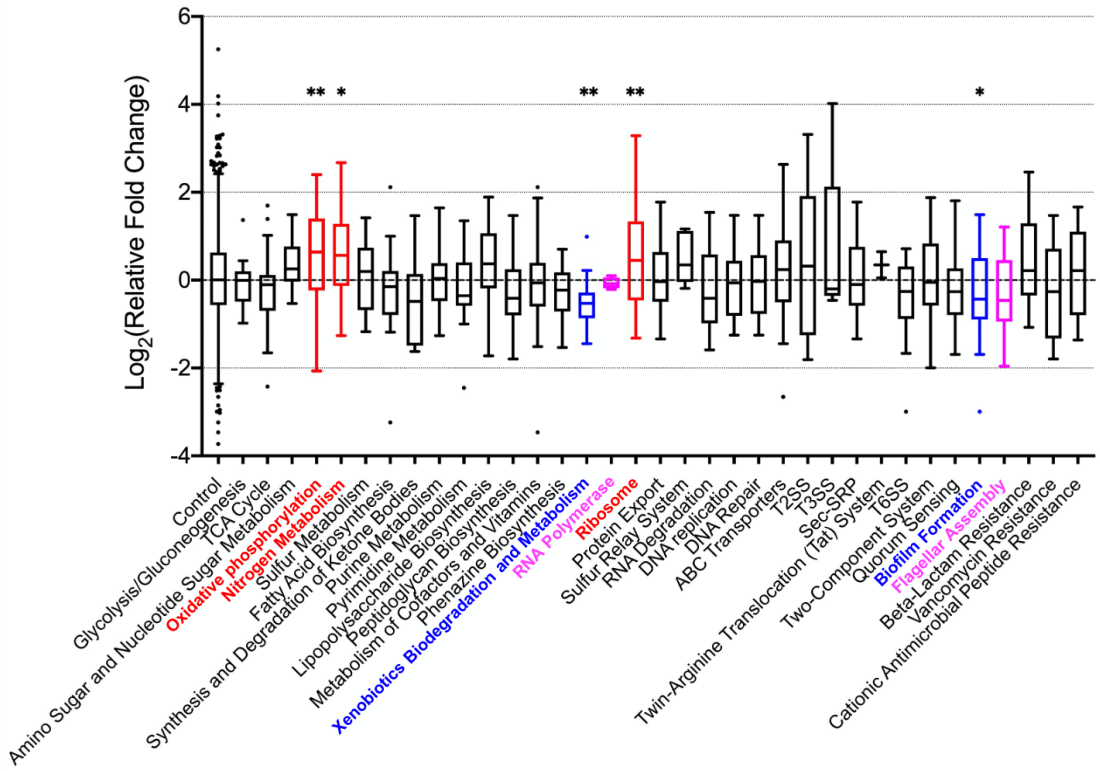

Figure 2.4 Regulation of cellular processes in *P. aeruginosa* biofilms. Proteins identified in both the P_{pvdF}*:nll-metRS* and P_{trc}*:nll-metRS* strains were categorized according to the Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathway database. Tukey's box plots show the relative fold change of proteins for each annotated pathway, with each box extending from the 25th to the 75th percentile of the distribution. The Mann-Whitney *U* test was applied to determine if a given pathway was regulated differently in the two samples (*, $p < 0.05$; **, $p < 0.01$).

concentration in the bulk fluid, and that it decreases rapidly with depth until oxygen is fully depleted 100 – 200 microns below the surface (4, 45). Many studies have validated the shift of *P. aeruginosa* to nitrogen metabolism under oxygen limitations characteristic of the biofilm interior, where our labeling studies indicate that cells are also subject to iron depletion (44, 46, 47). Consistent with this interpretation, P_{pvdF}*:nll-metRS* samples showed evidence of

significant upregulation of genes in the denitrification pathway, including *narK1*, *narK2*, and *narG*, which lie within the *narK1K2GHJI* operon encoding one of the main nitrate reductase complexes in the denitrification pathway, and *norB* and *norC*, which encode nitric oxide reductase subunits B and C (Figure 2.5A).

Interestingly, we found evidence of significant upregulation of the ribosomal protein synthesis and assembly pathway in the iron-depleted subpopulation; the median relative abundance of the ribosomal proteins in the subpopulation was approximately 1.4-fold higher than in the film as a whole. Of the 52 ribosomal proteins identified in the BONCAT analysis, 15 were found at significantly higher relative abundance in the subpopulation, and the 50S ribosomal protein L20 (RplT) was nearly 10-fold more abundant. As cells in the biofilm interior often exhibit reduced metabolic rates (4), this result may be surprising. There is, however, an earlier finding that several ribosomal proteins are found at higher abundance in *P. aeruginosa* biofilms under anaerobic conditions than under aerobic conditions (47). Also, it has been shown that genes encoding ribosomal proteins are hyper-expressed in biofilms of *P. aeruginosa* (7) (as well as *E. coli* (48) and *Xylella fastidiosa* (49)) as compared to the corresponding planktonic cultures. Therefore, the metabolic rates and ribosomal protein abundance are not as closely related to each other as people normally perceive. And our results suggest that *P. aeruginosa* iron-starved biofilm subpopulations, while having reduced metabolic rates, are actually characterized with significant upregulation of ribosomal proteins.

The oxidative phosphorylation pathway in P_{pvdF}*:nll-metRS* samples was also upregulated.

Specifically, multiple subunits of NADH dehydrogenase I (NDH-1), succinate dehydrogenase, the *cbb*$_3$-type cytochrome *c* oxidase (*cbb*$_3$-Cco), and the f-type ATPase, were upregulated. Upregulation of the succinate dehydrogenase complex encoded by the *sdhABCD* operon and of the f-type ATPase encoded by the *atpIBEFHAGDC* operon is consistent with the results of previous studies of anoxic *P. aeruginosa* biofilms (9). Furthermore, the genes *nuoJ*, *nuoL* and *nuoM*, which encode subunits of NDH-1, and the genes *PA14_10500*, *PA14_44340*, *PA14_44370* and *PA14_44380*, which encode subunits of *cbb*$_3$-Cco, were highly upregulated (Figure 2.5B). Recent studies have shown that mutations in genes of the *nuo* operon result in severe growth defects under anaerobic conditions (50-52), and that *cbb*$_3$-Cco promotes *P. aeruginosa* biofilm formation under anoxic denitrification, probably through NO accumulation (53, 54). Upregulation of these genes in P_{pvdF}*:nll-metRS* samples thus suggests that regions of the biofilm interior that are depleted in iron are also characterized by low oxygen concentrations, and oxidative phosphorylation pathway is vital for *P. aeruginosa* biofilm subpopulations under anaerobic conditions.

The observed downregulation of xenobiotic biodegradation and metabolism is most simply interpreted as further evidence for an overall reduction in metabolic activity in the biofilm interior. In similar fashion, downregulation of the biofilm formation pathway reflects the maturity of the four-day old films that were subject to Anl labeling. Downregulation of these two pathways supports the idea that the iron-depleted subpopulation of cells adopts a conservative survival strategy that allocates scarce resources to a limited number of pathways.

The flagellar assembly pathway, although not significantly downregulated as determined by the Mann-Whitney *U* test, had a median relative abundance 1.4-fold lower in the subpopulation than in the entire biofilm. The flagellar biosynthesis protein FlhF, required for swimming and swarming, is the most strongly downregulated of the shared proteins, more than 13-fold less abundant in the iron-depleted subpopulation (Figure 2.3E). FlhF determines the flagellar pole location and is required for initiation of flagellar assembly (55, 56). Downregulation of *flhF* is consistent with the report by Greenberg and coworkers showing that flagellar genes are repressed during biofilm maturation (7); the results shown here suggest that downregulation is especially acute in iron-depleted subpopulations of cells. Moreover, there are four other flagellar assembly proteins (FlgK, FlhA, FliF, MotA) that were identified only in the P_{trc}*:nll-metRS* samples (in all three replicates), and which therefore could not be included in the analysis of relative abundance. Taking into consideration all of the identified flagellar assembly proteins, we conclude that the flagellar assembly pathway is downregulated in the iron-depleted subpopulation (Figure 2.5C). This observation is consistent with accepted models of biofilm development that suggest that flagella are involved in early-stage cell attachment and then repressed during biofilm maturation (5, 7, 57). Interestingly, despite the general downregulation of proteins involved in flagellar assembly, the flagellin type B protein FliC was significantly upregulated in the subpopulation. FliC is part of the flagellar two-component signaling system and is a marker for biofilm dispersion (58). Upregulation of FliC in the context of downregulation of the flagellar assembly pathway merits further investigation.

2.3.5 Altered expression of proteins in the iron-starved subpopulation

In addition to the pathways for which differential regulation was found to be statistically significant, some protein regulations are also worth noting. We focused first on proteins that were found at significantly higher relative abundance in P_{pvdF}*:nll-metRS* than in P_{trc}*:nll-metRS* samples (relative fold change > 2 and FDR adjusted p-value < 0.05), and on proteins identified only in P_{pvdF}*:nll-metRS* samples (in at least two of three P_{pvdF}*:nll-metRS* replicates but not in any of the P_{trc}*:nll-metRS* samples). We used GO analysis to categorize these proteins according to molecular function (Table 2.2). Proteins involved in ribosomal assembly and activity, energy metabolism, and transport systems for ion exchange and nutrient uptake were significantly enriched. This analysis suggests that iron-depleted cells reallocate resources toward basic metabolic processes, ion homeostasis, and acquisition of nutrients. The number of proteins associated with DNA binding and transcriptional regulation was significantly less than would be expected on the basis of the number of proteins encoded by the annotated genes in the whole *P. aeruginosa* genome, suggesting an overall decrease in transcriptional activity.

Multiple proteins that provide protection against peroxides are found to be significantly upregulated in the subpopulation (Figure 2.3E). For example, the cytochrome *c* 551 peroxidase CcpR, which provides resistance towards peroxides, is found to be 3.7-fold more abundant in the subpopulation (59); the superoxide dismutase (SOD) SodB, which is shown to provide resistance to superoxide-generating agents, is more than 4.4-fold upregulated in the iron-starved subpopulation even though it is iron-cofactored (60); the alkyl hydroperoxide reductase AhpC is 2.9-fold upregulated; and a putative glutathione peroxidase BtuE is 2.4-fold

upregulated in the subpopulation. These observations suggest that, even though the subpopulation is subject to low-oxygen conditions, the cells regulate protein expression to defend against reactive oxygen species.

We found several outer membrane channels involved in protein secretion to be significantly upregulated in the iron-depeleted subpopulation (61, 62) (Figure 2.3E). The Type III secretion system protein PscC was 16-fold more abundant in the subpopulation, and the third most highly upregulated protein in the profile. The general secretion pathway protein D XcpQ, which is involved in Type II secretion, was 10-fold more abundant and the fifth most strongly upregulated protein. The esterase EstA involved in the Type Va secretion system was 7-fold upregulated (62); the type 4 fimbrial biogenesis outer membrane protein PilQ (required for piliation) was approximately 5-fold upregulated in the subpopulation (63); and an outer membrane efflux protein PA14_65750, suggested to be the TolC protein involved in the Type I secretion system in *P. aeruginosa*, was 3.2-fold upregulated in the subpopulation (62). Most of the other secretion-system proteins were not significantly upregulated in the subpopulation. The upregulation of multiple secretion-related outer membrane proteins in iron-starved *P. aeruginosa* cells has not been demonstrated previously, and merits further investigation.

Among the downregulated proteins (those characterized by relative fold change < 0.5 and FDR adjusted p-value < 0.05, plus those identified only in two or more of the P_{trc}:*nll-metRS* samples), several merit special note. The cell division protein FtsZ was significantly downregulated in the iron-starved subpopulation, while ZipA (another cell division protein) and MinE (a cell

Table 2.2 Gene Ontology enrichment analysis of proteins upregulated or found exclusively in P_{pvdF}:*nll-metRS* samples. The GO molecular function database suggested upregulation of genes involved in ribosomal assembly and activity, energy metabolism, and transport systems, and downregulation of genes for DNA binding and transcriptional regulation.

GO molecular function	GO ID	# proteins	Fold Enrichment	FDR adjusted p-value
cation: cation antiporter activity	GO:0015491	5	9.54	3.39E-02
NADH dehydrogenase (quinone) activity	GO:0050136	6	9.00	1.32E-02
porin activity	GO:0015288	6	9.00	1.27E-02
solute: cation symporter activity	GO:0015294	5	8.75	3.83E-02
solute: proton antiporter activity	GO:0015299	7	8.16	7.62E-03
rRNA binding	GO:0019843	15	8.07	1.58E-06
structural constituent of ribosome	GO:0003735	19	7.12	2.17E-07
heme binding	GO:0020037	14	4.59	9.14E-04
metal ion transmembrane transporter activity	GO:0046873	9	4.20	3.51E-02
electron transfer activity	GO:0009055	17	3.22	5.33E-03
DNA binding	GO:0003677	10	0.37	1.82E-02
transcription regulator activity	GO:0140110	5	0.28	3.77E-02

division topological specificity factor) were identified only in P_{trc}:*nll-metRS* samples, indicating repression of cell division under conditions of iron depletion (Figure 2.3E). The downregulation of the transcripts of *ftsZ* and *minE* genes at the bottom part of *P. aeruginosa* biofilms was also observed in previous transcriptomic profiling study (15). The decreased transcriptional activity as discussed above and the significant downregulation of division-related genes are consistent with the slow-growing—and perhaps dormant—nature of cells at the bottom of biofilms (4). BfrA and BfrB, two heme-binding bacterioferritins involved in iron storage, were 2.7-fold and 11.0-fold less abundant in the subpopulation, respectively (Figure 2.3E). The significant downregulation of BfrA and BfrB is consistent with their reported repression under iron starvation (64) (Figure 2.3E). Previous study of gene expressions in *P. aeruginosa* biofilms via LCM and transcriptome profiling also reveals that the mRNA levels

of these two genes are significantly higher in the upper regions of biofilms compared to the lower regions (15). Finally, Fur was found at reduced relative abundance in the subpopulation (Figure 2.3E). We have found no prior evidence for iron-sensitive regulation of *fur* in *P. aeruginosa*. Ochsner and coworkers have suggested that the promoters *fur* P1 and *fur* P2 are not regulated by iron (65), and Chuanchuen and Schweizer report that treatment of *P. aeruginosa* with triclosan, which causes iron accumulation, does not lead to changes in *fur* transcript levels (66). The observed downregulation of Fur in our iron-starved cellular subpopulations of *P. aeruginosa* biofilms suggests that there may be post-transcriptional mechanisms that couple the abundance of Fur to the intracellular concentration of iron.

We also note that all the four RNA polymerases in *P. aeruginosa* show little difference in relative abundance in the two populations (Figure 2.4). In contrast, ribosomal proteins are differentially regulated, suggesting that the transcriptional and translational machineries of the iron-depleted cells are regulated differently. Comparison of the proteomic profile obtained in our study with previous studies of *P. aeruginosa* biofilm heterogeneity via LCM and transcriptome profiling also reveals that while there is generally strong correlation between transcript and protein levels, discrepancies do exist. Pérez-Osorio and coworkers showed that the mRNA level of the master transcriptional stress response regulator *rpoS* is approximately 2 orders of magnitude higher at the top of biofilms than that at the bottom (67). However, our data indicate that the relative abundance of RpoS in the subpopulation is similar to that in the whole biofilm, with a fold change of 1.1. Another transcriptome study done by Williamson and coworkers found that transcripts involved in ATP metabolism (*atpIBEFHAGDC*) were 30-fold

more abundant near the top of biofilms, while our data indicate different regulation patterns, with AtpF 2.1-fold and AtpH 4.5-fold upregulated in the subpopulation. Only AtpG was found significantly downregulated in the subpopulation with a fold change of 0.24. We noted another striking difference in the expression pattern of the hypothetical protein PA14_57950 (PA4463 ortholog in PAO1 strain). We found that PA14_57950 was more than 2-fold downregulated in the subpopulation, while a previous study found a high abundance of the protein (PA4463) near the bottom of the biofilm (15). PA4463 is reported to be homologous to the ribosome-associated inhibitor RaiA of *E. coli*, suggesting its function in maintaining the stability of ribosomes while keeping them in a hibernating state. Because high levels of mRNA abundance do not necessarily indicate high levels of the corresponding proteins, the proteomic profiling method developed here can provide unique, direct insight into the programs of protein synthesis characteristic of cellular subpopulations.

Surprisingly, we did not detect any of the iron-regulated proteins involved in pyoverdine biosynthesis, pyochelin biosynthesis, or iron transport systems when we labeled four-day old P_{pvdF}*:nll-metRS* samples for 3 h. Because the labeling method relies on P_{pvdF}-driven expression of the NLL-MetRS, the absence of proteins involved in pyoverdine biosynthesis was especially surprising, and suggested that the levels of expression of iron-regulated genes were too low to be detected by LC-MS/MS. Therefore, we examined whether extending the Anl labeling time in the biofilm experiment would allow us to detect iron-controlled proteins. After growing P_{pvdF}*:nll-metRS* and P_{trc}*:nll-metRS* biofilms for 4 days, we labeled each sample with 1 mM Anl for 12 h. Consistent with our expectations, we found evidence for expression of iron- controlled

genes in the P_{pvdF}*:nll-metRS* samples. For example, the proteins encoded by *pvcA*, a paerucumarin biosynthesis gene positively regulated by PvdS and negatively by iron whose mutation inhibits pyoverdine biosynthesis (68); *pchD*, required for pyochelin biosynthesis; *exbD*, a tonB-dependent transporter gene involved in ferripyoverdine transport (69); *feoB*, a ferrous iron transporter; and *hitB*, the ferric iron-transport system permease, were all identified in P_{pvdF}*:nll-metRS* samples. When we compared relative protein abundances in P_{pvdF}*:nll-metRS* and P_{trc}*:nll-metRS* samples labeled for 12 h, we observed 1.5- fold upregulation of ExbD, 4.0-fold upregulation of HitB, and 2.8-fold upregulation of FeoB. We could not determine relative abundances for PvcA and PchD because of their low absolute abundances; we found only one peptide for each of these proteins in P_{pvdF}*:nll-metRS* samples, with low iBAQ values in each case (70). These results support the hypothesis that the levels of expression of many iron-regulated genes were low under the conditions used for BONCAT labeling. Consistent with expectation, none of the major iron-related genes was found to be significantly downregulated in P_{pvdF}*:nll-metRS* samples, with only one protein HitA (a ferric iron-binding periplasmic protein for transport) having a relative fold change of 0.59. Full proteomic results from the extended labeling biofilm experiment are provided in Data Set A.5.

2.3.6 The iron-starved cellular subpopulation in P. aeruginosa biofilms experience a unique microenvironment

Because our data suggest that the iron-depleted cells in *P. aeruginosa* biofilms are subject to other stresses, we compared our proteomic profile with the results of a previous biofilm study by Babin and coworkers, who used the stress promoter *rpoS* to target expression of the NLL-

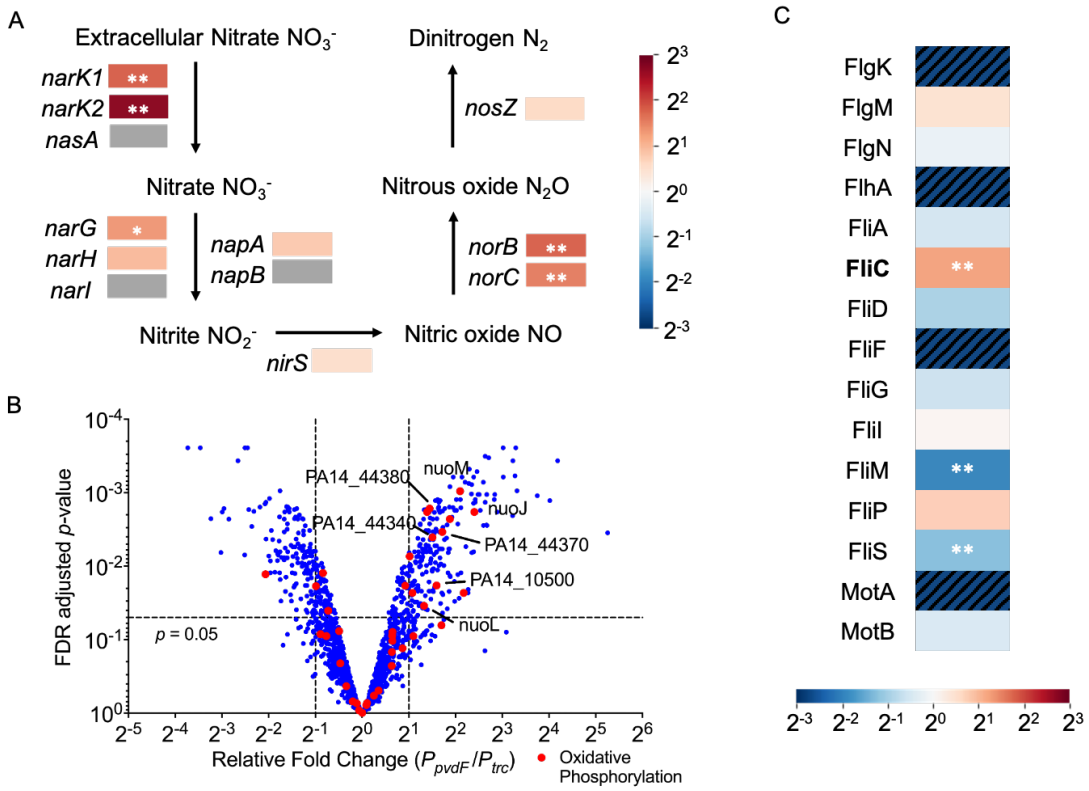

Figure 2.5 Differentially regulated pathways in the biofilm subpopulation. (A) The denitrification pathway. The heatmap indicates the relative fold changes (P_{pvdF}/P_{trc}) of the proteins in the process (*, $p < 0.05$; **, $p < 0.01$). The grey boxes indicate that the protein was not identified in any sample or identified only in one sample. (B) Volcano plot showing the relative fold changes of proteins from the oxidative phosphorylation pathway. (C) Heatmap of the relative fold changes (P_{pvdF}/P_{trc}) of proteins in the flagellar assembly pathway (*, $p < 0.05$; **, $p < 0.01$). The hatched dark blue boxes denote proteins identified only in P_{trc}:*nll-metRS* samples (in at least 2 replicates).

MetRS for BONCAT analysis of stressed cells (71). In the Babin study, 15 proteins were found to be significantly upregulated in the subpopulation (fold change > 2, FDA-adjusted *p*-value <0.05). Three of these proteins (PA14_47120, PA14_00480, and PA14_04300) were also found among the 281 proteins significantly upregulated in or unique to our P_{pvdF} samples. All three

are uncharacterized proteins (Table 2.3). The low extent of overlap indicates that regulation of protein expression in the iron-starved subpopulation probed in this work is distinct from that in the subpopulation characterized by expression of *rpoS*. Data Set A.6 lists the proteins identified both in the Babin study and in this work, along with the full proteomic results obtained here.

Table 2.3 Proteins significantly upregulated in both the iron-starved subpopulation and the RpoS-elevated subpopulation.

PA14 Locus ID	Gene Name	Protein function	Fold change (P_{pvdF} /P_{trc})	FDR adjusted *p*-value	Fold change (P_{rpoS} /P_{trc})	FDR adjusted *p*-value
PA14_00480	NA	Uncharacterized protein	5.78	1.74E-03	5.00	1.16E-02
PA14_04300	NA	Uncharacterized protein	9.67	5.91E-04	3.60	4.08E-02
PA14_47120	NA	Uncharacterized protein	3.06	1.13E-02	4.84	7.79E-03

We also compared our results with those of an earlier study the effect of oxygen limitation on *P. aeruginosa* physiology, where the proteomic profile of *P. aeruginosa* planktonic culture grown under anaerobic conditions was compared to the one grown under aerobic conditions (72). We would like to see which proteins significantly upregulated in our results are potentially regulated by oxygen availability. Among the 50 proteins either significantly upregulated or exclusively identified in the anaerobic condition in the previous study, 14 are also either significantly upregulated or exclusively found in our iron-starved subpopulation (Table 2.4). From this second comparison, we again observed that the iron-starved subpopulation is under a unique microenvironment that is under but not only under oxygen limitation. Data Set A.7

lists the proteins identified both in the oxygen limitation study and in this work, along with the full proteomic results obtained here.

Table 2.4 Proteins significantly upregulated in both the iron-starved subpopulation and the entire *P. aeruginosa* biofilm under anaerobic growth condition.

PA14 Locus ID	Gene Name	Protein function	Fold change (P_{pvdF} /P_{trc})	FDR adjusted *p*-value	Fold change (anaerobic /aerobic)	FDR adjusted *p*-value
PA14_00480	NA	Uncharacterized protein	5.78	1.74E-03	10.45	3.92E-04
PA14_01490	NA	Putative hemolysin	2.60	6.36E-03	3.88	3.95E-02
PA14_04300	NA	Uncharacterized protein	9.67	5.91E-04	10.35	2.19E-05
PA14_28450	eco	Ecotin	2.45	5.80E-03	18.41	2.19E-03
PA14_43680	fabA	3-hydroxydecanoyl-ACP dehydratase	2.00	1.36E-02	9.35	4.27E-04
PA14_61650	pagL	Lipid A 3-O-deacylase	3.32	2.44E-02	6.93	3.15E-02
PA14_66875	phaF	Polyhydroxyalkanoate synthesis protein	5.07	1.04E-03	3.57	3.99E-02
PA14_73280	atpH	F0F1 ATP synthase subunit delta	4.52	2.31E-02	5.35	4.55E-02
PA14_03490	NA	Uncharacterized protein	P_{pvdF} only	NA	Anaerobic only	NA
PA14_06740	nirM	Cytochrome c-551	3.64	2.21E-02	Anaerobic only	NA
PA14_07890	NA	ABC transporter permease	3.50	4.14E-03	Anaerobic only	NA
PA14_09030	rpmD	50S ribosomal protein L30	4.91	4.77E-04	Anaerobic only	NA
PA14_60450	rpmA	50S ribosomal protein L27	9.33	2.05E-03	Anaerobic only	NA
PA14_60700	ccpR	Cytochrome c551 peroxidase	3.68	8.29E-04	Anaerobic only	NA

2.3.7 Uncharacterized protein PA14_52000 enriched in the iron-starved subpopulation is iron-related

Among the proteins found to be either highly enriched or exclusively identified in P_{pvdF}*:nll-metRS* samples, those of unknown function constitute a candidate pool for the identification of new iron-related genes. We tested seven uncharacterized genes that encode proteins that were either exclusively found in at least two replicates of the P_{pvdF}*:nll-metRS* samples or that were at least 3-fold upregulated (FDR adjusted *p*-value < 0.05) in those samples. The protein encoded by uncharacterized gene *PA14_52000*, which was 3.7-fold more abundant in the

biofilm subpopulation, proved to be of special interest. In comparisons with wild-type PA14 cells, a transposon insertion mutant (*PA14_52000::tn*) from a PA14 transposon insertion library constructed by Liberati and coworkers showed no growth defect in planktonic culture, but exhibited increased pyoverdine production under low-iron conditions and a significant reduction in biofilm formation (Figure 2.S10) (73). These results suggest that PA14_52000 merits further investigation with respect to its roles in biofilm development and iron uptake. PA14_52000 is a hydrophobic, 139-amino acid protein predicted to localize at the cell membrane (Table 2.S1) (74, 75). Both RaptorX and IntFOLD predict that PA14_52000 should form a structure dominated by four major helices (76, 77). Thus, it is likely that PA14_52000 is a membrane protein involved in ferripyoverdine transport or other aspects of iron transport. Most germane to this work, however, is the potential of BONCAT analysis to facilitate discovery of genes and proteins associated with specific cell states in phenotypically heterogeneous cell populations. The transcript of the *PA14_52000* gene ortholog *PA0948* of PAO1 strain was previously found significantly downregulated at the bottom part of *P. aeruginosa* biofilms, rendering it overlooked (15). Moreover, according to the Pseudomonas Genome database, *PA14_52000* is within an operon of three poorly characterized genes (*PA14_51980, PA14_51990, PA14_52000*), and the promoter region does not contain any known motifs (75). However, specifically targeting the iron-starved subpopulation and employing cell-selective BONCAT allow us to successfully identify this new iron-related protein despite that limited information is available for the entire operon.

2.4 Discussion

The results reported here demonstrate a straightforward method for distinguishing and elucidating the programs of protein synthesis of distinct cellular subpopulations in complex biological systems. In addition, this study provides new insight into the physiology of iron-depleted cells of *P. aeruginosa* biofilms. We find evidence that the transcriptional and translational machineries of the cell respond differently to iron depletion, and that iron-starved cells allocate metabolic resources to a few essential metabolic pathways, including ribosome synthesis and assembly, oxidative phosphorylation, nitrogen metabolism, and synthesis of transporters for ion exchange and nutrient uptake. Several aspects of the observed response suggest that iron-depleted cells are also subject to low oxygen concentrations, highlighting the need for caution in assigning a direct causal link between iron concentration and protein abundance. The reduced abundance of Fur in iron-depleted cells was unexpected, and may point to a post-translational mechanism for control of this important regulator. Finally, we found that inactivation of the previously uncharacterized protein PA14_52000 leads to increased pyoverdine production and reduced biofilm formation.

Important extensions of the work reported here are readily imagined. The temporal resolution of the BONCAT method should allow the developmental programs of biofilm formation, maturation and dispersal to be thoroughly characterized. Increased labeling times will reveal additional low-abundance proteins, and careful selection of promoters will enable BONCAT labeling to be directed to other cellular subpopulations that contribute to biofilm growth and

survival. The generality of the method points to straightforward extension to other microbial communities and higher-order systems.

2.5 Materials and Methods

2.5.1 Bacterial strains and growth conditions

The PA14 strain of *Pseudomonas aeruginosa* was used in this study. To construct PA14 mutants, pUC18T mini-Tn7T plasmids bearing previously described gene cassettes were transposed into the Tn7 site of wild-type PA14 genome through four-parental mating conjugation (29). The 1.5 kb region upstream of the *pvdF* CDS was amplified from *P. aeruginosa* genomic DNA, which was prepared using the GenElute bacterial DNA kit (Sigma-Aldrich). The gene encoding the mutant *E. coli* methionyl-tRNA synthetase was cloned from plasmid BMB8 (71). The P_{trc} promoter element was cloned from plasmid BMB4 (71). Enzymes used for cloning and assembly were from New England Biolabs. The transposon insertion mutants (including *PA14_52000::tn*) were obtained from the PA14 transposon insertion mutants library (73).

Media used in this study were LB (5 g yeast extract, 10 g tryptone, 10 g NaCl per liter), 1% TSB (10 g tryptic soy broth per liter) with 1 mM $MgCl_2$, and M63 minimal medium with 1 mM $MgSO_4$, 0.2% glucose (w/v) and 0.5% casamino acid (w/v). For experiments involving planktonic cultures, 1% TSB medium was treated with Chelex 100 resin (Bio-Rad; 5 g per liter medium) for 16 h at 4°C under moderate stirring to remove trace iron from the medium. After filtration to remove the resin, $MgCl_2$ was added to bring the concentration to 1 mM.

To measure rates of growth and pyoverdine production, overnight cultures were diluted 1:500 into fresh medium and transferred to 96-well plates. OD600 was used as a measure of cell density; fluorescence emission (ex/em: 400nm/455nm) was used to detect pyoverdine production. In proteomic studies of the planktonic state, overnight cultures of the P_{pvdF}*:nll-metRS*, P_{trc}*:nll-metRS*, and wild-type strains in 1% TSB plus 1 mM $MgCl_2$ were diluted 1:100 into fresh medium with or without supplementation with 50 μM $FeCl_3$ as described previously. Cultures were grown for 7 h and subjected to BONCAT labeling (see the ***BONCAT labeling and enrichment*** section for details).

To visualize the spatial character of biofilm labeling, films were grown according to the protocol previously described except that the bubble traps were removed (78). Overnight cultures of the wild-type, P_{pvdF}*:nll-metRS*, and P_{trc}*:nll-metRS* strains in LB medium were diluted 1:100 into 0.9% NaCl and inoculated into flow cells (1 mm × 4 mm × 40 mm; Stovall). For proteomic profiling of the biofilms, diluted overnight cultures of the P_{pvdF}*:nll-metRS* and P_{trc}*:nll-metRS* strains were inoculated into silicone rubber tubing (6 mm interior diameter, 20 cm length; McMaster-Carr). Biofilms were grown in 1% TSB plus 1 mM $MgCl_2$ and 100 μM $FeCl_3$ at 37°C with medium flow rates of 0.06 mL/min for imaging and 0.25 mL/min for proteomic profiling.

To quantify biofilm formation, a microtiter plate crystal violet assay was employed following a protocol previously described (30). Overnight cultures (in LB) of cells were diluted 1:100 into fresh M63 minimal medium supplemented with $MgSO_4$, glucose and casamino acids. Cells

were incubated for 24 h at 37°C before biofilm quantification. Each strain was examined in eight replicates.

2.5.2 Biofilm imaging

All treatments of biofilms in flow cells were introduced through syringe injection (except for PBS, which was substituted for medium flow). Before injection, the peristaltic pump was turned off, the tubes upstream of the flow cells were clamped, and the downstream tubes were unclamped; reagents were injected through the tubes into the flow cell channels with 30G needles. Four-day-old biofilms in flow cells were incubated with the same medium (1% TSB plus 1 mM $MgCl_2$ and 100 μM $FeCl_3$) supplemented with 1 mM Anl (Iris Biotech) for 1.5 h at 37°C. Films were then fixed with 4% formaldehyde for 30 min and permeabilized with 70% ethanol for 5 min on ice. Flow cells were flushed with PBS for 5 min at 0.06 mL/min to remove formaldehyde and ethanol. To block free thiol groups, biofilms were incubated with 100 mM chloroacetamide in the dark at 37°C for 30 min. Fixed biofilms were incubated with a solution containing 25 μM TAMRA-alkyne (Click Chemistry Tools), 500 μM tris(3-hydroxypropyltriazolylmethyl)amine (THPTA; Click Chemistry Tools), 100 μM $CuSO_4$, 5 mM sodium ascorbate, and 5 mM aminoguanidine hydrochloride in the dark at 37°C for 30 min. Excess dye was removed by washing biofilms extensively with PBS for 30 min at ~0.3 mL/min. Biofilms were counterstained with 1 μM SYTO 9 and imaged on a Zeiss LSM 800 confocal microscope with a 40× objective. Images were processed with ImageJ 64-bit (v 2.0.0) (79).

2.5.3 BONCAT labeling and enrichment

In proteomic studies of the planktonic state, diluted cell cultures were grown for 7 h at 37°C with shaking and labeled with 1 mM Anl for 1 h. Cells were pelleted at 4℃. For biofilm labeling studies, after four days of biofilm growth in silicone rubber tubing at 37°C, 1 mM Anl was added to the medium flow for either 3 h or 12 h at 37°C. Loosely attached cells were collected along with the medium within each tube. Remaining cells were collected by cutting the tubing into small pieces (1-2 cm long) and vortexing in 0.9% NaCl. Cell fractions were combined and cells were pelleted at 4℃.

Cell pellets were resuspended in lysis buffer (4% SDS in 1×PBS), and treated with protease inhibitor (cOmplete, Mini, EDTA-free Protease Inhibitor Cocktail; Roche). To fully lyse the cells, lysates were heated at 95℃ for 10 min, sonicated with a microtip probe for 1 min with an amplitude of 25% and pulse 5 s on/5 s off (Qsonica), and heated at 95℃ for 3 min. Lysates were clarified by centrifuging at 10000 *g* for 15 min. Protein concentration was quantified using bicinchoninic acid (BCA) kits (Pierce).

To visualize the extent of labeling via in-gel fluorescence imaging, cell lysates were incubated in the dark with 15 μM TAMRA-alkyne, 250 μM $CuSO_4$, 1.25 mM THPTA, 5 mM aminoguanidine hydrochloride, and 5 mM sodium ascorbate for 1 h at room temperature. Labeled lysates were loaded on protein gels (NuPAGE Novex 4-12% Bis-Tris Protein Gels; Thermo Fisher Scientific) and separated via SDS-PAGE. Gels were fixed and destained with 40% methanol, 10% acetic acid, and 50% DI water overnight (80), rehydrated with DI water

for 1 h and imaged on a Typhoon Trio Variable Mode Imager (GE Healthcare). Gels were stained with InstantBlue (Expedeon) Coomassie protein stain to verify equal protein loading. To detect NLL-MetRS expression by Western blotting, each lysate (30 μg protein) was loaded and separated by SDS-PAGE, transferred to a nitrocellulose membrane (GE Healthcare), treated with mouse anti-Penta·His Alexa Fluor 488 Conjugate (Qiagen), and imaged on the Typhoon imager.

For enrichment, 3 mg of each lysate was incubated in the dark with 100 mM chloroacetamide for 30 min at 65°C with shaking (1200 rpm) to alkylate free thiol groups. An equal volume of freshly made 8 M urea and 0.85 M NaCl in 1×PBS was added to each lysate. DBCO-agarose resin (25 μL of a 50% slurry; Click Chemistry Tools) was washed three times with 0.8% SDS in 1×PBS (w/v), resuspended in 25 μL of the same SDS wash buffer, and added to each lysate. Resin-treated lysates were rotated end-over-end for 16 h in the dark at room temperature. Resin samples were washed with water, treated with 1 mM dithiothreitol (DTT) in the dark for 15 min at 70°C, treated with 40 mM chloroacetamide in the dark for 30 min at room temperature, transferred to Poly-prep chromatography columns (Bio-Rad), and washed extensively with the following solutions: (i) 8 × 5 mL 0.8% SDS in 1×PBS (w/v); (ii) 8 × 5 mL 8 M urea in 100 mM Tris pH 8.0; and (iii) 8 × 5 mL 20% acetonitrile (I) in water (v/v). During the second wash with each solution, resin samples were incubated with the washing solution for either 10 min (solutions I and iii), or 30 min (solution ii). Resin samples were resuspended in 10% I with 50 mM ammonium bicarbonate. Supernatant was removed to yield a sample volume of 100 μL and the resin fraction was treated with 50 ng LysC for 4 h and with 100 ng trypsin for 16 h at

37°C. Supernatant was collected. Resins were washed twice with 150 µL 20% I and the washes were combined with the supernatant. Peptides were dried and desalted with C_{18} ZipTips (EMD Millipore).

2.5.4 LC-MS/MS

Liquid chromatography-tandem-mass spectrometry (LC-MS/MS) experiments were carried out as previously described (81). For biofilm experiments with 3 h Anl labeling times, digested samples were subjected to LC-MS/MS analysis on a nanoflow LC system, EASY-nLC 1000, (Thermo Fisher Scientific) coupled to an LTQ Orbitrap Elite mass spectrometer (Thermo Fisher Scientific, Bremen, Germany). For the EASY-nLC 1000 system, solvent A consisted of 97.8% H_2O, 2% I, and 0.2% formic acid (FA); solvent B consisted of 19.8% H_2O, 80% I, and 0.2% FA. Samples were loaded directly onto a 25-cm analytical HPLC column (50 µm ID) packed in-house with ReproSil-Pur C18AQ 3 µm resin (12 nm pore size, Dr. Maisch, Ammerbuch, Germany). The column was heated to 55°C. The peptides were separated with a 70-min gradient at a flow rate of 220 nL/min. The gradient was as follows: 2–30% Solvent B (60 min), 30–100% B (1 min), and 100% B (9 min). Eluted peptides were then ionized using a Nanospray Flex ion source (Thermo Fisher Scientific) and introduced into the mass spectrometer. The LTQ Orbitrap Elite was operated in a data-dependent mode, automatically alternating between a full-scan (m/z 400-1600, 120K resolution) in the Orbitrap and subsequent MS/MS scans of the 20 most abundant peaks in the linear ion trap (Top20 method). Data acquisition was controlled by Xcalibur 2.2 and LTQ Tune Plus 2.7 software (Thermo Fisher Scientific). For biofilm experiments with 12 h Anl labeling times, the digested samples were

subjected to LC-MS/MS analysis on a nanoflow LC system, EASY-nLC 1200, (Thermo Fisher Scientific) coupled to a Q Exactive HF Orbitrap mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) equipped with a Nanospray Flex ion source. Samples were directly loaded onto a C18 Aurora series column (Ion Opticks, Parkville, Australia). The 25 cm x 50 µm ID column (1.6 µm) was heated to 50° C. Peptides were separated with a 120 min gradient at a flow rate of 350 nL/min. The gradient was as follows: 2–6% Solvent B (7.5 min), 6-25% B (82.5 min), and 25-40% B (30 min), to 100% B (1 min) and 100% B (12 min). The Q Exactive HF Orbitrap was operated in data dependent mode. Spray voltage was set at 1.8 kV, S-lens RF level at 50, and heated capillary at 275 °C. Full scan resolution was set to 60,000 at m/z 200. Full scan target was 3×10^6 with a maximum injection time of 15 ms (profile mode). Mass range was set to 400−1650 m/z. For data dependent MS2 scans the loop count was 12, target value was set at 1×10^5, and intensity threshold was kept at 1×10^5. Isolation width was set at 1.2 m/z and a fixed first mass of 100 was used. Normalized collision energy was set at 28. Peptide match was set to off, and isotope exclusion was on. MS2 data was collected in centroid mode.

2.5.5 Proteomic data analysis

Thermo raw data files were analyzed using MaxQuant (v 1.6.1.0) (70) and were searched against the *P. aeruginosa* PA14 UniProt entries and a contaminant database (246 sequences). Methionine oxidation and protein N-terminal acetylation were set as variable modifications. The carbamidomethylation of cysteine was specified as a fixed modification. Trypsin was set as the digestion enzyme with up to two missed cleavages. Protein abundances were calculated

with MaxLFQ (38) and match between runs was enabled. The relative abundances of newly synthesized proteins among samples and the corresponding *p*-values were calculated using the limma package in R (v 3.5.1). Normalized LFQ values were used to calculate relative abundance. The Benjamini-Hochberg procedure was used to adjust *p*-values for false discovery. Spearman's rank correlation coefficients were calculated between samples using iBAQ values and principal component analysis was performed using log2 LFQ values of proteins identified in all 6 MS runs (n = 1162) in Matlab (v R2018a). Gene ontology (GO) enrichment analysis was performed by the PANTHER Overrepresentation Test (v 14.0, released on 2018-11-13) using the GO ontology database (released on 2019-01-01).

2.5.6 Software and database used

Additional software used included the following. The Mann-Whitney *U* test was performed in Prism 8 (v 8.0.2). Bar plots, volcano plots and boxplots were generated using Prism 8 (v 8.0.2). Venn diagrams and heatmaps were generated by Python (v 3.6.4). The chemical structures in Fig. S1 were drawn in ChemDraw (v 16.0). The Kyoto Encyclopedia of Genomes and Genomics (KEGG) Pathway database for the PA14 strain was used for system-level regulation analysis (42). The Gene Ontology PANTHER database (v 14.0) was used for analysis of genes and biological processes (32, 82). The physicochemical properties of PA14_52000 protein were calculated using ExPASy ProtParam (74). The 3D structure of PA14_52000 was predicted by RaptorX and IntFOLD (76, 77).

2.6 Acknowledgements

We acknowledge Roxana Eggleston-Rangel and Brett Lomenick of the Caltech Proteome Exploration Laboratory for assistance with liquid chromatography-tandem mass spectrometry and Dr. Andres Collazo and Caltech Biological Imaging Facility for training in and use of confocal microscopy. We thank Dr. Dianne Newman's laboratory for the kind gift of *P. aeruginosa* wild-type and transposon insertion mutant strains.

2.7 References

1. Mulcahy LR, Isabella VM, Lewis K (2013) *Pseudomonas aeruginosa* biofilms in disease. *Microb Ecol* 68(1):1–12.
2. Delden CV, Iglewski BH (1998) Cell-to-cell signaling and *Pseudomonas aeruginosa* infections. *Emerg Infect Dis* 4(4):551–560.
3. Cross A, et al. (1983) Nosocomial infections due to *Pseudomonas aeruginosa*: review of recent trends. *Clin Infect Dis* 5 (Supplement_5):S837-S845.
4. Stewart PS, Franklin MJ (2008) Physiological heterogeneity in biofilms. *Nat Rev Microbiol* 6(3):199–210.
5. Hall-Stoodley L, Costerton JW, Stoodley P (2004) Bacterial biofilms: from the natural environment to infectious diseases. *Nat Rev Microbiol* 2(2):95–108.
6. Waite RD, et al. (2006) Clustering of *Pseudomonas aeruginosa* transcriptomes from planktonic cultures, developing and mature biofilms reveals distinct expression profiles. *BMC Genomics* 7(1):162.
7. Whiteley M, et al. (2001) Gene expression in *Pseudomonas aeruginosa* biofilms. *Nature* 413(6858):860–864.
8. Hentzer M, Eberl L, Givskov M (2005) Transcriptome analysis of *Pseudomonas aeruginosa* biofilm development: anaerobic respiration and iron limitation. *Biofilms* 2(1):37–61.
9. Tata M, et al. (2016) RNAseq based transcriptional profiling of *Pseudomonas aeruginosa* PA14 after short- and long-term anoxic cultivation in synthetic cystic fibrosis sputum medium. *PloS One* 11(1):e0147811.
10. Mukherjee S, Moustafa D, Smith CD, Goldberg JB, Bassler BL (2017) The RhlR quorum-sensing receptor controls *Pseudomonas aeruginosa* pathogenesis and biofilm development independently of its canonical homoserine lactone autoinducer. *PloS Pathog* 13(7):e1006504.
11. Rusconi R, Guasto JS, Stocker R (2014) Bacterial transport suppressed by fluid shear. *Nat Phys* 10(3):212–217.

12. Drescher K, Shen Y, Bassler BL, Stone HA (2013) Biofilm streamers cause catastrophic disruption of flow with consequences for environmental and medical systems. *Proc Natl Acad Sci USA* 110(11):4345–4350.
13. Hochbaum AI, Aizenberg J (2010) Bacteria pattern spontaneously on periodic nanostructure arrays. *Nano Lett* 10(9):3717–3721.
14. Lenz AP, Williamson KS, Pitts B, Stewart PS, Franklin MJ (2008) Localized gene expression in *Pseudomonas aeruginosa* biofilms. *Appl Environ Microb* 74(14):4463–4471.
15. Williamson KS, et al. (2012) Heterogeneity in *Pseudomonas aeruginosa* biofilms includes expression of ribosome hibernation factors in the antibiotic-tolerant subpopulation and hypoxia-induced stress response in the metabolically active population. *J Bacteriol* 194(8):2062–2073.
16. Skaar EP (2010) The battle for iron between bacterial pathogens and their vertebrate hosts. *PloS Pathog* 6(8):e1000949.
17. Singh PK, Parsek MR, Greenberg EP, Welsh MJ (2002) A component of innate immunity prevents bacterial biofilm development. *Nature* 417(6888):552–555.
18. Singh PK (2004) Iron sequestration by human lactoferrin stimulates *P. aeruginosa* surface motility and blocks biofilm formation. *BioMetals* 17(3):267–270.
19. Banin E, Vasil ML, Greenberg EP (2005) Iron and *Pseudomonas aeruginosa* biofilm formation. *Proc Natl Acad Sci USA* 102(31):11076–11081.
20. Nairz M, Schroll A, Sonnweber T, Weiss G (2010) The struggle for iron—a metal at the host—pathogen interface. *Cell Microbiol* 12(12):1691–1702.
21. Venturi V, Weisbeek P, Koster M (1995) Gene regulation of siderophore-mediated iron acquisition in *Pseudomonas*: not only the Fur repressor. *Mol Microbiol* 17(4):603–610.
22. Vasil ML, Ochsner UA (1999) The response of *Pseudomonas aeruginosa* to iron: genetics, biochemistry and virulence. *Mol Microbiol* 34(3):399–413.
23. Visca P, Leoni L, Wilson MJ, Lamont IL (2002) Iron transport and regulation, cell 47 yoto 47 ing and genomics: lessons from *Escherichia coli* and *Pseudomonas*. *Mol Microbiol* 45(5):1177–1190.

24. Tanrikulu IC, Schmitt E, Mechulam Y, Goddard WA, Tirrell DA (2009) Discovery of *Escherichia coli* methionyl-tRNA synthetase mutants for efficient labeling of proteins with azidonorleucine in vivo. *Proc Natl Acad Sci USA* 106(36):15285–15290.
25. Link AJ, et al. (2006) Discovery of aminoacyl-tRNA synthetase activity through cell-surface display of noncanonical amino acids. *Proc Natl Acad Sci USA* 103(27):10180-10185.
26. McMorran BJ, Kumara HMCS, Lamont IL, Sullivan K (2001) Involvement of a transformylase enzyme in siderophore synthesis in *Pseudomonas aeruginosa*. *Microbiology+* 147(6):1517–1524.
27. Wilson MJ, Mcmorran BJ, Lamont IL (2001) Analysis of promoters recognized by 48yot, an extracytoplasmic-function sigma factor protein from *Pseudomonas aeruginosa*. *J Bacteriol* 183(6):2151–2155.
28. Gasser V, Guillon L, Cunrath O, Schalk I (2015) Cellular organization of siderophore biosynthesis in *Pseudomonas aeruginosa*: evidence for siderosomes. *J Inorg Biochem* 148:27–34.
29. Choi K-H, Schweizer HP (2006) mini-Tn7 insertion in bacteria with single *att*Tn7 sites: example *Pseudomonas aeruginosa*. *Nat Protoc* 1(1):153–161.
30. O'Toole GA (2011) Microtiter dish biofilm formation assay. *JoVE-J Vis Exp* (47):e2437.
31. Glatter T, et al. (2012) Large-scale quantitative assessment of different in-solution protein digestion protocols reveals superior cleavage efficiency of tandem Lys-C/trypsin proteolysis over trypsin digestion. *J Proteome Res* 11(11):5145-5156.
32. Ashburner M, et al. (2000) Gene ontology: tool for the unification of biology. *Nat Genet* 25(1):25.
33. Heinrichs DE, Poole K (1993) Cloning and sequence analysis of a gene (*pchR*) encoding an AraC family activator of pyochelin and ferripyochelin receptor synthesis in *Pseudomonas aeruginosa*. *J Bacteriol* 175(18):5882–5889.

34. Ochsner UA, Vasil AI, Vasil ML (1995) Role of the ferric uptake regulator of *Pseudomonas aeruginosa* in the regulation of siderophores and exotoxin A expression: purification and activity on iron-regulated promoters. *J Bacteriol* 177(24):7194–7201.
35. Heinrichs DE, Poole K (1996) PchR, a regulator of ferripyochelin receptor gene (*fptA*) expression in *Pseudomonas aeruginosa*, functions both as an activator and as a repressor. *J Bacteriol* 178(9):2586–2592.
36. Reimmann C, Serino L, Beyeler M, Haa D (1998) Dihydroaeruginoic acid synthetase and pyochelin synthetase, products of the *pchEF* genes, are induced by extracellular pyochelin in *Pseudomonas aeruginosa. Microbiology+* 144(11):3135–3148.
37. Schwanhäusser B, et al. (2011) Global quantification of mammalian gene expression control. *Nature* 473(7347):337–342.
38. Cox J, et al. (2014) Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol Cell Proteomics*13(9):2513–2526.
39. Xu KD, Stewart PS, Xia F, Huang CT, McFeters GA (1998) Spatial physiological heterogeneity in *Pseudomonas aeruginosa* biofilm is determined by oxygen availability. *Appl Environ Microb 64*(10):4035-9.
40. Fux C, Costerton J, Stewart P, Stoodley P (2005) Survival strategies of infectious biofilms. *Trends Microbiol* 13(1):34–40.
41. Walters MC, Roe F, Bugnicourt A, Franklin MJ, Stewart PS (2003) Contributions of antibiotic penetration, oxygen limitation, and low metabolic activity to tolerance of *Pseudomonas aeruginosa* biofilms to ciprofloxacin and tobramycin. *Antimicrob Agents Ch* 47(1):317–323.
42. Kanehisa M, Goto S (2000) KEGG: 49yoto encyclopedia of genes and genomes. *Nucleic Acids Res* 28(1):27-30.
43. Mann HB, Whitney DR (1947) On a test of whether one of two random variables is stochastically larger than the other. *Ann Math Stat* 50-60.

44. Alst NEV, Picardo KF, Iglewski BH, Haidaris CG (2007) Nitrate sensing and metabolism modulate motility, biofilm formation, and virulence in *Pseudomonas aeruginosa*. *Infect Immun* 75(8):3780–3790.
45. Zhang TC, Fu Y-C, Bishop PL (1995) Competition for substrate and space in biofilms. *Water Environ Res* 67(6):992–1003.
46. Southey-Pillig CJ, Davies DG, Sauer K (2005) Characterization of temporal protein production in *Pseudomonas aeruginosa* biofilms. *J Bacteriol* 187(23):8114–8126.
47. Yoon SS, et al. (2002) *Pseudomonas aeruginosa* anaerobic respiration in biofilms: relationships to cystic fibrosis pathogenesis. *Dev Cell* 3(4):593–603.
48. Schembri MA, Kjaergaard K, Klemm P (2003) Global gene expression in *Escherichia coli* biofilms. *Mol Microbiol* 48(1):253–267.
49. Souza AA, et al. (2004) Gene expression profile of the plant pathogen *Xylella fastidiosa* during biofilm formation in vitro. *FEMS Microbiol Lett* 237(2):341–353.
50. Filiatrault MJ, Picardo KF, Ngai H, Passador L, Iglewski BH (2006) Identification of *Pseudomonas aeruginosa* genes involved in virulence and anaerobic growth. *Infect Immun* 74(7):4237–4245.
51. Schurek KN, et al. (2008) Novel genetic determinants of low-level aminoglycoside resistance in *Pseudomonas aeruginosa*. *Antimicrob Agents Chem* 52(12):4213–4219.
52. Torres A, et al. (2019) NADH dehydrogenases in *Pseudomonas aeruginosa* growth and virulence. *Front Microbiol* 10:75.
53. Hamada M, Toyofuku M, Miyano T, Nomura N (2014) *cbb*$_3$-type cytochrome *c* oxidases, aerobic respiratory enzymes, impact the anaerobic life of *Pseudomonas aeruginosa* PAO1. *J Bacteriol* 196(22):3881–3889.
54. Yoon MY, Lee K-M, Park Y, Yoon SS (2011) Contribution of cell elongation to the biofilm formation of *Pseudomonas aeruginosa* during Anaerobic respiration. *PloS One* 6(1):e16105.
55. Dasgupta N, et al. (2003) A four-tiered transcriptional regulatory circuit controls flagellar biogenesis in *Pseudomonas aeruginosa*. *Mol Microbiol* 50(3): 809-824.

56. Murray TS, Kazmierczak BI (2006) FlhF is required for swimming and swarming in *Pseudomonas aeruginosa*. *J Bacteriol* 188(19):6995–7004.

57. O'Toole GA, Kolter R (1998) Flagellar and twitching motility are necessary for *Pseudomonas aeruginosa* biofilm development. *Mol Microbiol* 30(2):295–304.

58. Sauer K, et al. (2004) Characterization of nutrient-induced dispersion in *Pseudomonas aeruginosa* PAO1 biofilm. *J Bacteriol* 186(21):7312–7326.

59. Trunk K, et al. (2010) Anaerobic adaptation in *Pseudomonas aeruginosa*: definition of the Anr and Dnr regulons. *Environ microbiol* 12(6):1719-1733.

60. Hassett DJ, Schweizer HP, Ohman DE (1995) *Pseudomonas aeruginosa sodA* and *sodB* mutants defective in manganese-and iron-cofactored superoxide dismutase activity demonstrate the importance of the iron-cofactored form in aerobic metabolism. *J Bacteriol* 177(22):6330-6337.

61. Filloux A, Michel G, Bally M (1998) GSP-dependent protein secretion in Gram-negative bacteria: the Xcp system of *Pseudomonas aeruginosa*. *FEMS Microbiol Rev* 22(3):177-198.

62. Filloux A (2011) Protein secretion systems in *Pseudomonas aeruginosa*: an essay on diversity, evolution, and function. *Front Microbiol 2*:155.

63. Wolfgang M, van Putten JP, Hayes SF, Dorward D, Koomey M (2000) Components and dynamics of fiber formation define a ubiquitous biogenesis pathway for bacterial pili. *EMBO J* 19(23):6408-6418.

64. Ma JF, et al. (1999) Bacterioferritin A modulates catalase A (KatA) activity and resistance to hydrogen peroxide in *Pseudomonas aeruginosa*. *J Bacteriol* 181:3730-3742.

65. Ochsner UA, Vasil AI, Johnson Z, Vasil ML (1999) *Pseudomonas aeruginosa fur* overlaps with a gene encoding a novel outer membrane lipoprotein, OmlA. *J Bacterial* 181(4):1099-1109.

66. Chuanchuen R, Schweizer HP (2012) Global transcriptional responses to triclosan exposure in *Pseudomonas aeruginosa*. *Int J Antimicrob Ag* 40(2):114-122.

67. Pérez-Osorio AC, Williamson KS, Franklin MJ (2010) Heterogeneous *rpoS* and *rhlR* mRNA levels and 16S rRNA/rDNA (rRNA gene) ratios within *Pseudomonas aeruginosa* biofilms, sampled by laser capture microdissection. *Journal Bacteriology* 192(12):2991-3000.

68. Stintzi A, et al. (1999) The *pvc* gene cluster of *Pseudomonas aeruginosa*: role in synthesis of the pyoverdine chromophore and regulation by PtxR and PvdS. *J Bacteriol* 181(13):4118-4124.
69. Ringel M, Brüser T (2018) The biosynthesis of pyoverdines. *Microb Cell* 5(10):424.
70. Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* 26(12):1367–1372.
71. Babin BM, et al. (2017) Selective proteomic analysis of antibiotic-tolerant cellular subpopulations in *Pseudomonas aeruginosa* biofilms. *mBio* 8(5):e01593-17.
72. Babin BM, et al. (2016) SutA is a bacterial transcription factor expressed during slow growth in *Pseudomonas aeruginosa*. *Proc Natl Acad Sci USA* 113(5):E597-E605.
73. Liberati N, et al. (2006) An ordered, nonredundant library of *Pseudomonas aeruginosa* strain PA14 transposon insertion mutants. *Proc Natl Acad Sci USA* 103(8):2833-2838.
74. Gasteiger E, et al. (2003) ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res* 31(13):3784-3788.
75. Winsor G, et al. (2015) Enhanced annotations and features for comparing thousands of *Pseudomonas* genomes in the *Pseudomonas* genome database. *Nucleic Acids Res* 44(D1):D646-D653.
76. Wang S, Li W, Liu S, Xu J (2016) RaptorX-Property: a web server for protein structure property prediction. *Nucleic Acids Res*, 44(W1):W430-W435.
77. McGuffin L, Atkins J, Salehe B, Shuid A, Roche D (2015) IntFOLD: an integrated server for modelling protein structures and functions from amino acid sequences. *Nucleic Acids Res* 43(W1):W169-W173.
78. Nielsen MW, Sternberg C, Molin S, Regenberg B (2011) *Pseudomonas aeruginosa* and *Saccharomyces cerevisiae* biofilm in flow cells. *JoVE-J Vis Exp* (47):e2383.
79. Schneider CA, Rasband WS, Eliceiri KW (2012) NIH image to imageJ: 25 years of image analysis. *Nat Methods* 9(7):671–675.

80. Hong V, Presolski SI, Ma C, Finn MG (2009) Analysis and optimization of copper-catalyzed azide-alkyne cycloaddition for bioconjugation. *Angew Chem Int Edit* 48(52):9879–9883.

81. Kalli A, Hess S (2012) Effect of mass spectrometric parameters on peptide and protein identification rates for shotgun proteomic experiments on an LTQ-orbitrap mass analyzer. *Proteomics* 12(1):21–31.

82. Mi H, Muruganujan A, Ebert D, Huang X, Thomas PD (2018) PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res* 47(D1):D419-D426.

2.8 Supplementary Figures

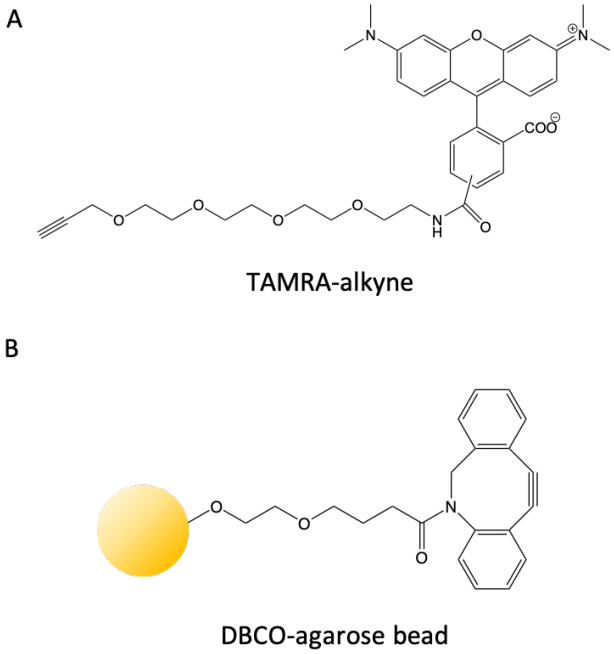

Figure 2.S1 Structures of reagents used in the study. (A) TAMRA-alkyne. (B) DBCO-agarose beads.

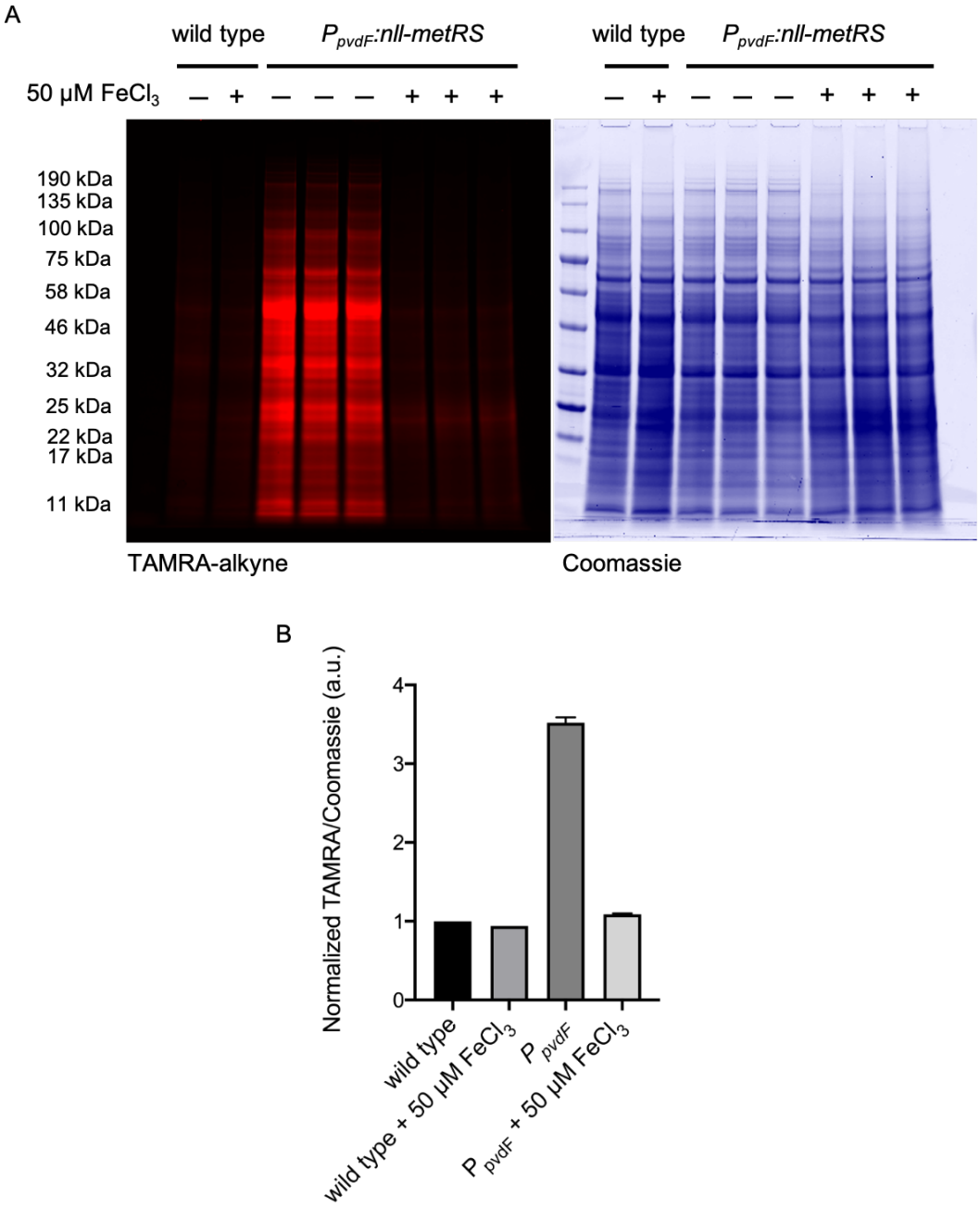

Figure 2.S2 Labeling of planktonic *P_{pvdF}:nll-metRS* cells. (A) Full gel images used to prepare Figure 1B. (B) Normalization of TAMRA signal for each sample with the corresponding Coomassie signal in Figure 2.S2. (Error bars: standard deviation, N = 3.)

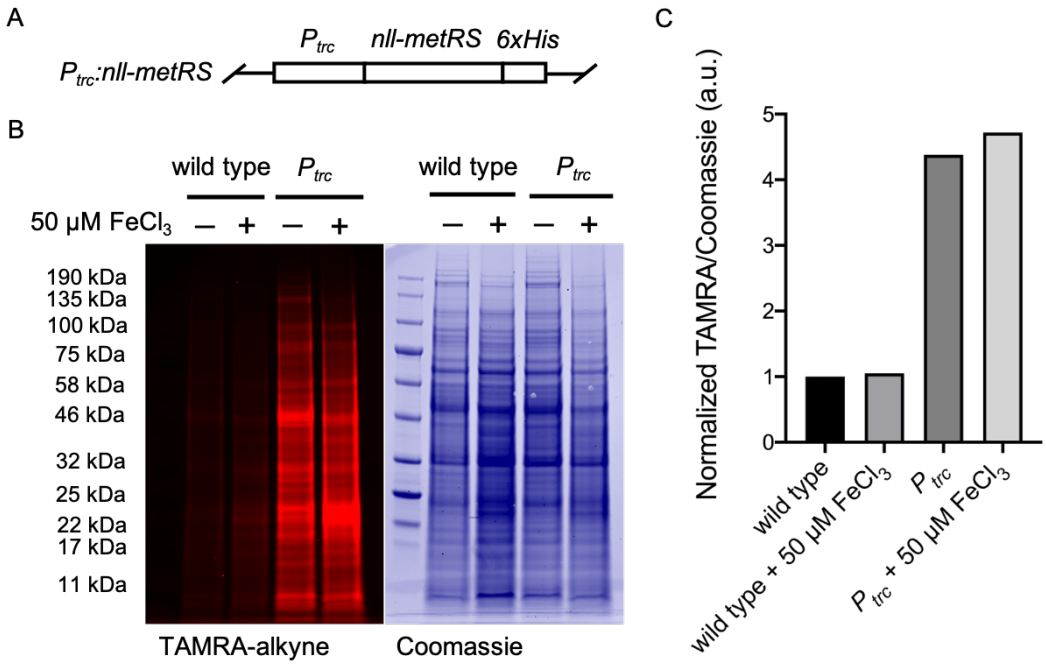

Figure 2.S3 Labeling of planktonic *P_{trc}:nll-metRS* cells. (A) The PA14 strain of *P. aeruginosa* was engineered to express the *E. coli* NLL-MetRS under control of the constitutive P_{trc} promoter. The expression cassette was transposed to the *att*Tn7 site in the PA14 genome. (B) The wild-type and *P_{trc}:nll-metRS* strains were cultured either with or without supplementation with 50 μM $FeCl_3$. Cell cultures were incubated with 1 mM Anl for 1 h. Lysates were treated with TAMRA-alkyne, separated by SDS-PAGE, and the gel was imaged by TAMRA fluorescence. Coomassie staining of the same gel verified equal protein loading among lanes. (C) Normalization of the TAMRA signal for each sample by the corresponding Coomassie signal revealed that Anl labeling was insensitive to iron concentration, as expected.

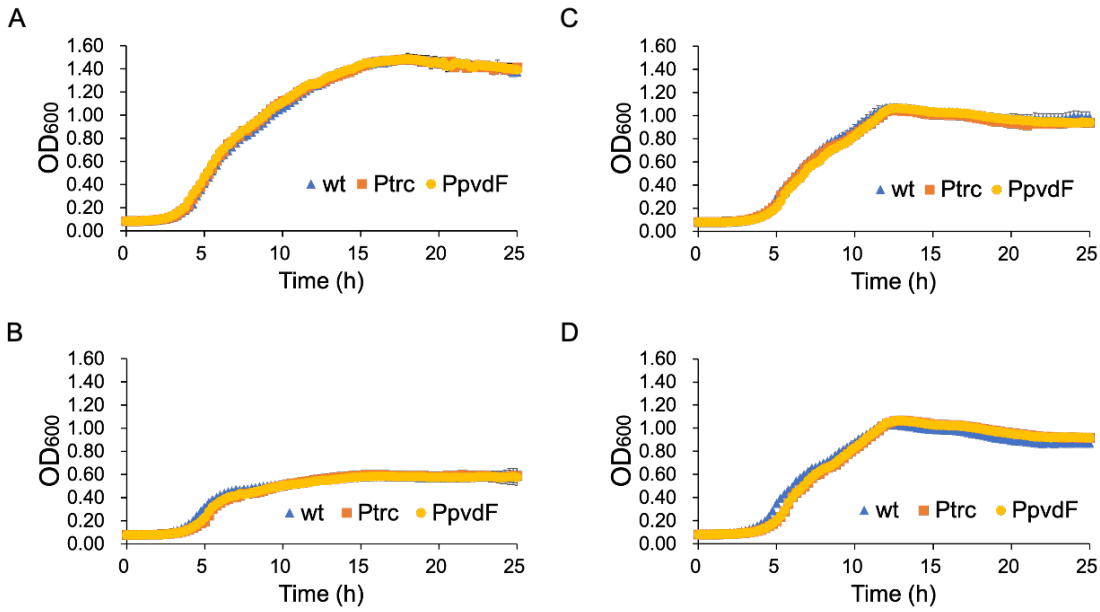

Figure 2.S4 Growth assays of the wild type, *P_{trc}:nll-metRS*, and *P_{pvdF}:nll-metRS* strains. All three strains showed similar planktonic growth in (A) LB, (B) 1% TSB + 1 mM $MgCl_2$, (C) 1% TSB + 1 mM $MgCl_2$ + 50 µM $FeCl_3$, and (D) 1% TSB + 1 mM $MgCl_2$ + 100 µM $FeCl_3$. (Error bars: standard deviation, N = 3.)

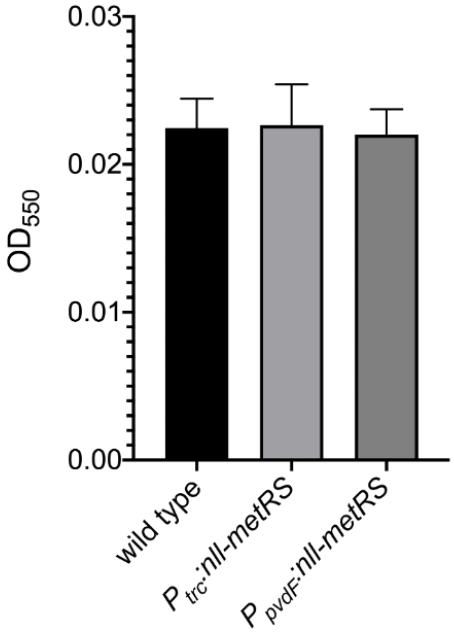

Figure 2.S5 Biofilm formation assays of the wild type, *P_{trc}:nll-metRS*, and *P_{pvdF}:nll-metRS* strains. The crystal violet assay showed no significant differences in biofilm formation among the three strains. (Error bars: standard deviation, N = 8.)

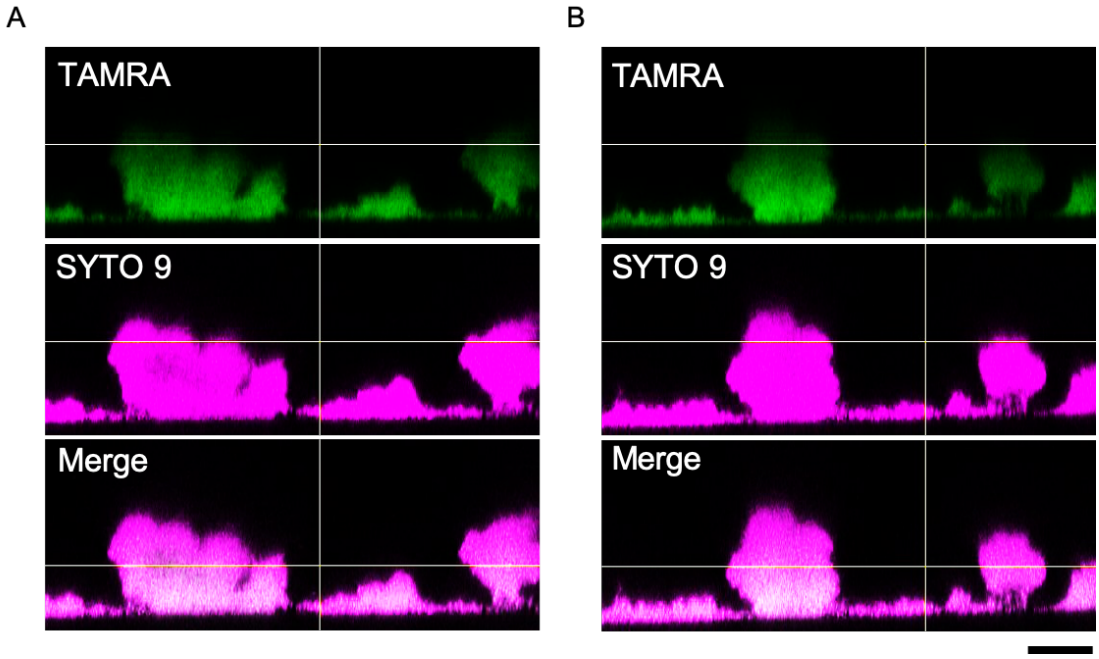

Figure 2.S6 Individual and merged images of labeled *P_{pvdF}:nll-metRS* biofilms. (A) View of the XZ plane. (B) View of the YZ plane. (Scale bar: 20 μm.)

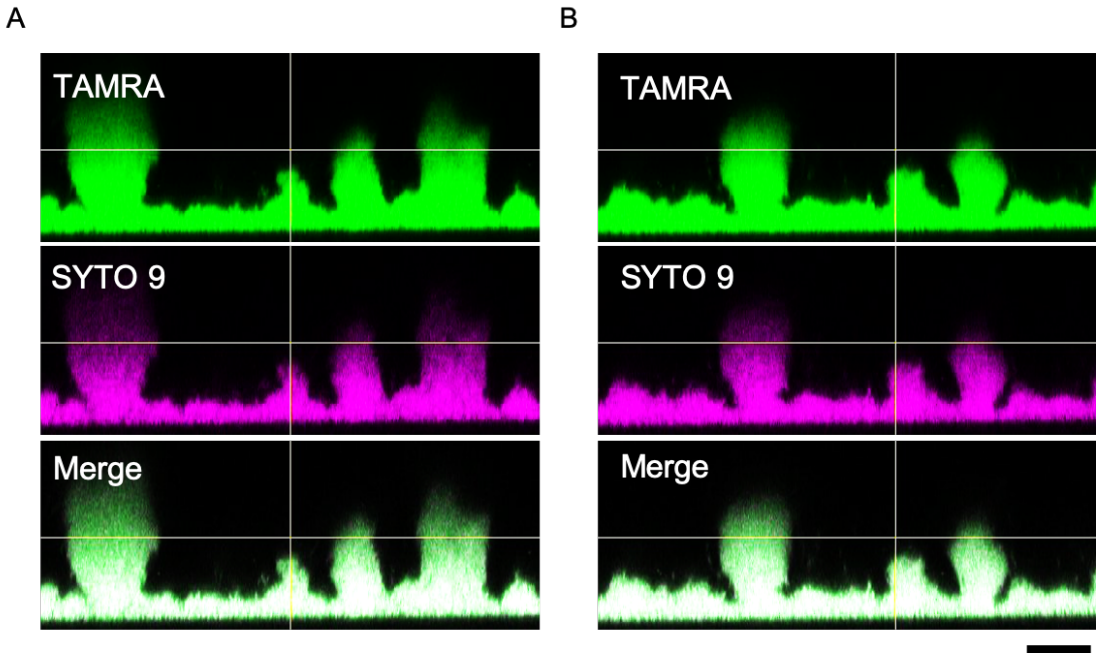

Figure 2.S7 Individual and merged images of labeled *P_{trc}:nll-metRS* biofilms. (A) View of the XZ plane. (B) View of the YZ plane. (Scale bar: 20 μm.)

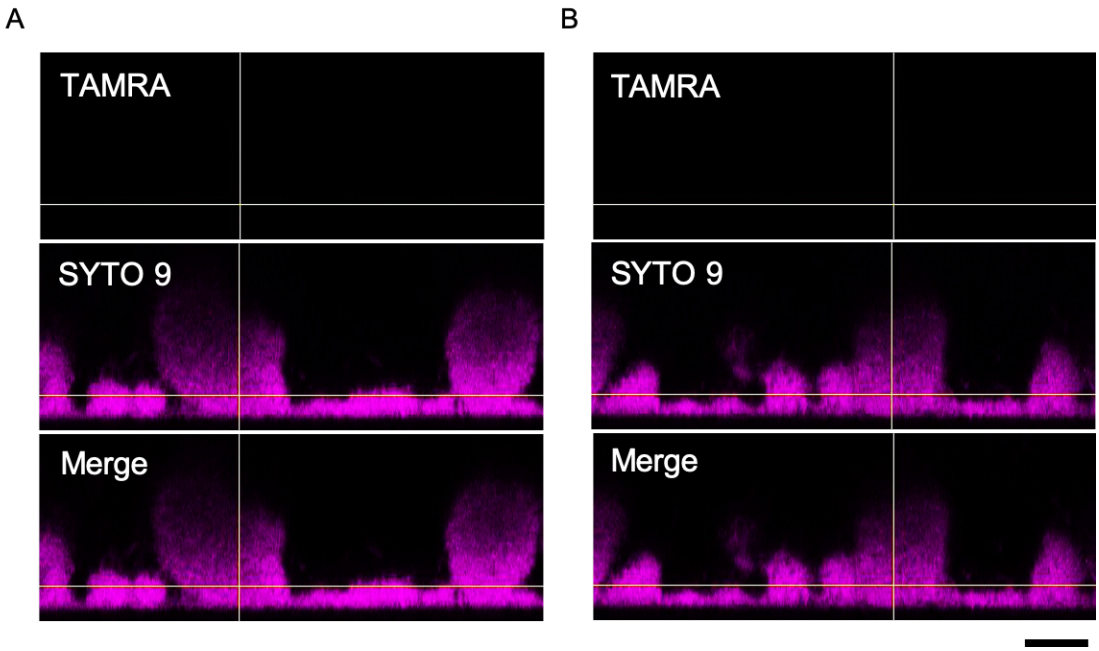

Figure 2.S8 Individual and merged images of labeled wild-type biofilms. (A) View of the XZ plane. (B) View of the YZ plane. (Scale bar: 20 μm.)

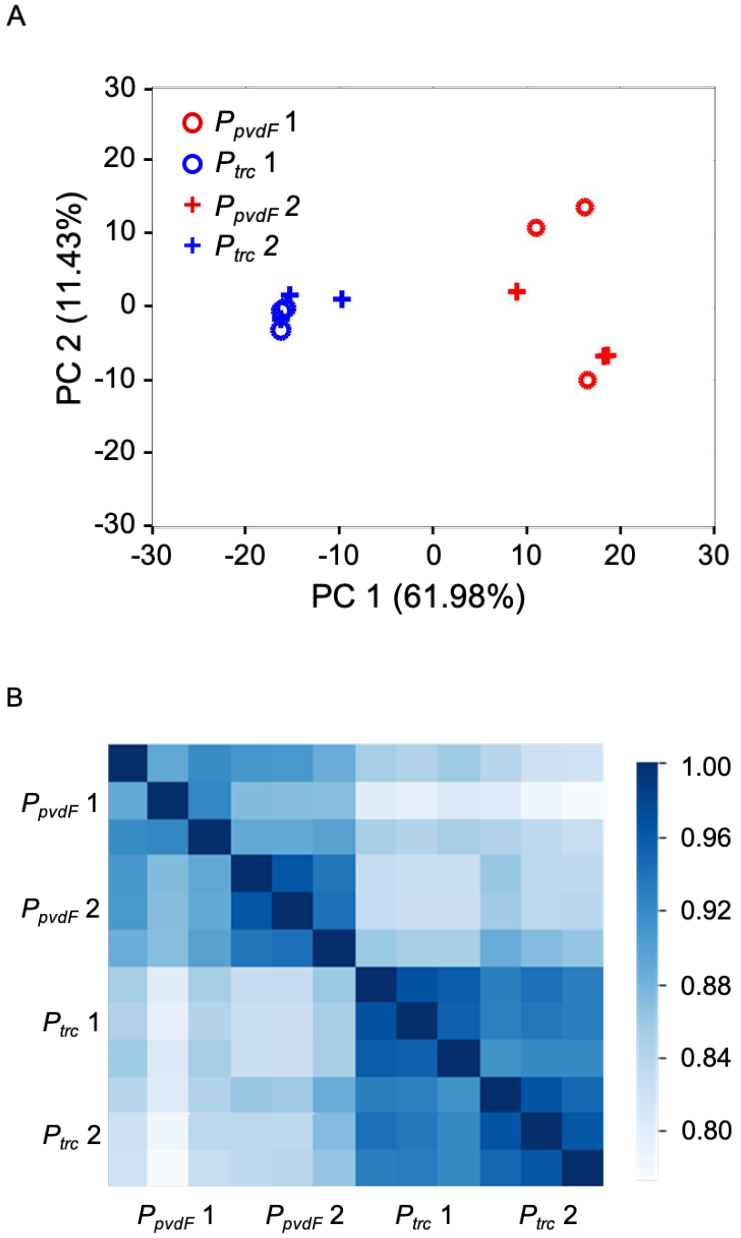

Figure 2.S9 Comparison of proteomic profiles obtained in replicates of the biofilm labeling experiment. Two sets of experiments were used in the analysis, with each set comprising three biological replicates. (A) Spearman's rank correlation coefficients of iBAQ values to visualize correlation among samples. (B) Principal component analysis using normalized LFQ to visualize the variance among samples. (1158 proteins were included in the analysis.)

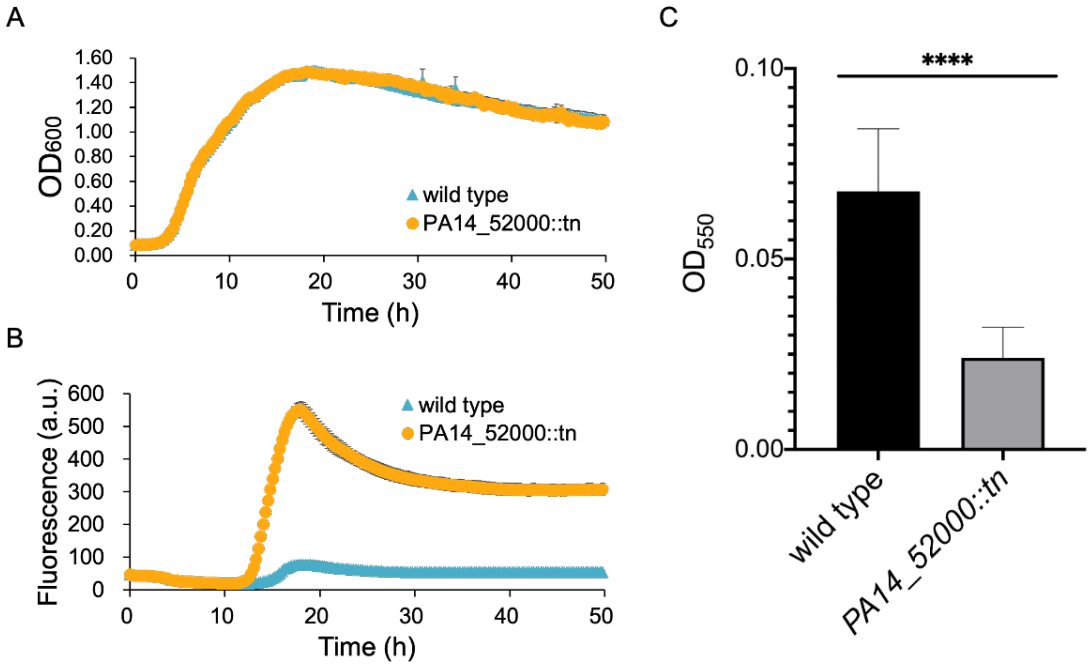

Figure 2.S10 Characterization of the transposon insertion mutant *PA14_52000::tn*. (A) Growth of wild type and mutant strains in LB medium. (Error bars: standard deviation, N = 3.) (B) Pyoverdine production as indicated by fluorescence emission (ex/em: 400 nm/455 nm) by the wild type and mutant strains in LB medium. (Error bars: standard deviation, N = 3.) (C) The crystal violet biofilm formation assay showed a significant defect in biofilm formation by the mutant strain. (Error bars: standard deviation, N = 8; ****, $p < 0.0001$.)

3.1 Supplementary Tables

Table 2.S1 Physico-chemical properties of PA14_52000 from ExPASy ProtParam (75).

Physico-chemical properties of PA14_52000	
Number of amino acids	139
Molecular weight	15566.63
Theoretical pI	10.01
Aliphatic index	130.58
Grand average of hydropathicity (GRAVY)	0.713

Table 2.S2 Strains and primers used in the study.

***Pseudomonas aeruginosa* strains**		
Name	Genotype	Source
DKN263	P. aeruginosa UCBPP-PA14	
P_{trc}:nll-metRS	UCBPP-PA14 *att*Tn7::mini-Tn7T-*Gm^R P_{trc}:nll-metRS*	This study
P_{pvdF}:nll-metRS	UCBPP-PA14 *att*Tn7::mini-Tn7T-*Gm^R P_{pvdF}:nll-metRS*	This study
PA14_52000::tn	UCBPP-PA14 gene::MAR2×T7	(73)
***Escherichia coli* strains**		
Name	Genotype	Source
	DH10B pTNS2	(29)
	HB101 pRK2013	(29)
XL1	Mach1 pUC18T-mini-Tn7T-*Gm^R P_{trc}:nll-metRS*	This study
XL2	Mach1 pUC18T-mini-Tn7T-*Gm^R P_{pvdF}:nll-metRS*	This study
Primers		
Name	Sequence	Purpose
Tn7T_F	GCATGCATATACCGGGCT	Tn7 vector construction
Tn7T_R	GAGCTCATGCATGATCGAAT	
P_{pvdF}_F	ATTCGATCATGCATGAGCTC TCTCGCCGACGATGAAGAC	P_{pvdF} construction
P_{pvdF}_R	GCGACTTGAGTCAT GGATCC TTCATCATTCCCAGGAGTGG	
P_{trc}_F	ATTCGATCATGCATGAGCTC ACTAGTGAGCTGTTGACAAT	P_{trc} construction
P_{trc}_R	GCGACTTGAGTCAT GGATCC TTTTTCTTCTCCTGAGCTCG	
*nll*_F	GGATCCATGACTCAAGTCGC	*nll-metRS* construction
*nll*_R	AGCCCGGTATATGCATGC TTAATGGTGGTGATGGTGGTG TTTAGAGGCTTCCACCAGTG	

Chapter 3

BIOORTHOGONAL NONCANONICAL AMINO ACID TAGGING (BONCAT) ENABLES CELL-SELECTIVE PROTEOMIC LABELING AND ANALYSIS OF *BACILLUS SUBTILIS* K-STATE SUBPOPULATION

3.1 Summary of Contributions

This work was performed in close collaboration with Yue Hui. Contributions that were primarily my own include BONCAT labeling and enrichment, protein quantification and analysis by LC-MS/MS. Yue's primary contributions include generation of plasmids and strains, and confocal microscopy. We shared the following work equally: experimental planning, adjusting K-state cell inducing and labeling conditions, data interpretation, and writing of manuscript.

3.2 Abstract

Bacteria have evolved with a wide range of strategies to survive under various fluctuations and stresses imposed by the environment. For the gram-positive bacterium *Bacillus subtilis*, entry into K-state is one of its characteristic adaptive responses to cope with environmental stresses during early stationary phase. The stochasticity of the phenomenon and the relatively low level of K-state inducing rate make the study of gene and pathway regulations hard to achieve. Here, we show that bioorthogonal noncanonical amino acid tagging (BONCAT) enables state-selective study of proteomic response in *B. subtilis*. By controlling the expression of a mutant methionyl-tRNA synthetase under an endogenous K-state-responsive promoter, we targeted BONCAT labeling exclusively to K-state cells. Compared to the entire population, proteomic response of labeled K-state subpopulation was characterized with system-level upregulation of pathways required for DNA repair and nutrient-scavenging, and with downregulation of those involved in ribosome biosynthesis, bacterial chemotaxis, and flagella assembly. The proteomic profile also provided evidence for gene regulations regarding cell division, cell wall synthesis, detoxification, and antibiotic synthesis that are characteristic of K-state cells. We also identified proteins that are uncharacterized or previously known for functions irrelevant of K-state highly enriched in the subpopulation, providing new insights toward K-state study. Together, we demonstrated that BONCAT can enhance our ability to systematically study a phenotypic subpopulation that arises stochastically.

3.3 Introduction

The ability of a genetically identical population to differentiate into distinct phenotypes in response to environmental perturbations is fundamental to bacteria (1-3). One of the most investigated stochastic phenomena is entry into competent state in the model organism *Bacillus subtilis* during early stationary phase, when a subset of cells, typically 10% - 20% of the total population, acquires the ability to uptake exogenous DNA as an adaptation to environmental stresses (4, 5). Fluctuations in the expression of ComK, the master competence transcriptional regulator, result in a bistable activation of multiple pathways that direct a subpopulation of cells to become competent (4, 5). Based on studies of ComK-dependent regulations, researchers have further found that the same cells that acquire competence might also exhibit differential expression patterns regarding cell division, cell shape maintenance, detoxification, central carbon metabolism, pH homeostasis, etc. (6, 7). The changes in diverse aspects of cellular physiology in addition to competence development distinguish the competent cells from their vegetative counterparts and spores, defining them as under a distinct cell state called K-state (6, 8). However, comprehensive characterizations of K-state and its gene and pathway regulations are yet to be achieved.

Previously, global transcriptional profiling of *B. subtilis* cells grown to K-state was performed via DNA microarray (6, 8). To interpret K-state-specific changes in gene expression, the abundance of respective mRNAs was directly compared between the wild type and the *comK* knockout strains, both under K-state-inducing conditions. Although around 150 genes were

reported to be under the control of ComK and upregulated in K-state cells, the resolution was significantly impaired due to high transcriptional background imposed by non-K-state cells which account for 80% - 90% of the entire population. The low abundance of mRNAs in cells further limited the ability of detecting transcripts from downregulated genes (6, 8). Moreover, the knockdown of *comK* inevitably disrupted the native cell state, biased the detection against genes that are affected but not completely dependent on ComK (6, 8). Therefore, probing changes in gene expression in a minor subpopulation, as in the case of *B. subtilis* K-state cells, is challenging to global profiling techniques.

Here, we describe the use of BONCAT to study proteomic response of *B. subtilis* cells as a result of entering K-state with spatiotemporal resolution. *B. subtilis* cells are engineered to express NLL-MetRS adapted from *Escherichia coli* machinery to incorporate Anl into protein synthesis (9). The promoter for the endogenous *comF* operon, required for DNA uptake, is adapted to regulate the synthetase expression, tightly restricting protein labeling within the subpopulation. We compared the proteomic profiles of K-state cells and the entire population and analyzed the dynamic proteomic response of the K-state subpopulation.

3.4 Results

3.4.1 P_{comF}*:nll-metRS* enables K-state-selective labeling in a heterogeneous population

To selectively label proteins in *B. subtilis* K-state cells through BONCAT, we first confirmed that NLL-MetRS can be expressed in W168 strain to efficiently incorporate Anl in protein

synthesis. The synthetase was controlled under a strong constitutive promoter P_{veg} and integrated into W168 genome (Figure 3.S1A). Cells were grown in rich medium and pulse-labeled with 1 mM Anl for 1 h during late-exponential phase. Cell lysates were treated with TAMRA-alkyne for detection and visualization of Anl incorporation via click chemistry. SDS-PAGE in-gel fluorescence of the treated lysates showed high labeling efficiency in *B. subtilis* W168 strain (Figure 3.S1B). K-state in W168 was induced following a two-step protocol and optimizations in growth conditions were performed (Figures 3.S2 and 3.S3) (10). K-state induction was maximized by growing cells in starvation medium 1 (SM1) for 4 h before diluting to fresh starvation medium 2 (SM2) and culturing for another 2 h (Figure 3.1A). The time point for dilution is defined as t_0. Using the optimized conditions, we pulse-labeled the culture of P_{comF}*:nll-metRS* strain, which expresses NLL-MetRS under the control of the endogenous *comF* promoter, with 1 mM Anl for 1 h (Figure 3.1B). Cell lysates were treated with TAMRA-alkyne and in-gel fluorescence imaging showed strong labeling (Figure 3.1C). To confirm accuracy in protein labeling, 1 mM Anl was added to P_{comF}*:nll-metRS* culture 2 h prior to t_0 and incubated for 1 h, during which cells were in mid-exponential phase and should not yet enter K-state. As a control, wild type W168 cells were grown and labeled in the same way both before and after t_0. Compared to after t_0, TAMRA signal from P_{comF}*:nll-metRS* samples labeled before t_0 was hardly detectable and was comparable to the W168 sample, suggesting negligible basal expression of NLL-MetRS in non-K-state cells and hence sufficient accuracy in protein labeling (Figures 3.1C and 3.S4A). Consistent with the trend in Anl incorporation, western blotting against 3×FLAG tagged to the C-terminal of NLL-MetRS verified the presence of the mutant synthetase only after cells began to enter K-state (Figures

3.1C and 3.S4A). The same cell lysates of P_{comF}*:nll-metRS* and W168 samples were used to enrich for labeled proteins which were subsequently analyzed with LC-MS/MS. The total LC-MS/MS ion intensities of P_{comF}*:nll-metRS* samples labeled after t_0 were around 6- to 18-fold higher than those labeled before t_0 which had similar level of intensity as the W168 sample (Figure 3.S4B). The relatively low level of background intensities should prove the enrichment to be efficient and have minimal effect on our results. The P_{veg}*:nll-metRS* strain, which expresses NLL-MetRS constitutively in the whole population regardless of cell states, was used to serve as a positive control. As expected, samples labeled before and after entry into K-state were both extensively labeled by Anl (Figure 3.S5). We also confirmed that the genetic manipulations do not change growth pattern, ComK expression, or genetic competence in the strains (Figures 3.S6 and 3.S7). Full LC-MS/MS results are provided in Data Set B.1.

3.4.2 Visualization of K-state labeling

To visualize the trend of P_{comF} activity over time, we expressed mRFP under the control of the promoter (P_{comF}*:mrfp*, Figure 3.S8A). Cells were grown under the same K-state inducing conditions and imaged over time. The appearance of fluorescent cells only after t_0 further indicated minimal activity of P_{comF} before cells entering K-state (Figures 3.S8B-S8G). W168 and P_{veg}*:mrfp* strains were used as controls (Figures 3.S8A, 3.S8H-S8S). The consistent strong fluorescence of P_{veg}:*mrfp* over time validated P_{veg} activity to be spatiotemporally non-differential.

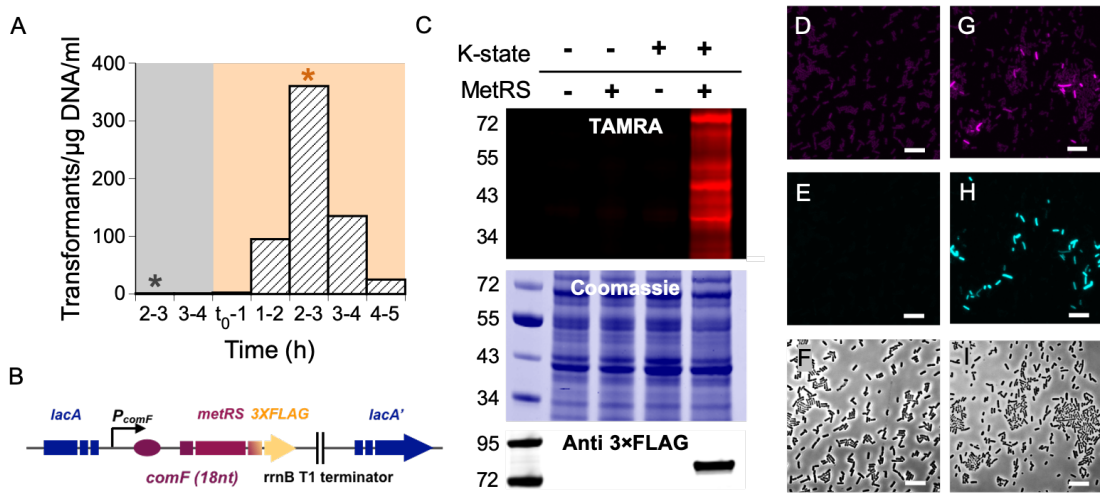

Figure 3.1 K-state-selective proteomic labeling of *P_{comF}:nll-metRS* strain. (A) Number of transformants per mL per µg purified plasmid DNA during the course of competence induction. Cells from saturated culture were inoculated in SM1 (grey). T_0 indicates the time of dilution into fresh SM2 (orange). For non-K-state control, cells were labeled with Anl from 2 to 3 h in SM1 (grey asterisk). Proteomic samples for K-state-selective labeling were prepared with cells grown in SM2 from 2 to 3 h after t_0 when maximum competence was reached (orange asterisk). (B) Genetic construct of *P_{comF}:nll-metRS* integrated into the *lacA* locus of W168 genome. (C) Visualization of Anl incorporation in W168 and *P_{comF}:nll-metRS* strains before and after entry into K-state by click chemistry with TAMRA-alkyne and in-gel fluorescence imaging. Coomassie staining of the same protein gel verified equal protein loading. Western blotting was applied to probe the 3×FLAG tag on NLL-MetRS. (D-I) Representative confocal laser scanning images of *P_{comF}:nll-metRS_ComK-mWasabi* strain clicked with DBCO-TAMRA: (D-F) before entry into K-state (labeled during 2 to 3 h in SM1); (G-I) after entry into K-state (2 to 3 h in SM2). (D, G) TAMRA fluorescence (magenta), (E, H) mWasabi fluorescence (cyan), and (F, I) phase contrast were acquired in separate tracks. (Scale bar: 10 µm.)

To visualize the labeling of K-state subpopulation and to further confirm that the labeled cells were under the desired state, we fused *mWasabi* 3' to *comK* (with both the promoter and the

coding sequence) and integrated the gene cassette into the P_{comF}*:nll-metRS* strain genome (Figure 3.S9A). Cells were grown under the K-state inducing conditions and labeled with 1 mM Anl before and after t_0 as described before. Cells were fixed and permeabilized, treated with DBCO-TAMRA (Figure 3.S10), and imaged with confocal microscopy. No mWasabi or TAMRA fluorescence significant than background was detected in cells labeled prior to t_0, while a subpopulation of the mWasabi-expressing cells was labeled by Anl after t_0 (Figures 3.1D-3.1I). We also integrated the gene cassette into the wild type W168 and the P_{veg}*:nll-metRS* strains which were cultured and treated in the same way as the negative and positive control, respectively (Figures 3.S9B-3.S9M).

3.4.3 Proteomic profiling in B. subtilis K-state subpopulation via BONCAT

To systematically investigate proteomic response as cells stochastically entering K-state, P_{comF}*:nll-metRS* and P_{veg}*:nll-metRS* (each in triplicates), along with the wild type W168 strain, were grown under K-state inducing conditions and treated with Anl 2 h after t_0 as described in previous section. Anl incorporation was visualized via click chemistry and in-gel fluorescence imaging (Figure 3.2A). Western blotting against 3×FLAG verified synthetase expression in the samples (Figure 3.2A). Unlike P_{veg}*:nll-metRS*, the extent of Anl incorporation in P_{comF}*:nll-metRS* samples differed among the three replicates, as indicated by the dissimilar TAMRA intensities. However, because of the stochastic fluctuation in ComK expression, percentage of cells entering K-state varies among populations (5). The extent of P_{comF} activation within each K-state cell, which directly controlled by ComK, should also be largely affected by the fluctuation. As a result, it is reasonable that the extent of protein labeling in P_{comF}*:nll-metRS*

varied from sample to sample. Anl-labeled proteins were enriched via BONCAT and analyzed by LC-MS/MS. Intensity Based Absolute Quantification (iBAQ), proportional to the absolute molar quantities of the identified proteins within each sample, was obtained (11). Label-free quantification (LFQ) with normalization was calculated as an indicator of the relative amount of each identified protein across all the samples (12). The normalization was such that the median relative abundance from each sample was approximately the same. In total, 2157 proteins were identified, among which 1261 were shared by the two strains; 446 and 450 proteins were exclusively identified in P_{veg}*:nll-metRS* and P_{comF}*:nll-metRS* samples, respectively (Figure 3.2B). Spearman's rank correlation coefficients calculated from iBAQ values of each MS run indicated high correlations among biological replicates within each strain (Figure 3.2C). Principal component analysis (PCA) was performed using the normalized LFQ values to visualize the variance among the samples (Figure 3.2D). According to PCA, proteomic profiles from replicates of the same strain had relatively low level of variance while those obtained from different strains were highly distinct. Therefore, as supported by the Spearman's rank correlation coefficients and the PCA, proteomic profiles obtained from replicates largely resembled each other despite the drastic difference in the extent of labeling, implying that the resolution of cell-state-selective protein identification provided by BONCAT is generally unaffected by the abundance of cells in the state of interest. The relative abundance (i.e. relative fold change) of newly synthesized proteins identified in both of the strains (P_{comF}*:nll-metRS*/P_{veg}*:nll-metRS*) and the corresponding FDR adjusted *p*-value were calculated using normalized LFQ values (Figure 3.2E). Full proteomic results are in Data Set B.2.

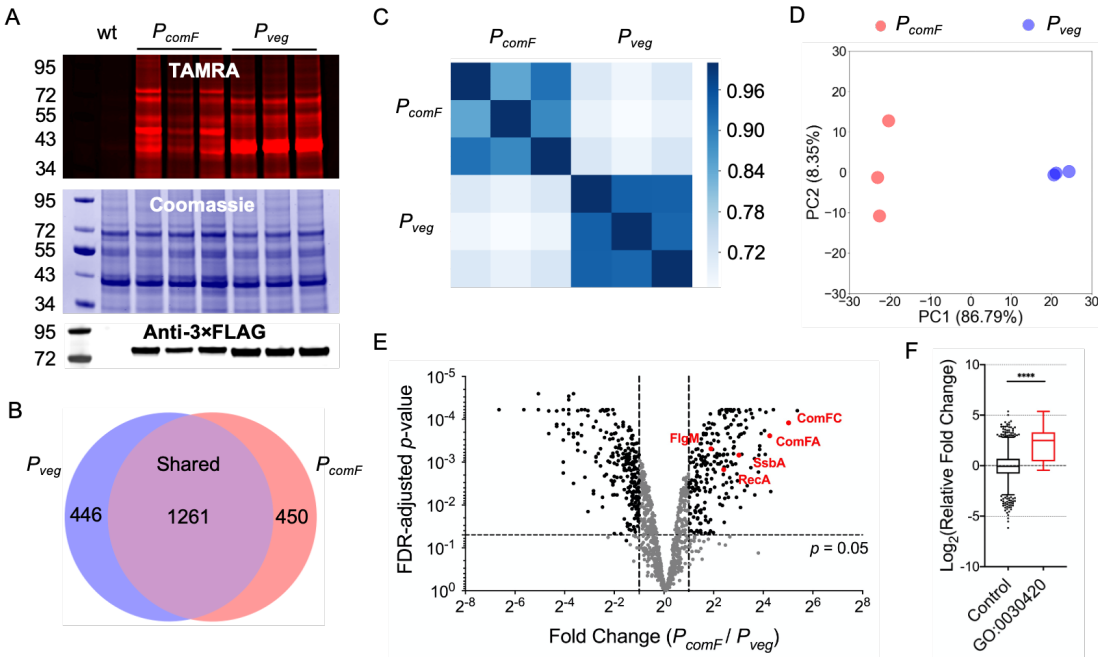

Figure 3.2 Selective proteomic labeling of *B. subtilis* K-state subpopulations. *P_{comF}:nll-metRS* strain was used for labeling of K-state cells and *P_{veg}:nll-metRS* for the whole population. (A) Visualization of Anl-labeling by treating cell lysates with TAMRA-alkyne. Coomassie staining of the same gel verified equal protein loading. Western blotting was applied to probe the 3×FLAG tag on NLL-MetRS. (B) Venn diagram of the number of proteins identified in *P_{comF}:nll-metRS* and *P_{veg}:nll-metRS* samples. (C) Spearman's rank correlation coefficients of iBAQs for each sample to show correlations among the replicates of each strain. (D) Principal component analysis (PCA) using normalized LFQs to show variance among samples. (E) Quantification of relative abundance and the corresponding FDR adjusted *p*-value for each shared protein. (F) The Gene Ontology biological process establishment of competence for transformation (GO: 0030420) was significantly upregulated in K-state subpopulations compared to the control.

3.4.4 Pathway regulations in B. subtilis K-state subpopulation

Because competence development is one of the signatures for K-state, the relative abundance of proteins encoded by genes categorized in the Gene Ontology (GO) term "establishment of competence for transformation" (GO: 0030420) was analyzed and used to verify the validity

of our proteomic profiles (13). With 52 genes in this GO term, 29 were identified in both the strains and 4 exclusively in P_{comF}:*nll-metRS* (at least in two replicates). There were only 3 proteins uniquely identified in P_{veg}:*nll-metRS* (at least in two replicates), of which one was a repressor for competence establishment (ComZ) and the other two were involved in multiple GO biological processes in addition to competence establishment (XlyA, AppA) according to the GO database (14). The relative abundance of 27 shared proteins with the annotation (MecA and MecB having negative regulatory function were excluded) was compared to the control where all the proteins shared by the two strains were included (Figure 3.2F). Because LFQ values were normalized as described above, median log2 relative fold change of the control was close to 0. Based on the Mann-Whitney *U* test, this GO biological process was significantly upregulated compared to the control, consistent with our expectation (15). It is worth noting that the 3 proteins encoded by the *comF* operon controlled by the promoter P_{comF} were either highly upregulated (ComFA, ComFC, Figure 3.2E) or only identified in P_{comF}:*nll-metRS* (ComFB), supporting that our genetic design of P_{comF}:*nll-metRS* to selectively label K-state subpopulations is appropriate. Full GO data are provided in Data Set B.3.

The above observations encouraged us to further investigate on how the regulations in K-state cells different from the whole population. The shared proteins were grouped based on Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathway database and pathway regulations within the subpopulation were analyzed (16). Relative abundance of proteins with a certain annotation was compared to the control as described above (Figure 3.3). Based on the Mann-Whitney *U* test and the comparison, pyrimidine metabolism, DNA replication, DNA repair and

ABC transporters pathways had higher expression levels in the subpopulation while ribosome, bacterial chemotaxis and flagella assembly pathways were significantly downregulated. Full KEGG Pathway analysis results are provided in Data Set B.4.

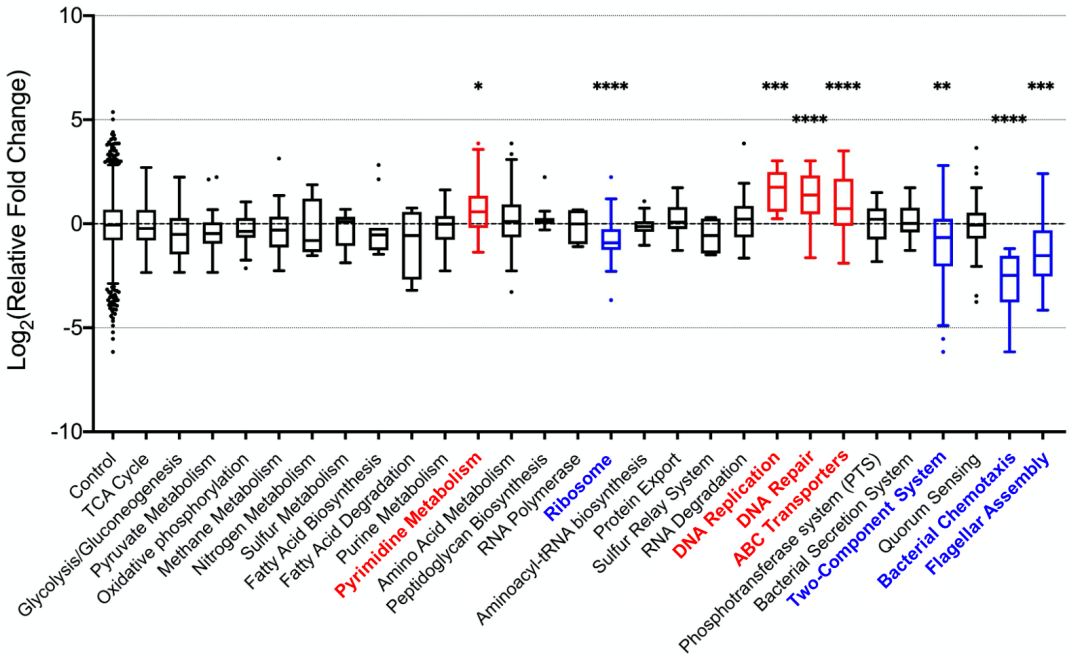

Figure 3.3 Pathway regulations in K-state subpopulations. Proteins identified in both the strains were categorized based on Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathway database. Tukey's box plot was used to show the relative abundance of proteins for each annotated pathway, with each box extended from the 25th to the 75th percentiles of the distribution. Mann-Whitney *U* test was employed to determine if the expression of a certain pathway was significantly different from the control (*, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.0001$). Pathways significantly upregulated were annotated in red while those significantly downregulated were in blue.

Among those significantly upregulated pathways, activation of DNA repair is characteristic of K-state cells in *B. subtilis* (17, 18). There is a hypothesis that the evolvement of competence

mechanism in *B. subtilis* is actually from the need for DNA damage repair, which has been supported by many studies (19-21). In addition to DNA repair, it is reported that, analogous to the *E. coli* SOS response, *B. subtilis* has an SOS system that is induced following the development of K-state to enhance DNA repair and mutagenesis (22-25). This SOS system is negatively regulated by LexA, a transcriptional repressor of the SOS regulon, while positively regulated by the multifunctional SOS repair factor RecA (25, 26). During K-state activation, ComK alleviates the repression effect of LexA on the expression of RecA, which in turn activates transcription of genes involved in the SOS system (27). We found that RecA was more than 5-fold more abundant in the K-state cells (Figure 3.2E), and proteins involved in SOS response (GO: 0009432) were either upregulated or exclusively identified in the subpopulation, consistent with the proposed regulation mechanism of SOS response in K-state cells. Full GO data are provided in Data Set B.3.

ABC transporters were also significantly upregulated in the subpopulation. Within the pathway, most of the upregulated proteins are involved in the transport of amino acids, peptides, and vitamins while two (TagG and TagH) are for teichoic acids (TAs) transport. Because *B. subtilis* cells enter K-state as a strategy to cope with multiple stresses including nutrient limitation, it is within expectation that the expression of transporters for nutrients and various building blocks was upregulated (28). For the two TAs transporters, a recent study has suggested that inhibiting TagG and TagH could result in significant decrease in transformation efficiency (29). Consistent with their study, we found TagG to be highly upregulated and TagH also slightly upregulated in the K-state subpopulation. Proteins involved in the biosynthesis of TAs were

also found to be highly abundant in the subpopulation, which will be discussed in more detail in the following section.

Interestingly, our results showed that the pyrimidine metabolism and DNA replication pathways were significantly upregulated compared to the control. Pyrimidine metabolism can provide building blocks for cells to synthesize DNA and RNA. The upregulation of both pathways indicates that K-state cells are active in DNA processing. *B. subtilis* K-state cells are supposed to be under growth arrest, and it is suggested that DNA replication and cell division are blocked (28, 30, 31). However, proteins involved in DNA replication pathway, which are mainly DNA polymerases, ribonucleases, and ligases, are not only responsible for genome replication but also actively involved in the processing of exogenous DNA (18). In fact, the most highly upregulated gene in the DNA replication pathway is *ssbA*, encoding for a single-stranded DNA-binding protein, that is directly involved in the exogenous DNA uptake machinery (Figure 3.2E) (32). Therefore, these upregulations, instead of being treated as a sign of active cell division, should be seen as a proactive strategy for K-state cells to make use of exogenous DNA and survive under various stresses.

Even though the amino acid metabolism pathway as a whole was not differentially regulated from the control, the valine, leucine and isoleucine biosynthesis pathway was highly activated (Figure 3.S11). These three amino acids are branched-chain amino acids (BCAAs) which can, along with GTP, activate the global regulator CodY (36). CodY mainly have a repressive effect on the transcription of various genes and operons expressed in early stationary phase, including

genes involved in BCAAs synthesis (34-39). The intracellular level of GTP, which drops with cell growth rate decrease, regulates the activity of CodY (40-43). Therefore, during the Anl labeling time, decrease in CodY activity due to GTP deficiency could lead to the derepression of the BCAAs biosynthesis pathway. Currently, the roles of BCAAs in K-state cells is poorly understood. Aside from the involvement in CodY regulation, BCAAs can be metabolized to produce the precursors for the synthesis of branched-chain fatty acids (BCFAs) which are important components of cell membrane to adjust membrane fluidity (44-46). However, the BCAAs metabolism pathway was not upregulated compared to the control (Figure 3.S11). The accumulation of BCAAs suggests that it may have other unknown functions important for K-state cells.

Among the significantly downregulated pathways, the negative regulation of both bacterial chemotaxis and flagellar assembly is an indication of cell motility repression. There is evidence showing that the transcription from the *comF* into the *flgM* operon is positively regulated by ComK (47). FlgM is an anti-sigma factor to inhibit σ^D, thus inhibiting σ^D-dependent genes involved in cell motility (48, 49). Consistent with previous studies, we found that FlgM was 3.7-fold more abundant in the subpopulation, resulting in the repression of σ^D-dependent genes from both bacterial chemotaxis and flagellar assembly pathways (Figure 3.2E). Ribosome pathway was also found to be significantly downregulated, which was likely due to the growth arrest in K-state cells.

3.4.5 Altered expression in genes and gene clusters with shared physiological functions

In addition to system-level regulations, we observed changes in expression of particular genes and gene clusters that might also contribute to the distinct phenotypes of K-state (Table 3.1). Consistent with previous studies, we found upregulation in proteins responsible for arginine biosynthesis (6, 8). Except for ArgG and ArgI, proteins encoded by *arg* genes were more than 2-fold upregulated in K-state cells. Interestingly, the fact that only the biosynthesis of arginine was enhanced but not its metabolism raised questions about the cellular destination and function of the built-up arginine. The K-state subpopulation also engaged in specialized metabolism compared to the rest of the cells. For example, the succinyl-CoA synthetase (beta subunit) SucC, which was found to be 6.5-fold upregulated, was recently related to acetate production in K-state cells, though its physiological role is not yet clear (7).

Our data suggested substantial genetic reprogramming regarding cell wall synthesis and organization in K-state as a response to nutritional stresses. Significant upregulation was observed for the *tag* operon, which regulate the synthesis of TAs. TAs are one of the main components of *B. subtilis* cell wall and are traditionally seen as important regulators for cell morphology and division, autolytic activity, ion homeostasis, and defense against antibiotics (50, 51). Recently, however, TAs are also found to be crucial in the development of competence in *B. subtilis* (29). Evidence suggests that wall teichoic acids, one of the two types of TAs that covalently attached to peptidoglycan, are enriched in competent cell wall and involved in mediating exogenous DNA binding (29). Our data showed consistency with this study regarding the importance of TAs in K-state subpopulations. K-state cells also directed more

Table 3.1 Summary of genes with shared cellular functions that are differentially regulated in K-state subpopulation. The corresponding proteins either exclusively identified in the subpopulation (in at least two replicates) or significantly upregulated (relative fold change > 2, *p*-value < 0.05) are in orange; those either exclusively identified in the P_{veg}*:nll-metRS* samples (in at least two replicates) or significantly downregulated (relative fold change < 0.5, *p*-value < 0.05) are in grey.

Protein Function	Name
Cell division	
Cell division arrest	*yneA* *maf* *minC*, *minD*, *minJ*
Division septum assembly	*divIVA* *sepF*
Cell wall biosynthesis and organization	
Teichoic acids and peptidoglycan biosynthesis	*tagA, tagB, tagC, tagE, tagF, tagO, tagT, tagV* *ykfA, ykfB, ykfC, ykfD* *dltB, dltC* *lytE*
Cell shape determination	*mreB, mreC* *pbp, pbpC, pbpE*
Detoxification	
Multidrug efflux transporter/permease	*yxkD* *yoeA* *yhbJ* *natB* *mdr* *yknX* *yvrP*
Multidrug efflux transcriptional activator	*mta*
Antibiotics biosynthesis	
Bacilysin biosynthesis	*bacA, bacB, bacC, bacD, bacE, bacF, bacG*
polykedite biosynthesis	*pksE, pksL, pksM, pksN, pksR*
Sporulation delaying killing factor biosynthesis	*sdpA, sdpC*

resources towards the synthesis and organization of other cell wall components: DltB and DltC, involved in D- alanylation of lipoteichoic acid, were both significantly upregulated; LytE, a cell wall hydrolase, was 2.6-fold more abundant in the K-state subpopulation. (52, 53). Notably, major cell shape determining proteins are significantly upregulated: MreB and MreC together form dynamic helical filaments that are critical for proper cell elongation (54); the penicillin

binding proteins Pbp, PbpC, and PbpE are involved in peptidoglycan crosslinking, affording structural integrity to living cells (55).

K-state-selective readjustments in gene expression related to cell division and chromosome segregation were observed. DivIVA, a membrane binding protein that recruits cell division machinery at division septum, was found to be more than 3-fold less abundant in the subpopulation (56). MinC and MinD, which together inhibit polar Z-ring formation, and thereby repressing cell division, were both more than 3-fold upregulated in K-state cells (57, 58). The nucleoid occlusion protein Noc, which inhibits cell division by preventing septum formation through the nucleoid, was also found to be slightly upregulated (1.8-fold) (57).

The expression of efflux pumps for detoxification and antibiotics synthesis was elevated in K-state cells. A notable group of proteins that were upregulated, though many were poorly characterized, is involved in the synthesis of efflux pumps that are usually membrane proteins extruding cytotoxic compounds out of cells (59, 60). Besides active detoxification, significant upregulation was observed for the expression of *bac* and *pks* operons. These two operons encode enzyme complexes that catalyze the biosynthesis of the dipeptide antibiotics bacilysin and the polyketide bacillaene respectively, both of which are part of the defense systems against predators (61-64). In addition, cells in K-state were active in cannibalism by increasing the expression of the sporulation delaying protein SdpC (14-fold upregulated) to collapse membrane proton motive force in their siblings for extra nutrients (65, 66). The regulations of efflux pumps genes in K-state cells can be seen as a possible strategy to outcompete their

neighbors, both siblings and other species, to maximize their survival rate.

3.5 Discussion

In this study, we demonstrated that the selective labeling and identification of newly synthesized proteins in a minor subpopulation of *B. subtilis* cells that stochastically entering K-state can be achieved by BONCAT. Previously, study of K-state was often performed using global transcriptional profiling which resulted in relatively low resolution as the transcriptional background was usually very high, limiting the information obtained (6, 8). Moreover, as recent studies have shown that the squared Pearson correlation coefficients between mRNA and protein levels are only around 0.4, the relative abundance of mRNAs in cells does not necessarily reflect the abundance of proteins, thus the translational regulations are usually hidden under transcriptional profiling (67, 68). Here, the adapted BONCAT technique largely resolved the problems. The K-state selective labeling and enrichment efficiently separated targeted proteins from rest of the protein pool, highly increased the resolution of LC-MS/MS in detecting proteins from cells of interest. As a result, we could identify over 2000 proteins among our samples, providing us a more detailed depiction of *B. subtilis* K-state. We found 396 proteins that were either exclusively identified (in at least two P_{comF}*:nll-metRS* replicates) or significantly upregulated (relative fold change > 2 and FDR adjusted p-value < 0.05) in K-state cells. More importantly, while genes repressed in K-state had been poorly documented due to extremely low abundance of mRNAs, we identified 432 proteins that were either exclusively found in the non-selective pool (at least in two replicates) or significantly

downregulated in K-state cells. In the meanwhile, as we applied proteomic profiling to our system, the concern over the low correlation between transcription and translation was extensively alleviated as proteomic profiling could well capture post-translational regulations and degradations in the system.

Our study has provided reliable and genuine results. The upregulation of DNA repair and ABC transporters mainly for nutrient scavenging, and the downregulation of ribosome, bacterial chemotaxis, and flagella assembly are highly consistent with previous studies. We also provided evidence for genetic reprogramming regarding cell division, cell wall synthesis, detoxification, and antibiotic synthesis that are characteristic of K-state cells. Our proteomic profiles showed consistency with the two most recent mRNA profiling studies: among the 36 genes identified as upregulated in both of the studies, 31 were found to be more than 2-fold upregulated in our profiles, with the remaining 5 unidentified (6, 8).

Our results provide key new insights into the development of K-state. Many of the proteins highly upregulated in the subpopulation, with known functions in other cell states, have not been correlated with K-state before. Our results suggest that these genes might have alternative K-state related functions yet to be identified. For example, it is the first time that BCAAs biosynthesis is correlated with K-state cells. Even though we found that the decrease in CodY activity could be a likely explanation on the upregulation, currently there is limited understanding of the exact regulations and roles of these BCAAs in K-state cells. A similar situation is applied to the arginine accumulation: there lacks a definite answer to the

upregulations even though multiple studies have also observed the phenomenon (6, 8). Spo0M, which was thought to be a sporulation protein, was recently found to be expressed also during vegetative growth and play a role in cell division (69). The upregulation of *spo0M* in this study could suggest yet other unknown function of the gene in K-state. Profound genetic readjustment was also observed for many genes that are completely uncharacterized. Although the exact biological functions of these genes and hence their contribution to the differentiated phenotypes in K-state are not addressed in this study, our proteomic results provide potential directions for further investigation into the development of K-state.

Cell-state-selective BONCAT should prove to be a powerful technique in studying translational regulations with high spatiotemporal resolution, especially for a phenotypically distinct subpopulation arising from stochastic gene expression as demonstrated in this study. The bioorthogonality of the technique can largely keep targeted systems intact, minimizing the perturbance and maximizing the accuracy of profiling results. As BONCAT uses the genetic machinery shared by all organisms, simultaneous proteomic labeling and comparison of various states or species in a multicellular community can also be achieved by employing multiple bioorthogonal mutant aaRSs controlled by different cell-selective promoters. Given the ubiquity of heterogeneity in biological systems, BONCAT is readily applicable in various contexts to efficiently target cells of interest for proteomic studies.

3.6 Materials and Methods

3.6.1 Strain construction

All genetic manipulations were done by chromosomal integration into the *lacA* or *amyE* locus via the integration vector pBS2E or pBS1C, respectively (70). Genetic construction was performed in *E.coli* DH10B (Zymo Research). Purified plasmid DNA with confirmed sequence (Laragen) was linearized by digesting with *BsaI* (New England Biolabs) before used to transform into *B. subtilis* strains following previous protocol (71). Transformants containing correct chromosomal insertion were selected on solid medium supplemented with appropriate antibiotics, either 1 μg/mL erythromycin (Erm) or 5 μg/mL chloramphenicol (Cm).

nll-metRS was amplified from plasmid BMB8 (72), using primers YH259 and YH260, and digested with *BsaI* and *XhoI*. The promoters of *comF* and *comK* (along with the complete UTR region and the first 18 coding base pairs) ware amplified from W168 chromosome using primer pairs YH225-226 (digested with *EcorI* and *BsaI*) and YH235-236 (digested with *HindIII* and *BsaI*). Sequence of P_{veg} along with the ribosome binding site of *SpoVG* was amplified by PCR using primers YH110 and YH119 and treated with *EcoRI* and *BsaI*. Fragments of each promoter were then ligated with the *nll-metRS* fragment, which then ligated with a 3×FLAG fragment containing *XhoI* and *SpeI* overhang by Golden Gate. Each of the ligated piece was then inserted into vector PBS2E to generate the final plasmid. To construct the mRFP expressing strains, the DNA fragment of codon-optimized *mrfp* was amplified with primers YH120 and YH162, digested with *EcoRI* and *BasI*, and ligated into vector PBS2E in a similar manner. For strains

with translational fusion of mWasabi and ComK, the DNA fragment of *comK*, including the promoter and the entire coding region, was obtained by PCR with primers YH235 and YH 278. The codon-optimized sequence of *mWasabi* was amplified with YH152 and YH279. The two fragments were then ligated and inserted into vector PBS1C. All restriction enzymes were purchased from New England Biolabs (NEB). All primers and gene blocks mentioned above were synthesized by Integrated DNA Technologies (IDT). Detailed information is listed in Table S1.

3.6.2 K-state Induction in B. subtilis

K-state in *B. subtilis* is induced following previous protocol with minor modification (10). Pre-cultures of *B. subtilis* strains were prepared by inoculating a single colony grown on antibiotics-supplemented agar plate in 5 mL Lysogeny broth (LB) at 37 °C under agitation (250 rpm) and growing for 16 h. To induce K-state in *B. subtilis*, saturated LB culture was diluted 1:25 into freshly prepared starvation medium 1 (SM1) and grown at 37 °C for 4 h under agitation (250 rpm). The culture was immediately diluted with an equal volume of starvation medium 2 (SM2) and further grown for 2 h to reach maximum competence.

3.6.3 Transformation assay

B. subtilis strains were grown to competence as previously described, 500 µL aliquots were withdrawn over the course of competence development, mixed with approximately 1 mg of purified plasmid DNA, incubated at 37 °C for 30 minutes, and further rescued with 300 µL of LB for 30 minutes before a known volume was plated on an LB agar plate supplemented with

chloramphenicol (5 μg/mL). The transformation efficiency (number of transformants per mL culture per μg DNA) was then back calculated from the number of isolated colonies after incubation at 37 °C overnight. All plasmid DNA was purified from *E. coli* DH10B (Zymo Research).

3.6.4 Confocal imaging

To image the expression of mRFP, *B. subtilis* strains were grown under K-state inducing conditions. 1 mL aliquots were withdrawn over the course of K-state development. Cells were harvested, resuspended in 4% paraformaldehyde in 1×PBS to reach an OD_{600} value of around 2, and incubated for 15 minutes under room temperature. Fixed cells were washed three times with 1×PBS and stored at 4 °C. 5 μL of cells were spotted onto 2 w/v % agarose pad for imaging. Samples were imaged by Zeiss LSM 800 with a 100× PC Plan-Apochromat lens. To detect mRFP expression, a 561 nm laser line (5.0%) was used for detection. Images spanning 5 μm in z-direction were taken as 25 stacks, with 0.2 μm per stack. 2D reconstruction was done using ImageJ 64-bit (v 2.0.0) (73).

For mWasabi-TAMRA imaging, *B. subtilis* strains were either cultured in SM1 medium for 2 h or cultured in SM1 medium for 4 h and diluted with same volume of SM2 medium and cultured for another 2 h. Cell cultures were incubated with 1 mM Anl (Iris Biotech) by dilution of 100 mM Anl stock (dissolved in double-distilled water [pH 7]) into the cell cultures for 1 h at 37°C with agitation (250 rpm). Cells were pelleted at 4°C, fixed with 4% paraformaldehyde/PBS for 15 min, and washed three times with 1×PBS. Cells were

permeabilized with 50%, 80% and 96% ethanol in sequence each for 3 min, washed with 100 mM Tris/HCl (pH 7.4), and incubated 1 h with 100 mM 2-chloroacetamide in 100 mM Tris/HCl (pH 7.4) at 46℃ in the dark to block free thiols with end-over-end rotation. DBCO-TAMRA (Click Chemistry Tools) was directly added to this solution to reach a final concentration of 100 nM. The click reaction was carried out at 46℃ in the dark for 30 min. To remove unbound DBCO-TAMRA, cells were washed with 1×PBS, incubated with 1×PBS for 10 min at 48℃ with rotation, incubated with 50% DMSO/PBS for 20 min in dark with rotation at room temperature, and washed with 1×PBS for three times. For all the above procedures, each washing step was followed by pelleting samples via centrifugation for 2 min at 16000 *g*. Cells were then concentrated and immobilized on 2 w/v % agarose pad topped with cover glass. Samples were imaged by Zeiss LSM 800 with a 100× PC Plan-Apochromat lens. For simultaneous imaging of TAMRA and mWasabi fluorescence, a 561 nm laser (4.0%) and a 488 nm laser (3.0%) were sequentially activated in two separate tracks. The images were processed with ImageJ.

3.6.5 BONCAT labeling and enrichment

For protein labeling, after cells reached maximum competence, cell cultures were labeled with 1 mM Anl for 1 h. Cells were pelleted at 4℃, resuspended in lysis buffer (4 w/v% SDS/PBS), and treated with protease inhibitor (cOmplete-, Mini, EDTA-free Protease Inhibitor Cocktail; Roche). To fully lyse the cells, lysates were heated at 95℃ for 10 min, sonicated with a microtip probe for 1.5 min with an amplitude of 30% and pulse 5 s on/5 s off (Qsonica), and heated at 95℃ for another 10 min. Lysates were clarified by centrifuging at 10000 *g* for 15 min. Protein

concentration of each lysate was quantified using BCA kit (Pierce).

To visualize Anl incorporation, cell lysates were incubated in dark with 15 μM alkyne-TAMRA, 250 μM $CuSO_4$, 1.25 mM THPTA, 5 mM aminoguanidine hydrochloride, and 5 mM sodium ascorbate for 1 h at room temperature. 10 μg of each treated lysate was loaded onto a protein gel (NuPAGE Novex 4-12% Bis-Tris Protein Gels; Thermo Fisher Scientific) and separated via SDS-PAGE. Gels were fixed and destained with 40% methanol, 10% acetic acid, and 50% DI water overnight, rehydrated with DI water for 1 h and imaged on a Typhoon Trio Variable Mode Imager (GE Healthcare). The gels were stained with InstantBlue Coomassie protein stain (Expedeon) to verify equal protein loading. For NLL-MetRS detection using Western blotting, each crude lysate containing 15 μg of protein was loaded and separated by SDS-PAGE, transferred to nitrocellulose membrane (GE Healthcare), conjugated with monoclonal ANTI-FLAG M2 antibody produced in mouse (Sigma Aldrich) and Goat anti-Mouse IgG (H+L) highly cross-adsorbed secondary antibody Alexa Fluor 647 (Thermo Fisher Scientific), and imaged with the Typhoon imager.

For BONCAT enrichment, each lysate containing 3 mg protein was incubated in dark with 100 mM chloroacetamide for 30 min at 65°C with shaking (1200 rpm) to block the free thiol group in cysteines. Equal volume of freshly made 8 M urea/0.85 M NaCl/PBS was added to each lysate. 25 μL of DBCO agarose resin 50% slurry (Click Chemistry Tools) was washed three times with 0.8% SDS/PBS (wt/vol), resuspended in original volume with 0.8% SDS/PBS, and added to each lysate. The resins were incubated in dark for 16 h at room temperature with end-

over-end rotation, washed with double-distilled water, and treated with 10 mM dithiothreitol (DTT)/0.8% SDS/PBS in dark for 15 min at 70 ℃, and incubated with 40 mM chloroacetamide in dark for 30 min at room temperature. Resins were transferred to Poly-prep chromatography columns (Bio-Rad) and extensively washed with the following solutions (i) 8 × 5 mL 0.8% SDS/PBS; (ii) 8 × 5 mL 8 M urea in 100 mM Tris pH 8.0; and (iii) 8 × 5 mL 20% acetonitrile (ACN) in water (vol/vol). During the second wash of each solution, the resins were incubated with the corresponding washing solution for either 10 min (solution i), or 30 min (solution ii and iii). Resins were resuspended in 10% ACN with 50 mM Ammonium Bicarbonate (AmmBic). The resins were digested with 50 ng LysC for 4 h followed with 100 ng trypsin for 16 h at 37℃. The supernatant containing the digested peptides was collected, which were dried and desalted with C_{18} ZipTips (EMD Millipore).

3.6.6 LC-MS/MS

The digested samples were subjected to LC-MS/MS analysis on a nanoflow LC system, EASY-nLC 1200, (Thermo Fisher Scientific) coupled to a Q Exactive HF Orbitrap mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) equipped with a Nanospray Flex ion source. Samples were directly loaded onto a C18 Aurora series column (Ion Opticks, Parkville, Australia). The 25cm x 50µm ID column (1.6 µm) was heated to 45° C. The peptides were separated with a 60 min gradient at a flow rate of 350 nL/min. The gradient was as follows: 2–6% Solvent B (3.5 min), 6-25% B (42.5 min), and 25-40% B (14.5min), to 100% B (1min) and 100% B (12min). Solvent A consisted of 97.8% H2O, 2% ACN, and 0.2% formic acid and solvent B consisted of 19.8% H2O, 80% ACN, and 0.2% formic acid. The Q Exactive HF

Orbitrap was operated in data dependent mode. Spray voltage was set to 1.8 kV, S-lens RF level at 50, and heated capillary at 275 °C. Full scan resolution was set to 60,000 at m/z 200. Full scan target was 3×10^6 with a maximum injection time of 15 ms (profile mode). Mass range was set to 300−1650 m/z. For data dependent MS2 scans the loop count was 12, target value was set at 1×10^5, and intensity threshold was kept at 1×10^5. Isolation width was set at 1.2 m/z and a fixed first mass of 100 was used. Normalized collision energy was set at 28. Peptide match was set to off, and isotope exclusion was on. MS2 data was collected in centroid mode.

3.6.7 Proteomic data analysis

Thermo raw data files were analyzed using MaxQuant (v 1.6.1.0) (74) and were searched against the *B. subtilis* W168 UniProt entries and a contaminant database (247 sequences). For MaxQuant setup, the variable modifications were methionine oxidation and protein N-terminal acetylation; the fixed modification was carbamidomethylation of cysteine; the digestion enzyme was trypsin with up to two missed cleavages. Protein abundances were calculated with MaxLFQ (12) and match between runs was enabled. The relative abundance of the identified proteins among samples and the corresponding *p*-values were calculated using the limma package in R (v 3.5.2). The Benjamini-Hochberg procedure was used to adjust the *p*-values for false discovery. Spearman's rank correlation coefficients calculated using iBAQ values and principle component analysis using log2 LFQ values of proteins identified in all 6 MS runs (n = 938) were performed in Matlab (v R2019a).

3.6.8 Software and database used

Aside from the software mentioned above, we used Prism 8 (v 8.2.1) to generate the bar plot, the volcano plot and the boxplots, and performed the Mann-Whitney *U* test. The Venn diagram was generated by Python (v 3.6.3). The chemical structures were drawn with ChemDraw (v 16.0). Kyoto Encyclopedia of Genomes and Genomics (KEGG) Pathway database for W168 strain was used for system-level regulations analysis (16). Gene Ontology (GO) PANTHER database (v 14.1) was used for genes and biological processes analysis (13, 75).

3.7 Acknowledgements

We acknowledge Roxana Eggleston-Rangel and Brett Lomenick (Proteome Exploration Laboratory) for the technical assistance with liquid chromatography-tandem-mass spectrometry; and Dr. Andre Collazo and Caltech Biological Imaging Center for the training and use of confocal microscope.

3.8 References

1. Losick R, Desplan C (2008) Stochasticity and cell fate. *Science* 320(5872):65-68.
2. Norman TM, Lord ND, Paulsson J, Losick R (2013) Memory and modularity in cell-fate decision making. *Nature* 503(7477):481.
3. Choi PJ, Cai L, Frieda K, Xie XS (2008) A stochastic single-molecule event triggers phenotype switching of a bacterial cell. *Science* 322(5900):442-446.
4. Maamar H, Raj A, Dubnau D (2007) Noise in gene expression determines cell fate in *Bacillus subtilis*. *Science* 317(5837):526-529.
5. Maamar H, Dubnau D (2005) Bistability in the *Bacillus subtilis* K-state (competence) system requires a positive feedback loop. *Mol Microbiol* 56(3):615-624.
6. Berka RM, et al. (2002) Microarray analysis of the *Bacillus subtilis* K-state: genome-wide expression changes dependent on ComK. *Mol Microbiol* 43(5):1331–1345.
7. Rosenthal AZ, et al. (2018) Metabolic interactions between dynamic bacterial subpopulations. *Elife* 7:e33099.
8. Ogura M, et al. (2002) Whole-genome analysis of genes regulated by the *Bacillus subtilis* competence transcription factor ComK. *J Bacteriol* 184(9):2344-2351.
9. Tanrikulu IC, Schmitt E, Mechulam Y, Goddard WA, Tirrell DA (2009) Discovery of *Escherichia coli* methionyl-tRNA synthetase mutants for efficient labeling of proteins with azidonorleucine in vivo. *Proc Natl Acad Sci USA* 106(36):15285–15290.
10. Bennallack PR, Burt SR, Heder MJ, Robison RA, Griffitts JS (2014) Characterization of a novel plasmid-borne thiopeptide gene cluster in *Staphylococcus epidermidis* strain 115. *J Bacteriol* 196(24):4344-4350.
11. Schwanhäusser B, et al. (2011) Global quantification of mammalian gene expression control. *Nature* 473(7347):337–342.
12. Cox J, et al. (2014) Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol Cell Proteomics*13(9):2513–2526.

13. Ashburner M, et al. (2000) Gene ontology: tool for the unification of biology. *Nat genet* 25(1):25.
14. Ogura M, Tanaka T (2000) *Bacillus subtilis comZ* (*yjzA*) negatively affects expression of *comG* but not *comK*. *J Bacteriol* 182(17):4992–4994.
15. Mann HB, Whitney DR (1947) On a test of whether one of two random variables is stochastically larger than the other. *Ann Math Stat* 50-60.
16. Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 28(1):27-30.
17. Spatz HC, Trautner TA (1970) One way to do experiments on gene conversion? Transfection with heteroduplex *SPP1* DNA. *Mol Gen Genet* 109(1):84–106.
18. Dubnau D (1991) Genetic competence in *Bacillus subtilis*. *Microbiol Rev* 55(3):395–424.
19. Michod RE, Wojciechowski MF, Hoelzer MA (1988) DNA repair and the evolution of transformation in the bacterium *Bacillus subtilis*. *Genetics* 118(1):31–39.
20. Wojciechowski M., Hoelzer MA, Michod RE (1989) DNA repair and the evolution of transformation in *Bacillus subtilis*. II. Role of inducible repair. *Genetics* 121(3):411-422.
21. Hoelzer MA, Michod RE (1991) DNA repair and the evolution of transformation in Bacillus subtilis. III. Sex with damaged DNA. *Genetics* 128(2):215-223.
22. Yasbin RE (1977) DNA repair in *Bacillus subtilis*. *Molecular and General Genetics* 153(2):219–225.
23. Love PE, Yasbin RE (1984) Genetic characterization of the inducible SOS-like system of *Bacillus subtilis*. *J Bacteriol* 160(3):910-920.
24. Love PE, Lyle MJ, Yasbin RE (1985) DNA-damage-inducible (*din*) loci are transcriptionally activated in competent *Bacillus subtilis*. *Proc Natl Acad Sci USA* 82(18):6201–6205.
25. Au N, et al. (2005) Genetic composition of the *Bacillus subtilis* SOS system. *J Bacteriol* 187(22):7655–7666.
26. Raymond-Denise A, Guillen N (1992) Expression of the *Bacillus subtilis dinR* and *recA* genes after DNA damage and during competence. *J Bacteriol*174(10):3171–3176.

27. Hamoen LW, Haijema B, Bijlsma JJ, Venema G, Lovett CM (2001) The *Bacillus subtilis* competence transcription factor, ComK, overrides LexA-imposed transcriptional inhibition without physically displacing LexA. *J Biol Chem* 276(46):42901–42907.
28. Hamoen LW, Venema G, Kuipers OP (2003) Controlling competence in *Bacillus subtilis*: shared use of regulators. *Microbiology* 149(1):9–17.
29. Mirouze N, Ferret C, Cornilleau C, Carballido-López R (2018) Antibiotic sensitivity reveals that wall teichoic acids mediate DNA binding during competence in *Bacillus subtilis*. *Nat Commun* 9(5072).
30. Dubnau D, Blokesch M (2019) Mechanisms of DNA uptake by naturally competent bacteria. *Ann Rev Genet* 53.
31. Haijema BJ, Hahn J, Haynes J, Dubnau D (2001) A ComGA-dependent checkpoint limits growth during the escape from competence. *Mol Microbiol* 40(1):52-64.
32. Chen I, Christie PJ, Dubnau D (2005) The ins and outs of DNA transfer in bacteria. *Science* 310(5753):1456–1460.
33. Tojo S, et al. (2005) Elaborate transcription regulation of the *Bacillus subtilis ilv-leu* operon involved in the biosynthesis of branched-chain amino acids through global regulators of CcpA, CodY and TnrA. *Mol Microbiol* 56(6):1560–1573.
34. Ferson AE, Wray Jr. LV, Fisher SH (1996) Expression of the *Bacillus subtilis gabP* gene is regulated independently in response to nitrogen and amino acid availability. *Mol Microbiol* 22(4):693–701.
35. Slack FJ, Serror P, Joyce E, Sonenshein AL (1995) A gene required for nutritional repression of the *Bacillus subtilis* dipeptide permease operon. *Mol Microbiol* 15(4):689–702.
36. Wray, Jr. LV, Ferson AE, Fisher SH (1997) Expression of the *Bacillus subtilis ureABC* operon is controlled by multiple regulatory factors including CodY, GlnR, TnrA, and Spo0H. *J Bacteriol* 179(17):5494–5501.
37. Serror P, Sonenshein AL (1996) Interaction of CodY, a novel *Bacillus subtilis* DNA-binding protein, with the *dpp* promoter region. *Mol Microbiol*20(4):843–852.

38. Mirel DB, et al. (2000) Environmental regulation of *Bacillus subtilis* σ^D-dependent gene expression. *J Bacteriol* 182(11):3055–3062.

39. Debarbouille M, Gardan R, Arnaud M, Rapoport G (1999) Role of *bkdR*, a transcriptional activator of the *sigL*-dependent isoleucine and valine degradation pathway in *Bacillus subtilis*. *J Bacteriol* 181(7):2059-2066.

40. Mitani T, Heinze JE, Freese E (1977) Induction of sporulation in *Bacillus subtilis* by decoyinine or hadacidin. *Biochem Bioph Res Co* 77(3):1118–1125.

41. Freese E, Heinze JE, Galliers EM (1979) Partial purine deprivation causes sporulation of *Bacillus subtilis* in the presence of excess ammonia, glucose and phosphate. *J Gen Microbiol* 115(1):193–205.

42. Lopez JM, Marks CL, Freese EB (1979) The decrease of guanine nucleotides initiates sporulation of *Bacillus subtilis*. *BBA-Gen Subjects* 587(2):238–252.

43. Ratnayake-Lecamwasam M, Serror P, Wong KW, Sonenshein AL (2001) *Bacillus subtilis* CodY represses early-stationary-phase genes by sensing GTP levels. *Genes Dev* 15(9):1093-1103.

44. Willecke K, Pardee AB (1971) Fatty acid-requiring mutant of *Bacillus subtilis* defective in branched chain α-keto acid dehydrogenase. *J Biol Chem* 246(17):5264–5272.

45. Kaneda T (1991) Iso- and anteiso-fatty acids in bacteria: biosynthesis, function, and taxonomic significance. *Microbiol Rev* 55(2):288–302.

46. Diomandé SE, Nguyen-The C, Guinebretière M-H, Broussolle V, Brillard J (2015) Role of fatty acids in *Bacillus* environmental adaptation. *Front Microbiol* 6:813.

47. Liu J, Zuber P (1998) A molecular switch controlling competence and motility: competence regulatory factors ComS, MecA, and ComK control σ^D-dependent gene expression in *Bacillus subtilis*. *J Bacteriol* 180(16):4243–4251.

48. Caramori T, Barillà D, Nessi C, Sacchi L, Galizzi A (1996) Role of FlgM in σ^D-dependent gene expression in *Bacillus subtilis*. *J Bacteriol*178(11):3113–3118.

49. Fredrick K, Helmann JD (1996) FlgM is a primary regulator of σ^D activity, and its absence restores motility to a *sinR* mutant. *J Bacteriol*178(23):7010–7013.

50. Formstone A, Carballido-López R, Noirot P, Errington J, Scheffers D-J (2008) Localization and interactions of teichoic acid synthetic enzymes in *Bacillus subtilis*. *J Bacteriol* 190(5):1812–1821.

51. Brown S, Maria, Jr. JPS, Walker S (2013) Wall teichoic acids of gram-positive bacteria. *Ann Rev Microbiol* 67:313–336.

52. Perego M, Glaser P, Minutello A, Strauch MA, Leopold K, Fischer W (1995) Incorporation of D-alanine into lipoteichoic acid and wall teichoic acid in *Bacillus subtilis* identification of genes and regulation. *J Biol Chem* 270(26):15598-15606.

53. Carballido-López R, et al. (2006) Actin homolog MreBH governs cell morphogenesis by localization of the cell wall hydrolase LytE. *Dev Cell* 11:399-409.

54. Jones L, Carballido-López R, Errington J (2001) Control of cell shape in bacteria. *Cell* 104:913-922.

55. Scheffers D, Jones L, Errington J (2003) Several distinct localization patterns for penicillin-binding proteins in *Bacillus subtilis*. *Mol Microbiol* 51:749-764.

56. Edwards DH, Errington J (1997) The *Bacillus subtilis* DivIVA protein targets to the division septum and controls the site specificity of cell division. *Mol Microbiol* 24(5):905-915.

57. Wu L, Errington J (2004) Coordination of cell division and chromosome segregation by a nucleoid occlusion protein in *Bacillus subtilis*. *Cell* 117:915-925.

58. Bramkamp M, et al. (2008) A novel component of the division-site selection system of *Bacillus subtilis* and a new mode of action for the division inhibitor MinCD. *Mol Microbiol* 70:1556-1569.

59. Baranova N, Danchin A, Neyfakh A (1999) Mta, a global MerR-type regulator of the *Bacillus subtilis* multidrug-efflux transporters. *Mol Microbiol* 31:1549-1559.

60. Ohki R, Murata M (1997) *bmr3*, a third multidrug transporter gene of *Bacillus subtilis*. *J Bacteriol* 179:1423-1427.

61. Steinborn G, Hajirezaei M, Hofemeister J (2005) *bac* genes for recombinant bacilysin and anticapsin production in *Bacillus* host strains. *Arch Microbiol* 183:71-79.

62. Mahlstedt S, Walsh C (2010) Investigation of anticapsin biosynthesis reveals a four-enzyme pathway to tetrahydrotyrosine in *Bacillus subtilis*. *Biochemistry-US* 49:912-923.
63. Butcher R, et al. (2007) The identification of bacillaene, the product of the PksX megacomplex in *Bacillus subtilis*. *Proc Natl Acad Sci USA* 104:1506-1509.
64. Müller S, et al. (2014) Bacillaene and sporulation protect *Bacillus subtilis* from predation by *Myxococcus xanthus*. *Appl Environ Microbiol* 80:5603-5610.
65. Gonzalez-Pastor J, Hobbs E, Losick R (2003) Cannibalism by sporulating bacteria. *Science* 301:510-513.
66. Lamsa A, Liu W, Dorrestein P, Pogliano K (2012) The *Bacillus subtilis* cannibalism toxin SDP collapses the proton motive force and induces autolysis. *Mol Microbiol* 84:486-500.
67. Vogel C, Marcotte EM (2012) Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet* 13(4):227–232.
68. de Sousa Abreu R, Penalva LO, Marcotte EM, Vogel C (2009) Global signatures of protein and mRNA expression levels. *Mol Biosyst* 5(12):1512–1526.
69. Vega-Cabrera LA, et al. (2017) Analysis of Spo0M function in *Bacillus subtilis*. *PloS One* 12(2):e0172737.
70. Popp PF, Dotzler M, Radeck J, Bartels J, Mascher T (2017) The *Bacillus* BioBrick Box 2.0: expanding the genetic toolbox for the standardized work with *Bacillus subtilis*. *Sci Rep* 7(1):15058.
71. Toymentseva AA, Schrecke K, Sharipova MR, Mascher T (2012) The LIKE system, a novel protein expression toolbox for *Bacillus subtilis* based on the *liaI* promoter. *Microb Cell Fact* 11(1):143.
72. Babin BM, et al. (2017) Selective proteomic analysis of antibiotic-tolerant cellular subpopulations in *Pseudomonas aeruginosa* biofilms. *mBio* 8(5):e01593-17.
73. Schneider CA, Rasband WS, Eliceiri KW (2012) NIH image to ImageJ: 25 years of image analysis. *Nat Methods* 9(7):671–675.

74. Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* 26(12):1367–1372.

75. Mi H, Muruganujan A, Ebert D, Huang X, Thomas PD (2018) PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res* 47(D1):D419-D426.

3.9 Supplementary Figures

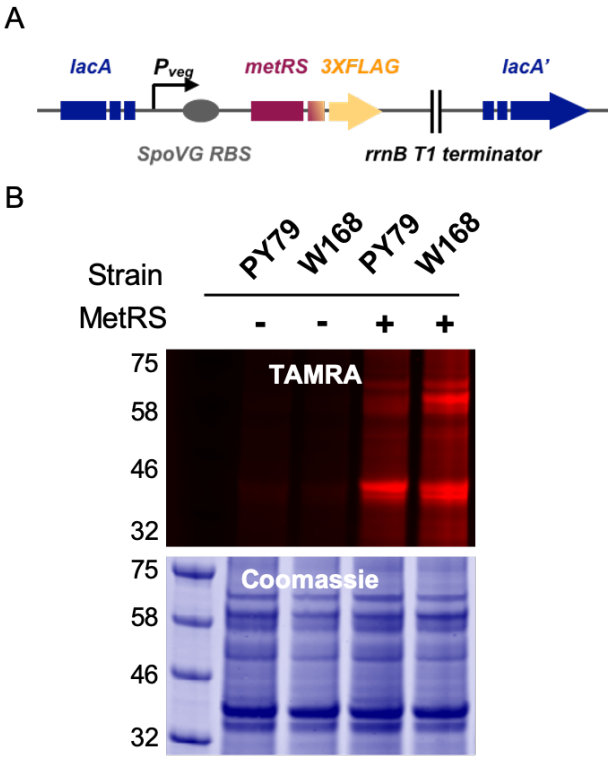

Figure 3.S1 The W168 strain incorporates Anl into protein synthesis using NLL-MetRS. (A) Genetic construct of P_{veg}*:nll-metRS* integrated into the *lacA* locus of W168 genome. (B) The genetic construct in (*A*) was integrated into two common lab strains, PY79 and W168. The two strains carrying the synthetase-expressing gene were grown in LB medium and pulse-labeled with Anl for 1 h during late-exponential phase. In-gel fluorescence imaging of TAMRA-alkyne-treated lysates showed both the strains could use NLL-MetRS to incorporate Anl into protein synthesis while W168 had higher labeling efficiency.

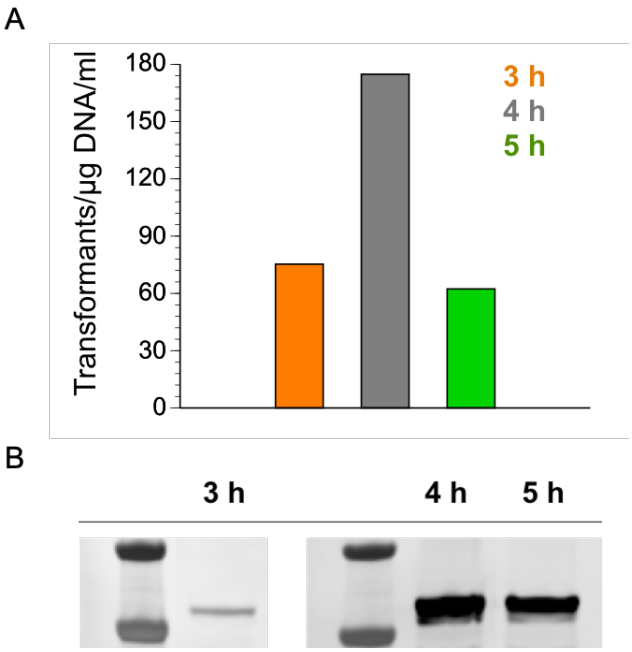

Figure 3.S2 Optimization of K-state inducing conditions. (A) Transformability of P_{comF}*:nll-metRS* grown in SM1 for 3, 4, and 5 h before diluted into SM2 and further cultured for 2 h. (B) Western blotting images of P_{comF}*:nll-metRS* cell lysates treated with anti-FLAG primary antibody and goat anti-mouse secondary antibody conjugated with Alexa Fluor 647. Culture conditions examined were the same as those in (A).

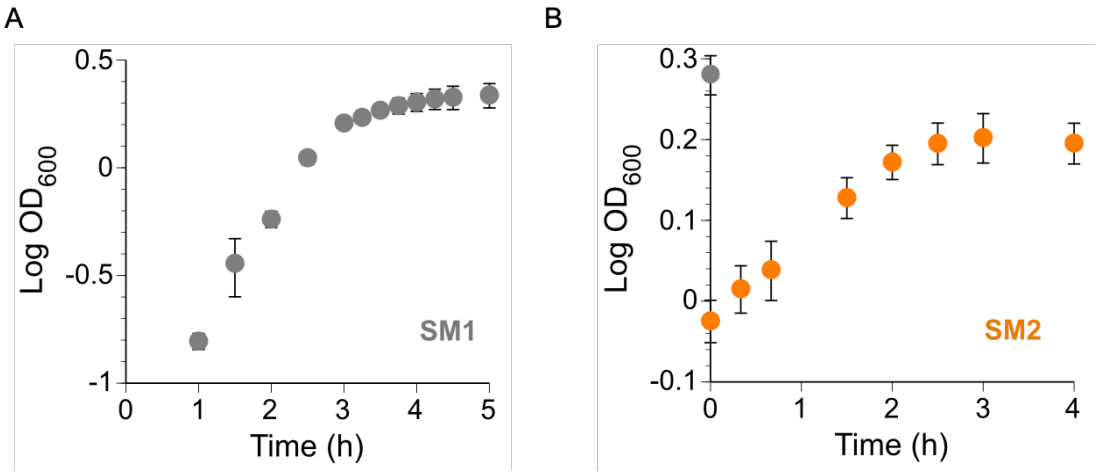

Figure 3.S3 Growth patterns of *P_{comF}:nll-metRS* strain. Growth curves of *P_{comF}:nll-metRS* (A) in SM1 and (B) after addition of equal volume of SM2 after culturing for 4 h in SM1. Overnight culture grown in LB was diluted 1:25 into 25 mL of fresh SM1 in a 250 mL Erlenmeyer flask. At each time point, 1 mL of culture was withdrawn for OD_{600} measurement. (N = 3.)

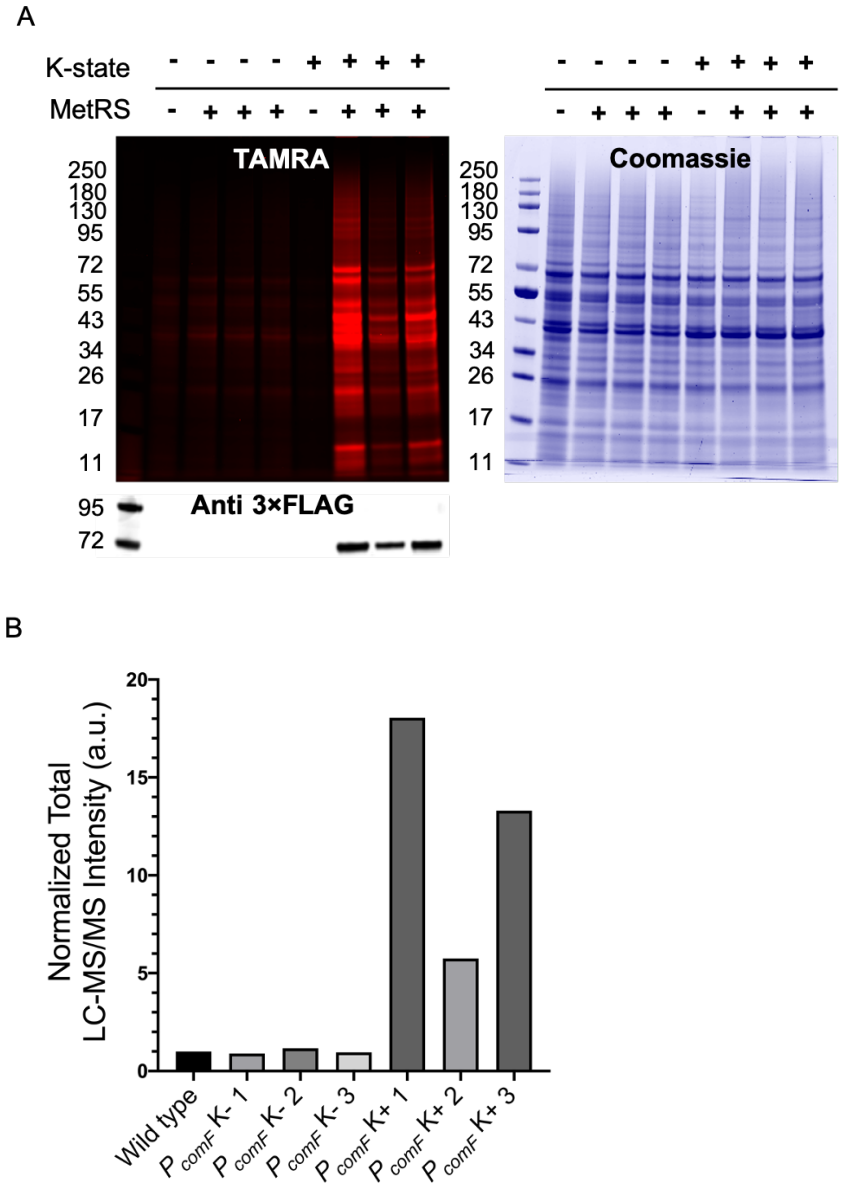

Figure 3.S4 Selective proteomic labeling of *B. subtilis* K-state subpopulation with the strain *P_{comF}: nll-metRS*. Cell cultures of W168 and *P_{comF}:nll-metRS* were incubated with 1 mM Anl for 1 h before and after t_0. (A) Lysates were treated with TAMRA-alkyne, separated by SDS-PAGE, and imaged by in-gel fluorescence. Coomassie staining of the same protein gel verified equal protein loading. Western blotting was applied to each lysate to probe the 3×FLAG tag on NLL-MetRS. (B) The same cell lysates from (A) were enriched for labeled proteins. LC-MS/MS was used to quantify the efficiency of enrichment and to further confirm low background of labeling.

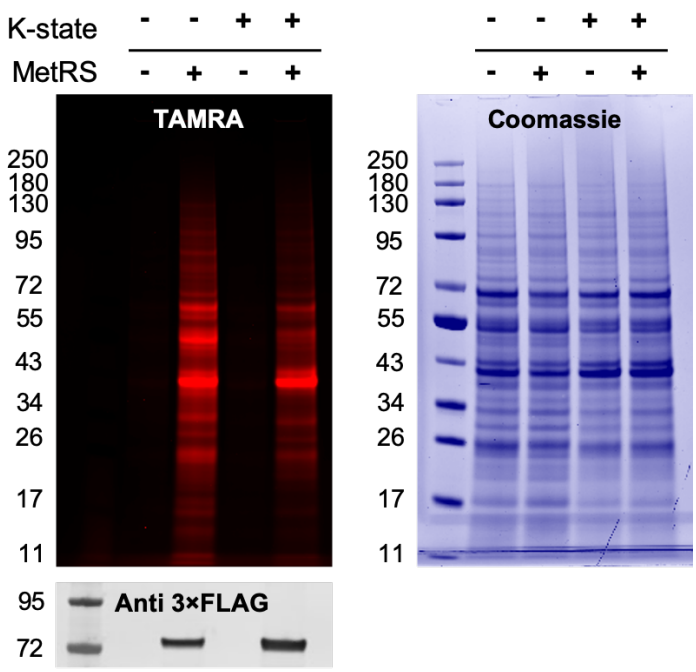

Figure 3.S5 Proteomic labeling of *B. subtilis* whole population with P_{veg}*: nll-metRS* before and after t_0. Samples labeled before and after entry into K-state both had strong TAMRA signal, confirming the validity of using P_{veg}*: nll-metRS* as a positive control.

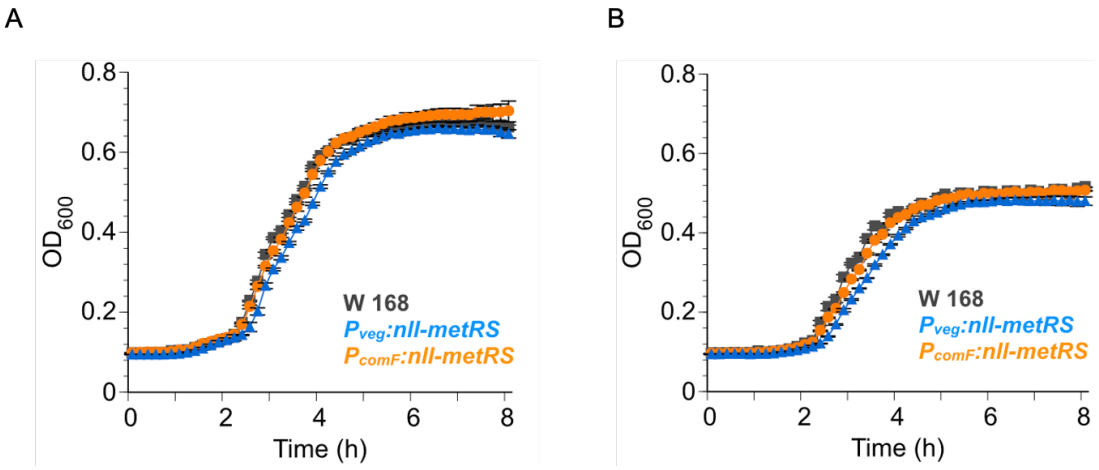

Figure 3.S6 Growth curves of W168, *P_{comF}:nll-metRS*, and *P_{veg}:nll-metRS*. (A) In SM1. (B) In SM2. Overnight culture of each strain grown in LB was diluted 1:100 into fresh medium. OD_{600} was monitored in 96-well plate over time. (N = 3.)

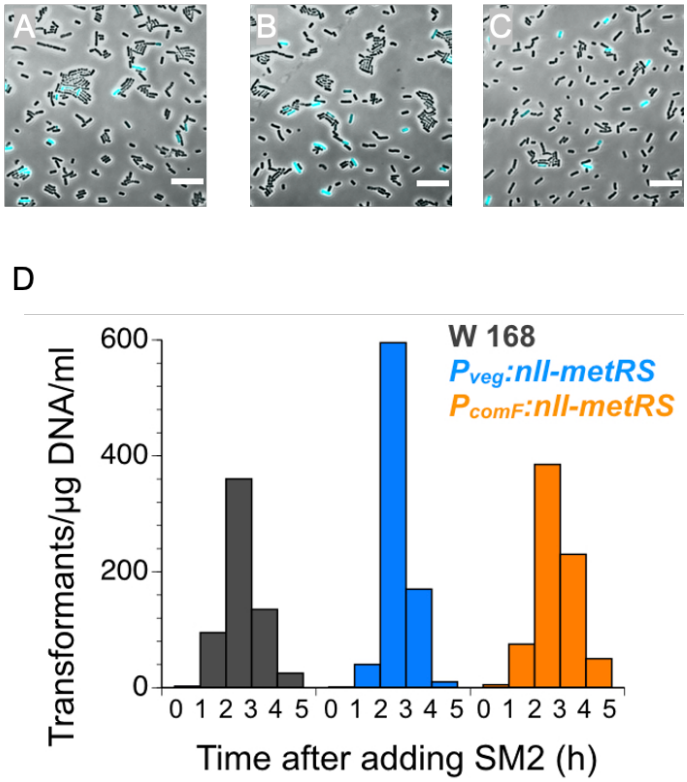

Figure 3.S7 Characterization of the effect of NLL-MetRS expression on ComK expression and competence induction. (A-C) Representative confocal laser scanning images of (A) *P_{comF}:nll-metRS_comK-mWasabi*, (B) *P_{veg}:nll-metRS_comK-mWasabi*, and (C) W168 *comK-mWasabi*. The strains were grown in SM1 for 4 h and further cultured in SM2 for 2 h. Phase contrast images were overlaid with mWasabi fluorescence images that were false-colored cyan. (Scale bar: 10 µm.) (D) Time-course transformation assay with W168, *P_{comF}:nll-metRS,* and *P_{veg}:nll-metRS* strains. Each strain was transformed with the same plasmid DNA and the number of transformants were determined at each of the indicated time point.

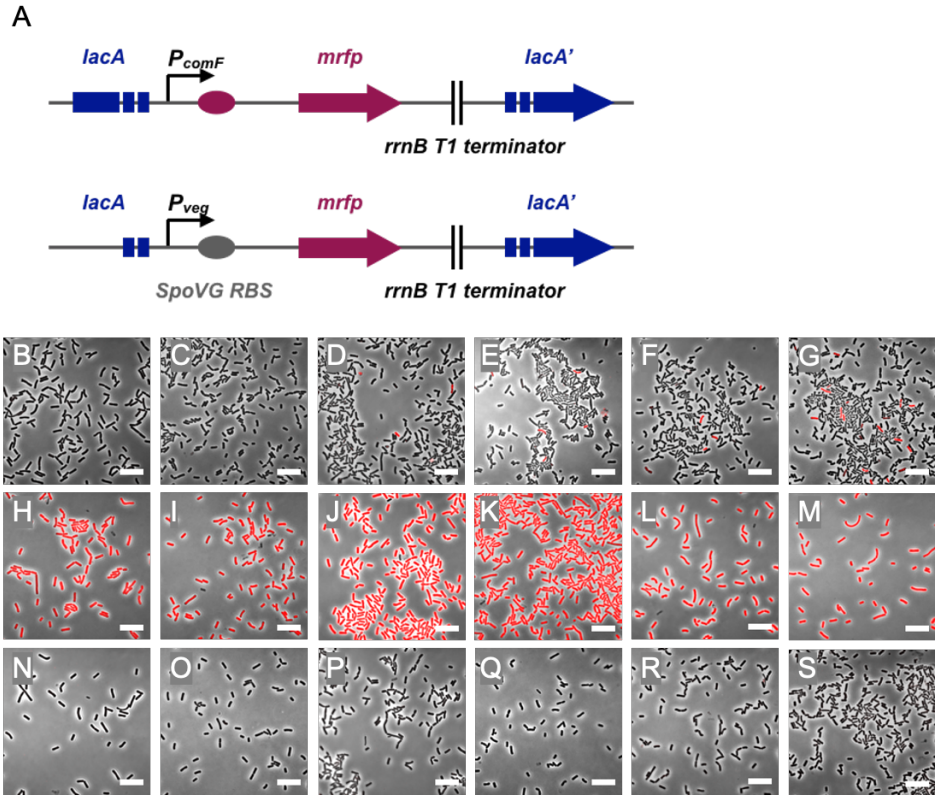

Figure 3.S8 Visualization of P_{comF} activity over time. (A) Genetic construct of P_{comF}:*mrfp*. (B-S) Representative confocal laser scanning images of (B-G) P_{comF}:*mrfp,* (H-M) P_{veg}:*mrfp*, and (N-S) W168. Phase contrast images overlaid with mRFP fluorescent images (false-colored red) were obtained for respective strain over the course of time: (B, H, N) 3 h in SM1, (C, I, O) 4 h in SM1, further cultured in SM2 for (D, J, P) 1 h, (E, K, Q) 2 h, (F, L, R) 3 h, and (G, M, S) 6 hr. (Scale bar: 10 μm.)

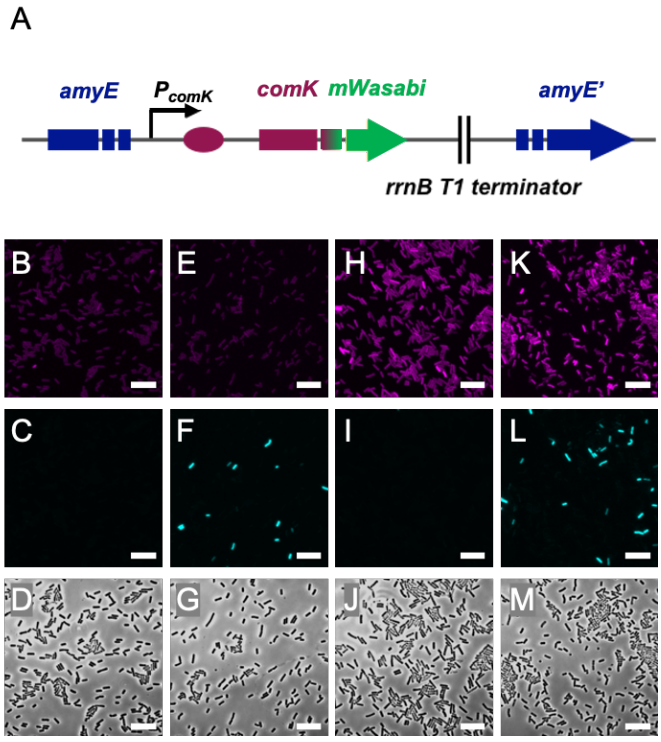

Figure 3.S9 Visualization of ComK-mWasabi expression and Anl-labeling in W168 *comK-mWasabi* and *P_{veg}:nll-metRS_comK-mWasabi* strains. (A) Genetic construct of *comK-mWasabi*. (B-M) Representative confocal laser scanning microscopy images of ComK-mWasabi and NLL-MetRS expressing strains. Images of (B, E, H, K) TAMRA fluorescence (magenta), (C, F, I, L) mWasabi fluorescence (cyan), and (D, G, J, M) phase contrast were obtained for the respective Anl-treated strains before and after entry into K-state: (B-D) W168 *comK-mWasabi* labeled with Anl from 2 to 3 h in SM1 and (E-G) in SM2; (H-J) *P_{veg}:nll-metRS_comK-mWasabi* labeled with Anl from 2 to 3 h in SM1 and (K-M) in SM2. (Scale bar: 10 μm.)

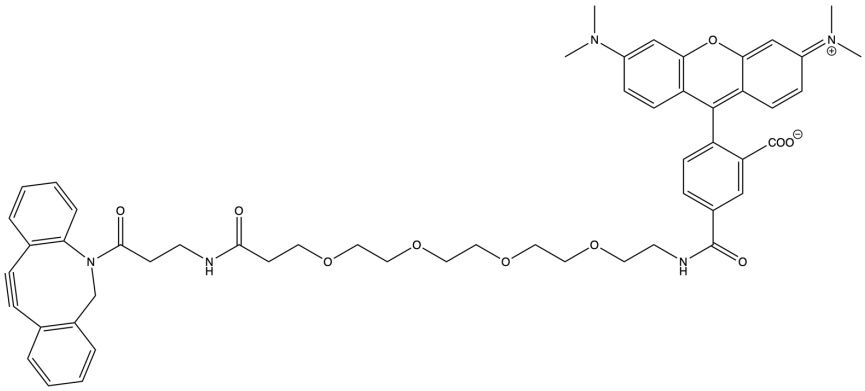

Figure 3.S10 Structure of TAMRA-DBCO.

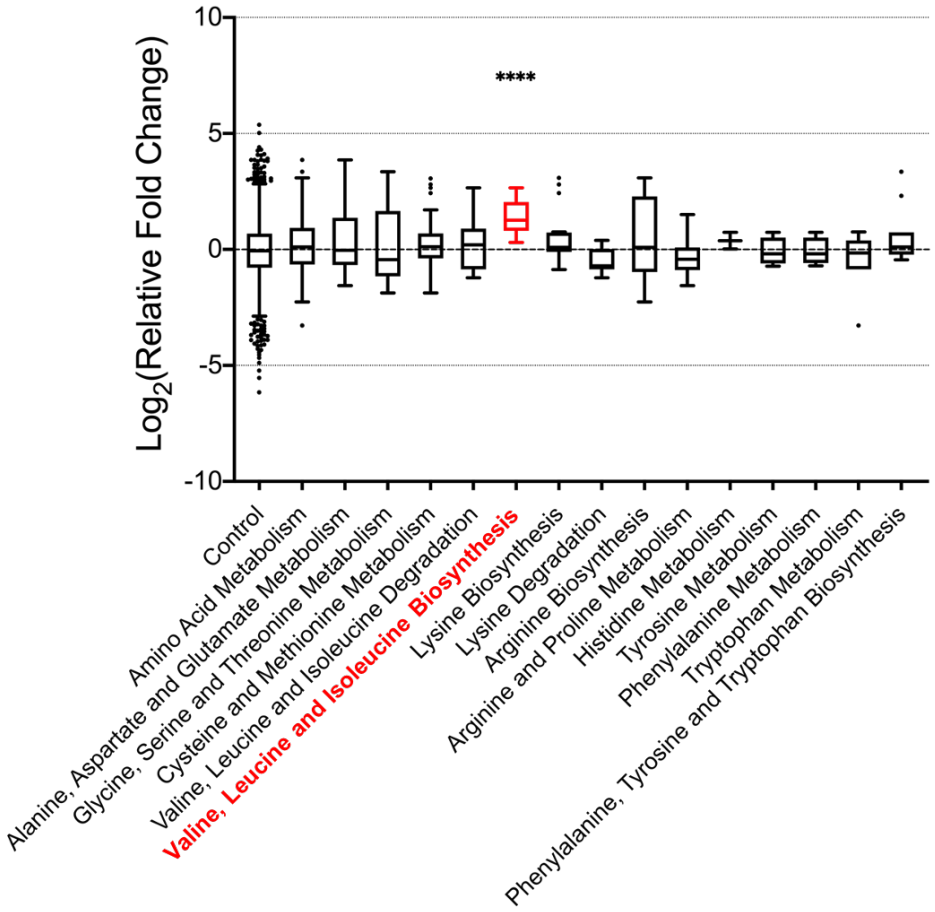

Figure 3.S11 Amino acid metabolism and degradation regulations in K-state subpopulation. Proteins identified in both the strains were categorized based on Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathway database. Tukey's box plot showed the relative fold change of proteins for each annotated pathway. Each box extended from the 25th to 75th percentiles of the distribution. Mann-Whitney *U* test was employed to determine if the expression of a certain pathway was significantly different from the control (****, $p < 0.0001$). The pathway significantly upregulated was annotated in red.

3.10 Supplementary Tables

Table 3.S1: Strains and primers used in this study.

***Bacillus subtilis* strains**		
Name	**Genotype**	**Source**
W168	*B. subtilis* wild type strain	Bacillus Genetic Stock Center
P_{veg}*:nll-metRS*	W168 *lacA*::P_{veg}*:nll-metRS; Erm*^{R}	This study
P_{comF}*:nll-metRS*	W168 *lacA*::P_{comF}*:nll-metRS; Erm*^{R}	This study
P_{veg}*:mrfp*	W168 *lacA*::P_{veg}*:mrfp; Erm*^{R}	This study
P_{comF}*:mrfp*	W168 *lacA*::P_{comF}*:mrfp; Erm*^{R}	This study
comK-mWasabi	W168 *amyE*::*comK-mWasabi; Cm*^{R}	This study
P_{comF}*:nll-metRS_comK-mWasabi*	W168 *amyE*::*comK-mWasabi; Cm*^{R} *lacA*::P_{comF}*:nll-metRS; Erm*^{R}	This study
P_{veg}*:nll-metRS_comK-mWasabi*	W168 *amyE*::*comK-mWasabi; Cm*^{R} *lacA*::P_{veg}*:nll-metRS; Erm*^{R}	This study

Primers	
Name	**Sequence**
YH110	GATCGAATTCGGAGTTCTGAGAATTGGT
YH119	GATCGGTCTCCCACCTTTCTCTAGTAACATTTATTGTACAACACGAGCC
YH152	GATCACTAGTTTATTAAGCACCGGTGGAGT
YH235	GATCAAGCTTATCGAATTCGGAACAATTGTGAACGGATAATAAA
YH236	GATCGGTCTCTGACTCATATTATGGCCTCCA
YH255	GATCGAATTCTGCGAGCTGCTGTTATCCAC
YH256	GATCGGTCTCTTGAGTTTCAACTGGCACATTCACATAGCAC
YH259	GATCGGTCTCACTCAAGTCGCGAAGAAAATTC
YH260	GATCCTCGAGTTTAGAGGCTTCCACCAGTG
YH278	GATCGGTCTCCTCCACCATACCGTTCCCCGAGCT
YH279	GATCGGTCTCGTGGAGGAGGATCCGTCAGTAAAGGGGAAGAAAC

Appendix A

SUPPLEMENTARY INFORMATION FOR CHAPTER 2

Supplementary Datasets

Data Set A.1 Full proteomic results from labeling of planktonic *P. aeruginosa* cultures. All proteins identified by LC-MS/MS from the BONCAT-enriched samples are listed. Columns A-E give the Uniprot ID of the protein, the protein description, the corresponding gene name if available, and the locus ID in both the PA14 and PAO1 strains. Columns F-M give LC-MS/MS ion intensities for each protein in the wild-type and P_{pvdF}*:nll-metRS* strains in either iron-depleted or iron-replete conditions. Columns N-U give LC-MS/MS normalized LFQ values for each protein in the wild-type and P_{pvdF}*:nll-metRS* strains in either iron-depleted or iron-replete conditions.

Data Set A.2 Full proteomic results from *P. aeruginosa* biofilm labeling experiments. All proteins identified by LC-MS/MS from the BONCAT-enriched samples are listed. Columns A-E give the Uniprot ID of the protein, the protein description, the corresponding gene name if available, and the locus ID in both the PA14 and PAO1 strains. Columns F-K give LC-MS/MS iBAQ values for each protein in the P_{pvdF}*:nll-metRS* (columns F-H) and P_{trc}*:nll-metRS* (columns I-K) strains. Columns L-Q give LC-MS/MS normalized LFQ values for each protein in the P_{pvdF}*:nll-metRS* (columns L-N) and P_{trc}*:nll-metRS* (columns O-Q) strains. Columns R-T give the log2-transformed LFQ ratios (fold change, P_{pvdF} / P_{trc}) for proteins shared between the

two strains, the fold change, and the corresponding FDR adjusted *p*-values.

Data Set A.3 Full proteomic results from two separate sets of biofilm labeling experiments. All proteins identified by LC-MS/MS from the BONCAT-enriched samples are listed. Columns A-E give the Uniprot ID of the protein, the protein description, the corresponding gene name if available, and the locus ID in both the PA14 and PAO1 strains. Columns F-K give LC-MS/MS iBAQ values for each protein identified in P_{pvdF}*:nll-metRS* samples prepared from two sets of experiments. Columns L-Q give LC-MS/MS iBAQ values for each protein identified in P_{trc}*:nll-metRS* samples prepared from two sets of experiments. Columns R-W give LC-MS/MS normalized LFQ values for each protein identified in P_{pvdF}*:nll-metRS* samples prepared from two sets of experiments. Columns X-AC give LC-MS/MS normalized LFQ values for each protein identified in P_{trc}*:nll-metRS* samples prepared from two sets of experiments.

Data Set A.4 Full results from the KEGG Pathway analysis. Each sheet contains LC-MS/MS data from a certain annotated KEGG pathway. Line 1 of each sheet gives the annotation of the KEGG pathway and the Mann-Whitney *U* test result compared with the control group where all shared proteins were included. Columns A-E give the Uniprot ID of the protein, the protein description, the corresponding gene name if available, and the locus ID in both the PA14 and PAO1 strains. Columns F-H give the log2-transformed LFQ ratios (fold change, P_{pvdF} / P_{trc}) for proteins shared between the two strains, the fold change, and the corresponding FDR adjusted *p*-values. Columns I and J show whether a certain protein was

only identified in at least two replicates of P_{pvdF}*:nll-metRS* and P_{trc}*:nll-metRS* samples, respectively.

Data Set A.5 Full proteomic results from 12 h labeling of P_{pvdF}*:nll-metRS* and P_{trc}*:nll-metRS* biofilms experiments. All proteins identified by LC-MS/MS from the BONCAT-enriched samples are listed. Sheet 1 gives the LC-MS/MS results of P_{pvdF}*:nll-metRS* samples. Columns A-E give the Uniprot ID of the protein, the protein description, the corresponding gene name if available, and the locus ID in both the PA14 and PAO1 strains. Columns F-H give LC-MS/MS iBAQ values for each protein identified in P_{pvdF}*:nll-metRS* samples. Columns I-K give LC-MS/MS iBAQ values for each protein identified in P_{pvdF}*:nll-metRS* samples. Sheet 2 gives the LC-MS/MS results of both the P_{pvdF}*:nll-metRS* and the P_{trc}*:nll-metRS* samples. Columns A-E give the Uniprot ID of the protein, the protein description, the corresponding gene name if available, and the locus ID in both the PA14 and PAO1 strains. Columns F-K give LC-MS/MS iBAQ values for each protein in the P_{pvdF}*:nll-metRS* (columns F-H) and P_{trc}*:nll-metRS* (columns I-K) strains. Columns L-Q give LC-MS/MS normalized LFQ values for each protein in the P_{pvdF}*:nll-metRS* (columns L-N) and P_{trc}*:nll-metRS* (columns O-Q) strains. Columns R-T give the log2-transformed LFQ ratios (fold change, P_{pvdF} / P_{trc}) for proteins shared between the two strains, the fold change, and the corresponding FDR adjusted *p*-values.

Data Set A.6 Proteins identified both in the Babin study and in this work, along with the full proteomic results obtained here. Combination of Data Set A.2 with the proteomic data from the Babin study. Columns A-T are the same as those in Data Set A.2. Columns U-W give

the log2-transformed LFQ ratios (fold change, P_{rpoS} / P_{trc}) for proteins shared between the two strains, the fold change, and the corresponding FDR adjusted *p*-values.

Data Set A.7 Proteins identified both in the oxygen limitation study and in this work, along with the full proteomic results obtained here. Combination of Data Set A.2 with the proteomic data from previous oxygen stress study. Columns A-T are the same as those in Data Set A.2. Columns U-W give the log2-transformed LFQ ratios (fold change, anaerobic/aerobic) for proteins shared between the two conditions, the fold change, and the corresponding FDR adjusted *p*-values.

Appendix B

SUPPLEMENTARY INFORMATION FOR CHAPTER 3

Supplementary Datasets

Data Set B.1: Full proteomic results from P_{comF} promoter activity validation experiments. All proteins identified by LC-MS/MS from the BONCAT-enriched samples are listed. Columns A-D give the Uniprot ID of the protein, the protein description, the corresponding gene name if available, and the locus ID in W168 strain. Column E gives LC-MS/MS ion intensities for each protein in the wild-type strain without K-state induction. Columns F-K give LC-MS/MS ion intensities for each protein in the P_{comF}*:nll-metRS* strain without (columns F-H) and with (columns I-K) K-state induction. Column L gives LC-MS/MS normalized LFQ values for each protein in the wild-type strain without K-state induction. Columns M-R give LC-MS/MS normalized LFQ values for each protein in the P_{comF}*:nll-metRS* strain without (columns M-O) and with (columns P-R) K-state induction.

Data Set B.2 Full proteomic results from *B. subtilis* K-state labeling experiments. All proteins identified by LC-MS/MS from the BONCAT-enriched samples are listed. Columns A-D give the Uniprot ID of the protein, the protein description, the corresponding gene name if available, and the locus ID in W168 strain. Columns E-J give LC-MS/MS iBAQ values for each protein in the P_{comF}*:nll-metRS* (columns E-G) and P_{veg}*:nll-metRS* (columns H-J) strains. Columns K-P give LC-MS/MS normalized LFQ values for each protein in the P_{comF}*:nll-metRS*

(columns K-M) and P_{veg}*:nll-metRS* (columns N-P) strains. Columns Q-S give the log2-transformed LFQ ratios (fold change, P_{comF} / P_{veg}) for proteins shared between the two strains, the fold change, and the corresponding FDR adjusted *p*-values.

Data Set B.3 Full results from the Gene Ontology (GO) analysis. Sheet 1 contains LC-MS/MS data from establishment of competence for transformation (GO: 0030420) pathway; sheet 2 contains LC-MS/MS data from SOS response (GO: 0009432) pathway. In each sheet, line 1 gives the annotation of the KEGG pathway and the Mann-Whitney *U* test result compared with the control group where all shared proteins were included. Columns A-D give the Uniprot ID of the protein, the protein description, the corresponding gene name if available, and the locus ID in W168 strain. Columns E-G give the log2-transformed LFQ ratios (fold change, P_{comF} / P_{veg}) for proteins shared between the two strains, the fold change, and the corresponding FDR adjusted *p*-values. Columns H and I show whether a certain protein was only identified in at least two replicates of P_{comF}*:nll-metRS* and P_{veg}*:nll-metRS* samples, respectively.

Data Set B.4 Full results from the KEGG Pathway analysis. Each sheet contains LC-MS/MS data from a certain annotated KEGG pathway. Line 1 of each sheet gives the annotation of the KEGG pathway and the Mann-Whitney *U* test result compared with the control group where all shared proteins were included. Columns A-D give the Uniprot ID of the protein, the protein description, the corresponding gene name if available, and the locus ID in W168 strain. Columns E-G give the log2-transformed LFQ ratios (fold change, P_{comF} / P_{veg})

for proteins shared between the two strains, the fold change, and the corresponding FDR adjusted *p*-values. Columns H and I show whether a certain protein was only identified in at least two replicates of P_{comF}*:nll-metRS* and P_{veg}*:nll-metRS* samples, respectively.