Global Analysis of Protein Synthesis and Degradation in *Escherichia coli*

Thesis by
Elliot James MacKrell

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

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2025
Defended June 20th, 2024

Elliot James MacKrell
ORCID: 0009-0006-5619-8548

ACKNOWLEDGEMENTS

I would first like to express gratitude to my advisor, Prof. David Tirrell, for extending exceptional trust, freedom, and encouragement as I integrated my intellectual curiosity into the laboratory's research program. My interests and skills have shifted substantially from my first year to the sunset of my graduate education, and this evolution was made possible only through Dave's patience and kindness. The independence I was permitted in my research endeavours has gifted me with a deep pride in the discoveries I have made. I am forever grateful.

I would next like to thank the other members of my thesis committee, namely Prof. Rustem Ismagilov, Prof. Mikhail Shapiro, and Prof. Dianne Newman, for their constructive comments on my research, particularly in my final year of study. Their perspectives have sharpened my own and have produced a welcomed, visible impact on my research.

I would like to thank current and former members of the Proteome Exploration Laboratory at the Beckman Institute for their guidance and expert knowledge; specifically, Prof. Tsui-Fen Chou, Dr. Brett Lomenick, Dr. Jeff Jones, and Dr. Ting-Yu Wang. My projects were indelibly impacted by their keen intellect and direction in mass spectrometry–based proteomics, and they are all genuinely delightful on a personal and professional level.

I would also like to thank colleagues of the Tirrell lab with whom I've had the fortune of overlapping. I will remember our spirited academic debate and many of our interactions fondly. My time mentoring Cindy Cao through her undergraduate thesis in chemical engineering was especially rewarding. It was truly a joy to experience her unwavering positivity—during a global pandemic, no less.

I will also provide a special thanks to Irina Meininger and Allison Kinard. Their proficiency in addressing the needs of graduate students in the department is outshined only by their kindheartedness. Their advice and support throughout my journey into parenthood have been invaluable.

Lastly, I would like to express my deepest gratitude to my family: my parents, Mary and Jim; my brother, Evan; and my wife and beloved son, Victoria and Simon. You are everything.

ABSTRACT

Protein synthesis and degradation shape the cellular proteome to drive homeostasis and physiological adaptation. Many fundamental aspects of protein regulation have been elucidated through investigation of the Gram-negative bacterium *Escherichia coli*, which remains a fruitful model organism for uncovering conserved regulatory mechanisms relevant to cell biology, biotechnology, and medicine. Here, we used bioorthogonal noncanonical amino acid tagging (BONCAT) for the time-resolved analysis of protein synthesis and degradation in this organism in several contexts. We profiled protein degradation on a proteome-wide scale in growing and growth-arrested cells, identifying instability in a diverse panel of regulators. Our identifications served as training data in the validation and deployment of a machine learning classifier of in vivo protein stability, which highlighted the role of active degradation in motility and surface adhesion. We then utilized an efficient system of active degradation in this organism to engineer the instability of the mutant methionyl-tRNA synthetase NLL-MetRS for the analysis of protein synthesis in transient physiological states. Destabilized NLL-MetRS variants exhibited half-lives on the order of hours, which improved the fidelity of metabolic labeling in growth-arrested cells. Additionally, we leveraged the sensitivity of BONCAT to investigate protein synthesis in growth-arrested cells expressing a well-studied but controversial member of the widespread toxin-antitoxin family, MazF. Our proteomic profiling suggests this toxin activates several endogenous stress response systems, most notably the cold shock response system. Taken together, these investigations highlight the advantage of time-resolved proteomics in characterizing proteome dynamics.

PUBLISHED CONTENT AND CONTRIBUTIONS

(1) Jones, J.; MacKrell, E. J.; Wang, T. Y.; Lomenick, B.; Roukes, M. L.; Chou, T. F. Tidyproteomics: an open-source R package and data object for quantitative proteomics post analysis and visualization. *BMC Bioinformatics* **2023**, *24*, 239, DOI: 10.1186/s12859-023-05360-7,
E.J.M. participated in software development, documentation distribution, source code distribution, application deployment, and preparation of the manuscript.

(2) Altshuller, M.; He, X.; MacKrell, E. J.; Wernke, K. M.; Wong, J. W. H.; Sellés-Baiget, S.; Wang, T. Y.; Chou, T. F.; Duxin, J. P.; Balskus, E. P.; Herzon, S. B.; Semlow, D. R. The Fanconi anemia pathway repairs colibactin-induced DNA interstrand cross-links. *bioRxiv* **2024**, DOI: 10.1101/2024.01.30.576698,
E.J.M. participated in software development, data analysis, and preparation of the manuscript.

(3) MacKrell, E. J.; Lomenick, B.; Chou, T. F.; Tirrell, D. A. Global protease substrate identification reveals instability of diverse regulators in Escherichia coli. *Manuscript to be submitted*,
E.J.M participated in the conception of the project, collection and analysis of data, and preparation of the manuscript.

(4) MacKrell, E. J.; Lomenick, B.; Jones, J.; Chou, T. F.; Tirrell, D. A. MazF activates the cold shock response in Escherichia coli. *Manuscript to be submitted*,
E.J.M participated in the conception of the project, collection and analysis of data, and preparation of the manuscript.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

NOMENCLATURE

AAA+. ATPases associated with diverse cellular activities.

Aha. Azidohomoalanine.

Anl. Azidonorleucine.

ANN. Artificial neural network.

ATC. A-type carrier.

aTc. Anhydrotetracycline.

AUC. Area under the curve.

BONCAT. Bioorthogonal noncanonical amino acid tagging.

c-di-GMP. cyclic di-GMP.

CSP. Cold shock protein.

DBCO. Dibenzocyclooctyne.

DGC. Diguanylate cyclase.

FDR. False discovery rate.

Fe-S. Iron-sulfur.

KEGG. Kyoto Encyclopedia of Genes and Genomes.

LC-MS/MS. Liquid chromatography-tandem mass spectrometry.

lmRNA. Leaderless mRNA.

ncAA. Noncanonical amino acid.

NTE. N-terminal extension.

PAS. Per-ARNT-Sim.

PDE. Phosphodiesterase.

PPI. Protein-protein interaction.

PR. Precision-recall.

PSK. Post-segregational killing.

ROC. Receiver operating characteristic.

ssRNA. Single-stranded RNA.

TA. Toxin-antitoxin.

TAMRA. Carboxytetramethylrhodamine.

THPTA. Tris(3-hydroxypropyltriazolylmethyl)amine.

TMT. Tandem mass tag.

Chapter 1

INTRODUCTION

My doctoral research examines proteome dynamics in *Escherichia coli* through a versatile and information-rich method in chemical biology: bioorthogonal noncanonical amino acid tagging (BONCAT).[1] This technique operates through the incorporation of azide- or alkyne-bearing noncanonical amino acids (ncAAs) into proteins, enabling researchers to leverage the bioorthogonality of azide-alkyne "click" reactions for applications such as the covalent conjugation of enrichment resins or fluorophores for proteomic identification or visualization, respectively, of these modified, or labeled, proteins.[1–3] In analogy to pulse-chase radiolabeling experiments with [^{35}S]methionine, incorporation is initiated by the investigator's addition of ncAA to the cell; therefore, signal derived from these labeled proteins provides information in the form of temporal resolution. This facet of the method is advantageous in dissecting biological systems rich in proteome complexity or poor in translational activity because the physical enrichment of labeled proteins from the pre-existing protein pool provides a direct view of protein synthesis. By leveraging the mutant methionyl-tRNA synthetase NLL-MetRS, which enables incorporation of the otherwise translationally inert methionine surrogate azidonorleucine, an extension of this method provides information along biological axes via cell or state selectivity.[4–6] In these approaches, expression of NLL-MetRS is restricted to a cell type or physiological state of interest, thereby enabling the proteomic profiling and visualization of protein synthesis in defined cell subpopulations.

My colleagues have demonstrated the capabilities of the BONCAT method across increasingly complex biological systems, each distinguished by marked temporal or spatial heterogeneity in protein synthesis.[7–9] In a complementary approach, investigators have illuminated the advantages of BONCAT through their examination of an interconnected but opposing process: protein destruction. For example, combined BONCAT and stable isotope labeling revealed widespread non-exponential degradation of nascent proteins in mammalian cells, which was attributed to the differential stability of complexed and free subunits of protein complexes.[10] The resolution of BONCAT was also utilized to label and measure the turnover of proteins synthesized in early S phase, revealing differential stabilities across protein functionalities and subcellular localizations.[11] Drawing inspiration from these works, I dedicated the

main body of my doctoral work to the integration of quantitative time-series global instability profiling with microbiology.

In Chapter II, we present an experimental and computational analysis of protein degradation aimed at globally identifying protease substrates in growing and growth-arrested *E. coli* cells. We identified instability across regulators mediating motility, biofilm development, iron homeostasis, and protein degradation itself. We then utilized the sensitivity of BONCAT to achieve the first proteome-wide analysis of nascent protein degradation in growth-arrested bacterial cells. A synthesis of these BONCAT analyses provided the basis for validating and deploying a machine learning classifier of native *E. coli* protein stability. We used this model to predict the stability of proteins that escaped identification in our proteomic profiling, revealing the active degradation of two key regulators of biofilm development.

In Chapter III, we present a strategy for the proteomic analysis of transient physiological states by engineering the destabilization of NLL-MetRS in *E. coli*. Such an approach may improve the resolution of state-selective analysis when the lifetime of NLL-MetRS far exceeds that of the physiological state of interest. This method relies on a degradation pathway widely conserved in bacteria—the ssrA pathway—for active degradation, which could in principle afford extension of the strategy to other species of bacteria. We demonstrate that the cellular abundance of NLL-MetRS can be tuned by active protein degradation to provide greater fidelity in state-selective protein labeling with respect to promoter activity.

In Chapter IV, we turn our efforts towards protein synthesis in a time-resolved analysis of cells expressing the well-studied but controversial endoribonuclease toxin MazF, which elicits growth arrest through a dramatic reduction in protein synthesis via widespread cleavage of cellular mRNAs. Our results suggest this toxin activates several stress response pathways consistent with translational inhibition and, more broadly, demonstrate that BONCAT provides sufficient sensitivity to characterize protein synthesis upon activation of translation-targeting toxin-antitoxin systems, which are widespread in bacteria.

References

(1) Dieterich, D. C.; Link, A. J.; Graumann, J.; Tirrell, D. A.; Schuman, E. M. Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT). *Proc Natl Acad Sci U S A* **2006**, *103*, 9482–9487.

(2) Beatty, K. E.; Xie, F.; Wang, Q.; Tirrell, D. A. Selective dye-labeling of newly synthesized proteins in bacterial cells. *J Am Chem Soc* **2005**, *127*, 14150–14151.

(3) Beatty, K. E.; Liu, J. C.; Xie, F.; Dieterich, D. C.; Schuman, E. M.; Wang, Q.; Tirrell, D. A. Fluorescence visualization of newly synthesized proteins in mammalian cells. *Angew Chem Int Ed Engl* **2006**, *45*, 7364–7367.

(4) Tanrikulu, I. C.; Schmitt, E.; Mechulam, Y.; Goddard, W. A.; Tirrell, D. A. Discovery of Escherichia coli methionyl-tRNA synthetase mutants for efficient labeling of proteins with azidonorleucine in vivo. *Proc Natl Acad Sci U S A* **2009**, *106*, 15285–15290.

(5) Ngo, J. T.; Champion, J. A.; Mahdavi, A.; Tanrikulu, I. C.; Beatty, K. E.; Connor, R. E.; Yoo, T. H.; Dieterich, D. C.; Schuman, E. M.; Tirrell, D. A. Cell-selective metabolic labeling of proteins. *Nat Chem Biol* **2009**, *5*, 715–717.

(6) Ngo, J. T.; Babin, B. M.; Champion, J. A.; Schuman, E. M.; Tirrell, D. A. State-selective metabolic labeling of cellular proteins. *ACS Chem Biol* **2012**, *7*, 1326–1330.

(7) Bagert, J. D.; van Kessel, J. C.; Sweredoski, M. J.; Feng, L.; Hess, S.; Bassler, B. L.; Tirrell, D. A. *Chem Sci* **2016**, *7*, 1797–1806.

(8) Babin, B. M.; Bergkessel, M.; Sweredoski, M. J.; Moradian, A.; Hess, S.; Newman, D. K.; Tirrell, D. A. SutA is a bacterial transcription factor expressed during slow growth in Pseudomonas aeruginosa. *Proc Natl Acad Sci U S A* **2016**, *113*, 597–605.

(9) Babin, B. M.; Atangcho, L.; van Eldijk, M. B.; Sweredoski, M. J.; Moradian, A.; Hess, S.; Tolker-Nielsen, T.; Newman, D. K.; Tirrell, D. A. Selective Proteomic Analysis of Antibiotic-Tolerant Cellular Subpopulations in Pseudomonas aeruginosa Biofilms. *mBio* **2017**, *8*, 10.1128/mbio.01593–17.

(10) McShane, E.; Sin, C.; Zauber, H.; Wells, J. N.; Donnelly, N.; Wang, X.; Hou, J.; Chen, W.; Storchova, Z.; Marsh, J. A.; Valleriani, A.; Selbach, M. Kinetic Analysis of Protein Stability Reveals Age-Dependent Degradation. *Cell* **2016**, *167*, 803–815.

(11) Chen, W.; Smeekens, J. M.; Wu, R. Systematic study of the dynamics and half-lives of newly synthesized proteins in human cells. *Chem Sci* **2016**, *7*, 1393–1400.

Chapter 2

GLOBAL IDENTIFICATION OF PROTEASE SUBSTRATES IN *ESCHERICHIA COLI*

2.1 Abstract

Regulated protein degradation underlies the timely execution of essential gene expression programs in bacteria. Here, we deployed time-resolved chemoproteomics, text mining of the PubMed and EcoCyc knowledge bases, and machine learning classification to discover proteolytic regulation in exponential and stationary phase *Escherichia coli* physiologies. We verified the instability of diverse homeostatic and stress response regulators, including the principal cyclic-di-GMP phosphodiesterase PdeH, the N-end rule substrate chaperone ClpS, and all four A-type domain iron-sulfur cluster carriers, IscA, ErpA, NfuA, and SufA. Mutagenesis of the PdeH N-terminal extension abolished ClpXP recognition, which otherwise assisted in stationary phase depletion of PdeH. Protease substrates with pronounced or exclusive degradation in stationary phase such as the morphology regulator BolA, RNA polymerase omega subunit RpoZ, and the biofilm regulator BssR have defined roles in quiescence. Finally, machine learning–assisted substrate identification revealed Lon-mediated degradation of two opposing key regulators of surface adhesion, the RpoS antagonist FliZ and the major curli fiber transcriptional activator CsgD, suggesting proteolysis may hasten transitions between motility and sessility. Together, these results highlight the role of regulated proteolysis in driving physiological adaptation for this model organism.

2.2 Introduction

Protein degradation is essential for ameliorating proteotoxicity and achieving full spatiotemporal precision in post-translational regulation. In bacteria, ATP-dependent protein degradation proceeds through a sequential process of substrate recognition, unfolding, translocation, and subsequent peptide bond cleavage by an AAA+ (ATPases associated with diverse cellular activities) protease.[1] Substrate recognition requires the presence of one or more amino acid sequences or structural motifs, or degrons, in the substrate and may further demand the activity of adaptor proteins, which chaperone substrates to their cognate protease(s) under permissive conditions, or endoproteolytic cleavage, which may expose cryptic internal degrons.[2–4] *Escherichia coli* bears one essential membrane-bound AAA+ protease, FtsH, and four nonessential cytosolic AAA+ proteases: Lon, ClpXP, ClpAP, and HslUV. Seminal radiolabeling studies demonstrated a small proportion (ca. 5%) of nascent protein is rapidly degraded under exponential growth, elucidating a rarity of proteolysis in this organism reflected in its high expression of chaperones dedicated to the renaturation, and not the elimination, of potentially harmful misfolded proteins.[5–7] Yet protein degradation is a demonstrated necessity in a broad range of homeostatic, adaptational, and developmental programs, thus establishing a need for identifying protease substrates in faithfully mapping regulatory networks.[8–11]

Several strategies exist for uncovering proteolytic regulation. Overexpression of catalytically inactivated AAA+ proteases for affinity purification and identification of trapped substrates has proven successful in several species of bacteria.[12,13] Indeed, a seminal ClpXP trapping analysis under SOS response induction revealed DNA damage–associated substrate enrichment, suggesting the active pool of protease substrates can vary with cell phenotype.[14] However, artificially elevated protease abundances may invoke toxicity or perturb endogenous protein abundances or substrate-protease interactions, and such experiments inherently limit the scope of substrate identification and kinetic analysis.[15] One study alternatively accomplished a global stability analysis in *E. coli* using stable isotope labeling, though it focused on protein attributes governing proteolysis with less regard to the regulatory implications of select candidate substrates' instabilities.[16] Additionally, while the ubiquity and clinical relevance of growth arrest has garnered significant attention, study of the nascent proteome in quiescent bacterial cells is complicated by low metabolic activity, which biases protein identification towards the pre-existing proteome and thereby limits conventional isotopic labeling strategies.[17] More broadly, though de-

velopments in acquisition and fractionation strategies have considerably advanced quantitative accuracy and proteome coverage, bottom-up mass spectrometry-based proteomics remains biased towards abundant species with permissive physicochemical attributes, which precludes full proteome coverage.[18]

To address these challenges, we validated and expanded the application of bioorthogonal noncanonical amino acid tagging (BONCAT), a time-resolved chemical labeling method, in the analysis of bacterial protein stability. The BONCAT method operates through the treatment of cells with an azide- or alkyne-bearing amino acid surrogate for azide-alkyne cycloaddition–based "click" conjugation to enrichment resins, thereby enabling time-resolved chemoproteomic identification of labeled proteins.[19] Owing to its sensitivity in labeling nascent proteomes, BONCAT has demonstrated potential in dissecting microbial dormancy through the analysis of protein synthesis in quiescent cells in vitro, in vivo, and in situ.[20–22] Further, a BONCAT analysis in mammalian cells revealed substantial heterogeneity in the degradation kinetics of nascent proteins, demonstrating the utility of this method in proteome-wide degradation profiling.[23] While the most widespread adoption of the BONCAT method involves the treatment of cells with azidohomoalanine (Aha), the mutant methionyl-tRNA synthetase NLL-MetRS efficiently charges $tRNA^{Met}$ with azidonorleucine (Anl), providing elevated methionine replacement rates and enabling state- or cell-selective analysis.[24–26] We leveraged the temporal resolution and sensitivity of these labeling strategies in combination with multiplexed isobaric tandem mass tag (TMT) labeling to profile the stability of nascent proteins in both exponential and stationary phase, a model of bacterial quiescence.[27,28] We identified instability in a diverse panel of homeostatic and stress response regulators and mapped select substrates to their cognate proteases. We then used our identifications to train and validate a machine learning classifier for predicting in vivo protein stability, identifying the active degradation of key regulators driving motility and surface adhesion. Taken together, our observations expand the catalog of established ATP-dependent degradation in diverse aspects of microbial physiology.

2.3 Results

Text mining establishes a reference set of actively degraded proteins

While the study of regulated proteolysis in bacteria is an established field, substrate identifications are not thoroughly reflected in existing knowledge bases. To consolidate knowledge on this topic, we first reviewed the literature to manually curate the identities of proteins bearing direct or inferential evidence of regulated proteolysis, which we refer to as annotated substrates. We then constructed an integrated corpus of 30,238 documents by querying PubMed abstracts or EcoCyc gene summaries for terms related to protein degradation in *E. coli* to augment our manual curation (Fig. S2.1).[29] Sentences were searched for query terms and *E. coli* gene names to yield candidate annotations that were then approved or rejected in a specialized Shiny application.[30] Frequently identified genes in this query were classically associated with protein degradation (Fig. 2.1A). Text mining expanded our set of manually curated protease annotations by 41%, with each knowledge base yielding unique identifications (Fig. S2.2). Concatenation of text-mined and manually curated annotations yielded a dense network of 375 proteins enriched in computational and experimental interactions, suggesting these proteins are functionally associated (Fig. 2.1B). Enriched annotations from the UniProt and Kyoto Encyclopedia of Genes and Genomes (KEGG) knowledge bases included functions that are prototypical of regulated proteolysis in bacteria such as DNA repair, metal homeostasis, and sigma factor activity in addition to others such as RNA degradation, acetylation, and protein export that are not broadly recognized in reviews of regulated proteolysis in bacteria (Fig. 2.1C).[7,31–34]

BONCAT identifies proteolytic degradation in exponential phase

We next sought to expand our annotation set by profiling the stability of nascent protein populations with the BONCAT technique. We treated exponential phase *E. coli* MG1655 cells with Aha for 30 min to label nascent proteins in this physiological state. To eliminate the limiting effect of growth on the quantitative dynamic range attributable to proteolysis in subsequent proteome-wide analysis, we employed the bacteriostatic translation inhibitor chloramphenicol as a chase treatment in analogous manner to cycloheximide chases in eukaryotic proteome turnover analyses.[35,36] Lysates of cells harvested at selected time points were treated with tetramethylrhodamine alkyne (TAMRA-alkyne) under copper-catalyzed click conditions, and proteins were resolved by mass via SDS-PAGE. In-gel fluorescence measurements of TAMRA-conjugated proteins suggest these Aha incorporation levels are not broadly

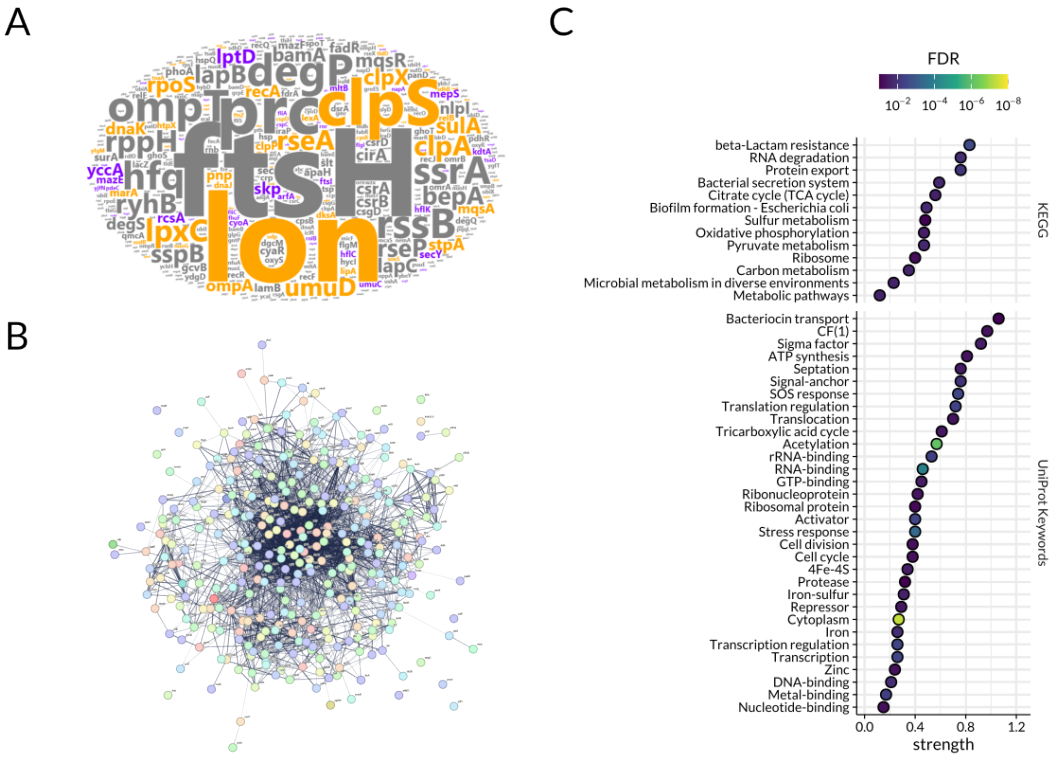

Figure 2.1: **Text mining of the PubMed and EcoCyc databases.** (A) Word cloud summarizing the gene counts across the query-satisfying sentences of the EcoCyc gene summaries. Gene names are scaled by their proportional representation. Orange names indicate evidence of proteolytic degradation in the literature from manual curation, grey names indicate a lack of this evidence, and purple names report genes absent in our manual curation bearing evidence of instability as reported in this knowledge base. (B) Protein-protein interaction (PPI) network for the 375 proteins bearing evidence of instability in the literature (PPI enrichment p-value < 1×10^{-16}). Edges indicate computational and experimental evidence of association. (C) Functional enrichment analysis in the UniProt and KEGG databases for annotated proteins.

destabilizing, as previously observed in mammalian cells, and recapitulate observations that most *E. coli* proteins are stable (Fig. 2.2A).[19,23]

To globally profile protein degradation kinetics, we prolonged the observation window (0, 0.5, 1, 2, and 4 h) in a large-scale experiment and covalently enriched labeled proteins with a dibenzocyclooctyne agarose (DBCO-agarose) resin. Enriched proteins were then digested and TMT labeled for liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis. We quantified 1607 proteins and fit their normalized time-series abundances to an exponential model, extracting the rate constant as an estimator for proteolytic instability (Fig. 2.2B). Density estimates of the rate

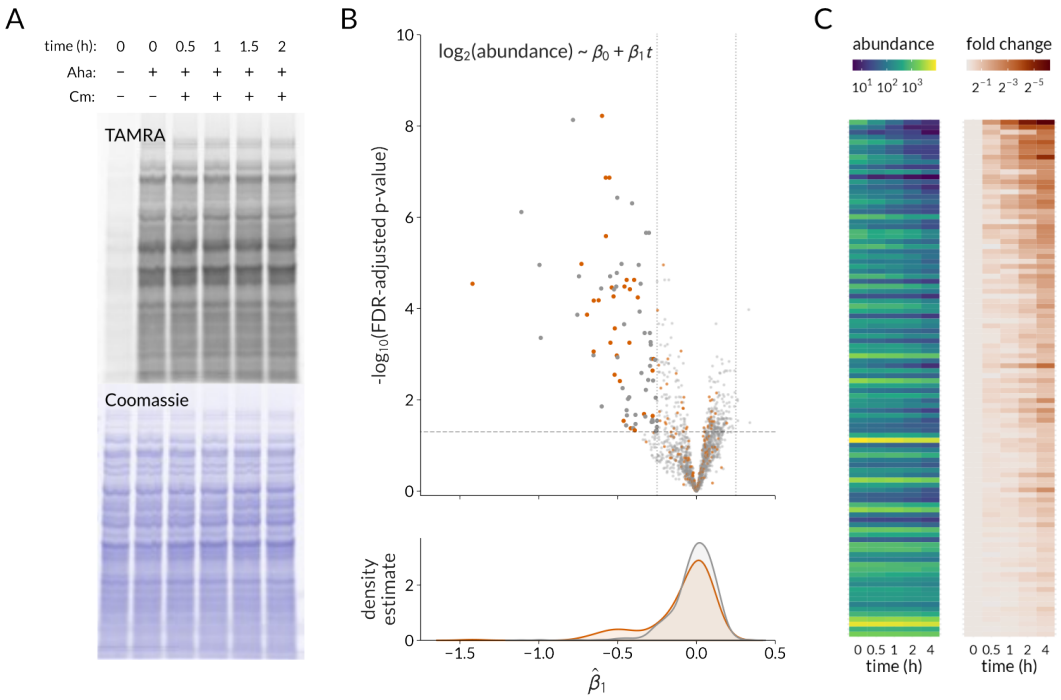

Figure 2.2: **Proteome-wide stability analysis in exponential phase cells.** (A) In-gel fluorescence of labeled proteins conjugated to TAMRA-alkyne. Coomassie staining reports total loading. (B) Statistical significance and kernel density estimates of fitted degradation rates stratified by annotation status. Vertical dotted lines indicate a fold change of two over the course of the experiment. Orange shading represents proteins bearing direct or inferential evidence of instability in the literature as assessed by manual curation or text mining, and grey shading indicates a lack of this evidence. (C) Heatmaps for proteins with evidence of degradation in the screen without (left) and with (right) normalization to their initial abundance.

constant distribution stratified by annotation status exhibited an expected bimodality consistent with a minority of proteins experiencing degradation. We identified a strong enrichment of annotated substrates (Fisher's exact test; p-value = 1.1×10^{-11}) among a set of 82 proteins with pronounced instability in our screen (fold change decrease > 2; FDR-adjusted p-value < 0.05), suggesting our method identifies established protease substrates. Time-series abundance decreases were seen across the quantitative dynamic range, with many candidate substrates exhibiting consistent chronological decreases in abundance (Fig. 2.2C).

The c-di-GMP phosphodiesterase PdeH is destabilized by ClpXP

The second messenger cyclic di-GMP (c-di-GMP) exerts transcriptional, translational, and post-translational regulation on its effectors to orchestrate diverse cel-

lular processes such as motility, surface adhesion, virulence, and cell division in bacteria.[37] In *E. coli*, c-di-GMP primarily influences motility and the production of the biofilm matrix components curli fiber and phosphoethanolamine cellulose, and the antagonistic activities of diguanylate cyclases (DGCs) and c-di-GMP phosphodiesterases (PDEs) determine steady-state c-di-GMP levels. We identified the PDE that exclusively licenses motility, PdeH, as a candidate substrate exhibiting a half-life of approximately 1.72 h (Fig. 2.3A). PdeH bears the simplest domain architecture of all 13 PDEs in *E. coli* MG1655, lacking accessory domains for localization or allosteric regulation and instead harboring solely a conserved EAL domain with an N-terminal extension poorly predicted by the AlphaFold model (Fig. 2.3B, Fig. S2.3).[38,39] To validate the instability of PdeH, we inserted its coding sequence fused to the FLAG epitope at either terminus into a medium-copy plasmid for arabinose-inducible expression in immunoblotting stability analysis.[40] The presence of a C-terminal ssrA-like tripeptide (LVA) in PdeH suggested proteolysis may initiate at this terminus. However, either blocking the N terminus with a FLAG tag or introducing successive N-terminal truncations stabilized PdeH, suggesting its N terminus mediates protease recognition (Figs. S2.4, S2.5). Ectopically expressed PdeH-FLAG in wild type MG1655 cells exhibited degradation that was nullified in *clpP* and *clpX* backgrounds, establishing PdeH as a client of ClpXP (Fig. 2.3C).

As a class III gene in the flagellar regulatory cascade, *pdeH* is transcriptionally activated by the ClpXP substrate and master regulatory complex FlhDC, thus rendering ClpXP a regulator of both PdeH synthesis and degradation.[41–44] The identification and subsequent mutagenesis of ClpXP recognition determinants in PdeH would allow us to isolate the role of proteolysis in directly shaping the endogenous PdeH abundance. Therefore, we subjected residues 2-21 of PdeH to alanine scanning or alanine stretch mutagenesis. The L19A point mutant and the X(8-13)A and X(15-19)A stretch mutants exhibited near-complete stabilization, further suggesting that this N-terminal extension confers ClpXP recognition (Fig. S2.6).

To ascertain the physiological relevancy of this proteolytic recognition, we then used a scarless Cas9 cloning system to generate strains bearing a stabilizing mutation in addition to a C-terminal 3xFLAG tag at the *pdeH* locus. Consistent with our proteomic analysis, PdeH-3xFLAG at endogenous expression levels exhibited a half-life of approximately 1.5 h, whereas the L19A and X(15-19)A mutants were completely stabilized (Fig. 2.3D). As a decrease in stationary phase PdeH abundance drives sessility and the expression of the central biofilm regulator CsgD, we then

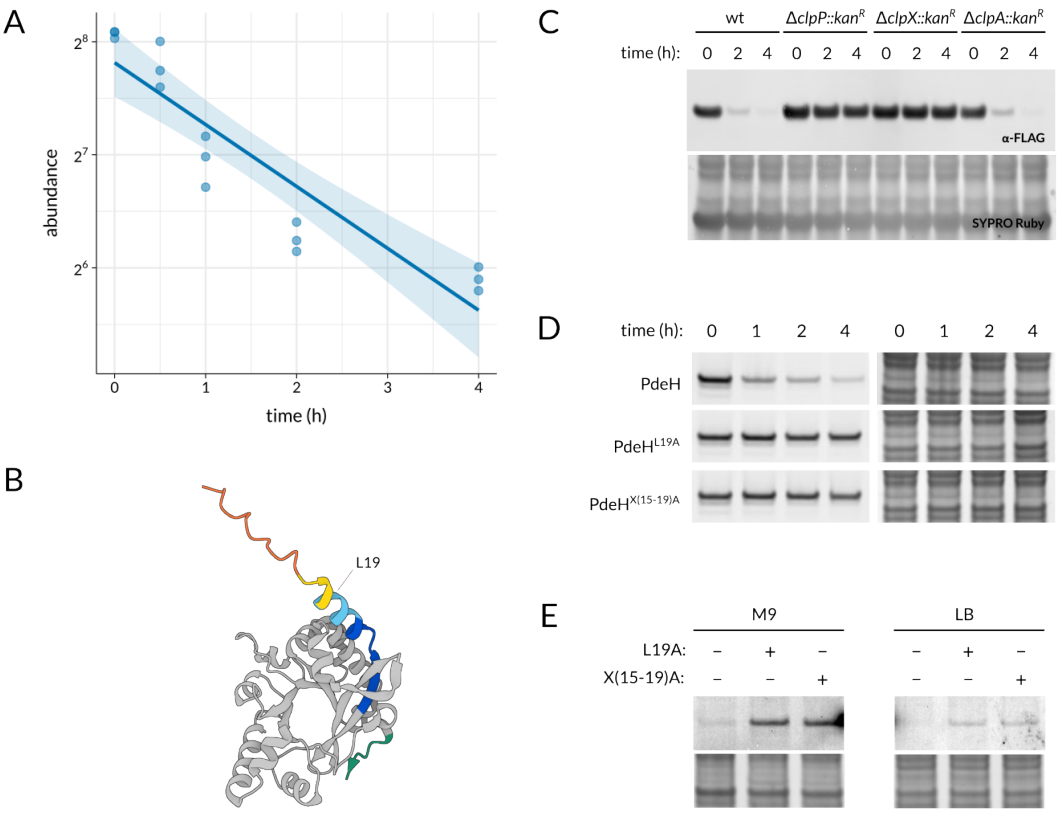

Figure 2.3: **ClpXP-mediated degradation of the principal c-di-GMP phosphodiesterase PdeH.** (A) Exponential fit of the PdeH abundance trace from our BONCAT analysis. The shaded region indicates a 95% confidence interval for the predicted values. (B) AlphaFold approximation of PdeH structure. The conserved c-di-GMP EAL domain is shaded in grey, and the N-terminal extension is colored by the prediction's per-residue model confidence score (pLDDT). The yellow and orange regions indicate low and very low model confidence scores, respectively. (C) Immunoblotting stability analysis of PdeH-FLAG expressed from a plasmid (pBAD33) in wild type or protease deletion mutant strains. (D) Immunoblotting stability analysis of wild-type or mutant PdeH-3xFLAG expressed from the *pdeH* locus. (E) Immunoblotting of wild type or mutant PdeH-3xFLAG expressed from the *pdeH* locus in stationary phase. Cells were cultured in M9 or LB medium.

monitored the abundance of this protein in stationary phase.[45] Cells bearing an L19A mutation at the *pdeH* locus exhibited an approximate six-fold increase in the abundance of PdeH in stationary phase when cultured in M9 (Fig. 2.3E). We conclude that the PdeH N-terminus is a destabilizing regulatory motif that facilitates ClpXP-mediated depletion of this enzyme at growth arrest.

The ClpAP modulator and N-degron recognin ClpS is degraded by ClpAP

The N-end rule pathway causally links N-terminal residue identity to proteolysis in bacteria and eukaryotes.[46,47] We observed degradation of the highly conserved N-degron recognin ClpS with a half-life of approximately 1.66 h (Fig. 2.4A).[48] ClpS chaperones N-end rule substrates for ClpAP-mediated degradation and attenuates ClpAP activity towards ssrA-tagged substrates, protein aggregates, and ClpA itself.[3,49] ClpS comprises a substrate- and ClpA-binding core domain preceded by an unstructured N-terminal extension (NTE) that reportedly moderates the ATPase activity and ssrA selectivity of ClpA by entering the ClpAP pore as a degron mimic (Fig. 2.4B).[50,51] In contrast with the documented in vitro stability of ClpS in ClpAP degradation reactions, immunoblotting stability analysis and protease mapping of ectopically expressed ClpS-3xFLAG demonstrated its in vivo ClpAP-mediated degradation (Fig. 2.4C).[49] Truncation of the NTE abolished degradation, suggesting engagement of the ClpAP pore by ClpS licenses its destruction.

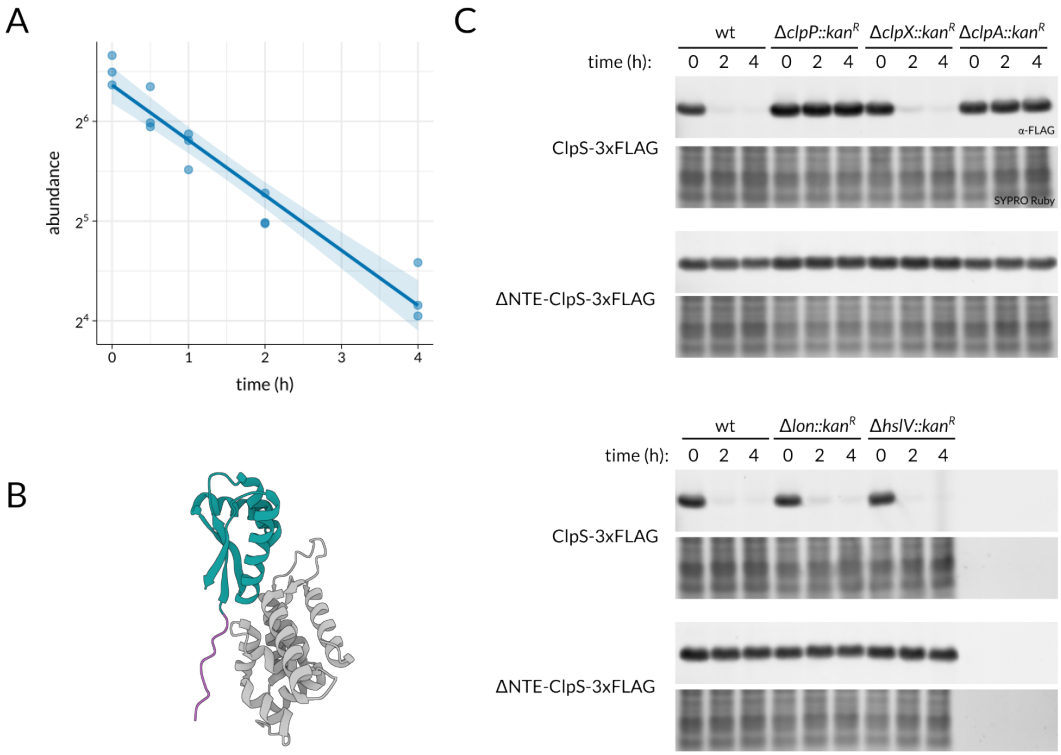

Figure 2.4: **ClpAP-mediated degradation of the N-degron recognin ClpS.** (A) Exponential fit of the ClpS abundance trace from proteome-wide screening. The shaded region indicates a 95% confidence interval for the predicted values. (B) Crystal structure of ClpS (green) bound to ClpA (grey).[52] Residues 16-25 of the unstructured ClpS NTE (purple) are indicated. (C) Immunoblotting stability analysis of ClpS-3xFLAG and ΔNTE-ClpS-3xFLAG expressed from a plasmid (pBAD33).

Fe-S cluster assembly proteins are differentially stable

E. coli stores approximately 30% of cellular iron in low-spin ferrous heme centers and prosthetic iron-sulfur (Fe-S) clusters, which are assembled on scaffold proteins and transferred to carrier proteins for allocation to their cognate apoproteins.[53,54] We observed instability across all four Fe-S cluster carrier proteins featuring the A-type domain, or A-type carriers (ATCs), including IscA, ErpA, SufA, and NfuA, which bears an A-type domain rendered defunct by mutation of conserved Fe-S cluster–liganding cysteines (Fig. 2.5A). We also observed instability in two components of the ISC and SUF Fe-S cluster biogenesis pathway scaffold complexes, IscU and SufD, which are essential for the activity of their respective scaffolds.

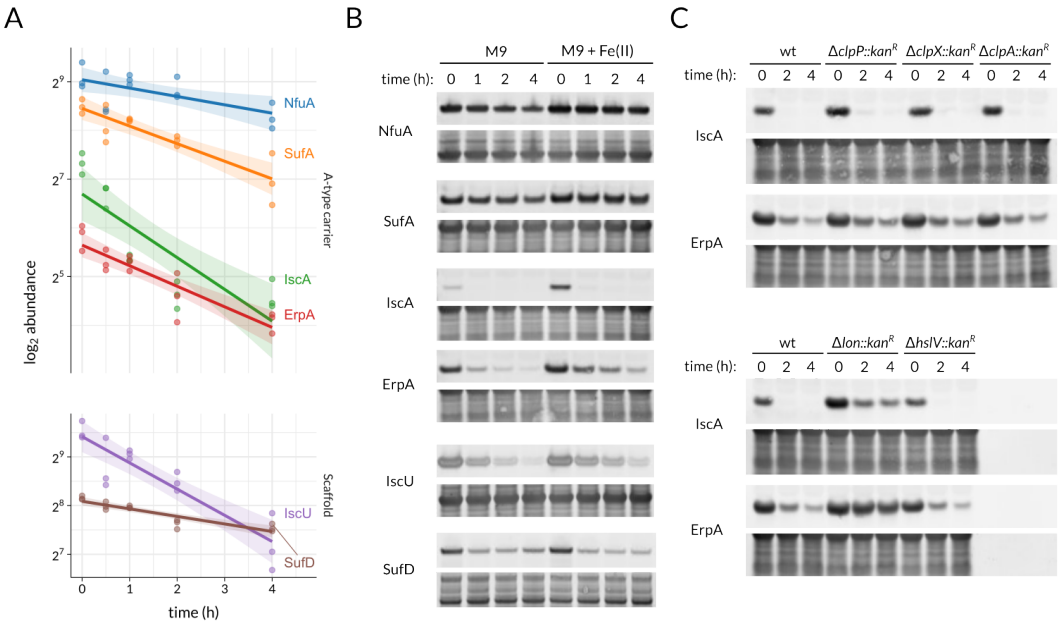

Figure 2.5: **Differential stability of A-type carriers and scaffold components in Fe-S cluster biogenesis pathways.** (A) Exponential fits of abundance traces for A-type carriers NfuA, ErpA, SufA, and IscA (upper panel) and scaffold components IscU and SufD (lower panel). The shaded regions indicate a 95% confidence interval for the predicted values. (B) Immunoblotting stability analysis of substrate candidates bearing an N-terminal 3xFLAG tag expressed from a vector (pBAD33) in cells grown with or without addition of 20 µM Fe^{2+} to the growth medium. (C) Protease mapping of ectopically expressed 3xFLAG-IscA and 3xFLAG-ErpA reveals partial and complete stabilization of these fusions in a Lon deletion background, respectively.

We fused the 3xFLAG epitope at either terminus of these ATCs and scaffold proteins, and we then inserted each fusion into a plasmid for ectopic expression in immunoblotting stability analysis. N-terminal 3xFLAG fusions reproduced the

abundance rank order and differential stability observed in our proteomic analysis (Fig. S2.7). While metal coordination is thought to influence protease engagement, the instabilities of IscA and the principal aerobic A-type carrier ErpA were independent of Fe^{2+} repletion (Fig. 2.5B) in M9 medium, which is deficient in iron.[55,56] We selected these two ATCs for protease mapping and discovered partial and complete stabilization in a *lon* background for IscA and ErpA, respectively (Fig. 2.5C). We conclude that the A-type domain may endow a carrier with Lon recognition that may be modulated by its sequence context.

BONCAT enables global protein stability profiling in stationary phase

Having established the fidelity of BONCAT in identifying protease substrates in bacteria, we next sought to leverage the sensitivity of this method by characterizing degradation in stationary phase, a model of microbial quiescence with established gene expression commonalities to the biofilm lifestyle.[57] We found that labeling stationary phase cultures with Aha produced insufficient methionine replacement for proteomic analysis (Fig. S2.8). To achieve adequate labeling, we transcriptionally fused a stringently regulated ribosomal RNA promoter, *rrnB* P1, to a gene encoding NLL-MetRS to enable its accumulation during exponential phase for Anl labeling in stationary phase.[58] Cells expressing the cassette from a low-copy plasmid were cultured for 24 h to stationary phase, pulse labeled with Anl for 4 h, resuspended in sterile-filtered spent medium collected from untreated cultures, and supplemented with Met as a chase treatment. We observed no growth resumption in stationary phase cultures upon treatment with Met or Anl with this protocol (Fig. S2.9). Lysates of cells removed at predetermined time points during the chase were treated with TAMRA-alkyne under copper-catalyzed click conditions, and proteins were then resolved by mass with SDS-PAGE for in-gel fluorescence visualization (Fig. 2.6A). We again observed widespread stability of the labeled proteins, suggesting Anl, in similar fashion to Aha, does not broadly destabilize recipient proteins.

We expanded this analysis for proteome-wide measurement by again scaling up this experiment and collecting cells over a prolonged observation window (0, 1, 2, 4, 8 h) for lysis. The Anl-labeled proteins were enriched, digested, and TMT labeled in analogous manner to the exponential phase screen. Samples in this protocol were additionally analyzed by the RTS-SPS-MS3 method to alleviate ratio compression arising from precursor ion cofragmentation.[59] Analysis of the enriched peptides yielded 1349 quantified proteins, and abundance traces again were fit to an exponential decay model for degradation rate constant estimation (Fig. 2.6B).

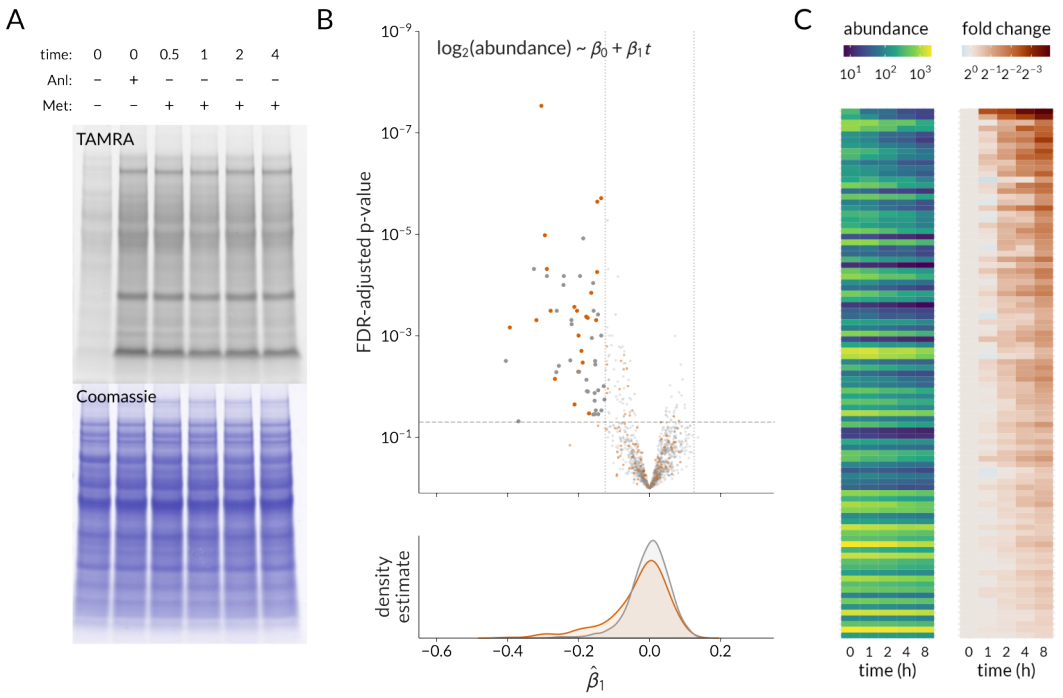

Figure 2.6: **Proteome-wide stability analysis in stationary phase cells.** (A) In-gel fluorescence of Anl-labeled proteins synthesized in stationary phase. (B) Statistical significance and kernel density estimates of fitted degradation rates stratified by annotation status. Vertical dotted lines indicate a fold change of two over the course of the experiment. Orange shading represents proteins bearing direct or inferential evidence of instability in the literature as assessed by manual curation or text mining, and grey shading indicates a lack of this evidence. (C) Heatmaps for proteins with evidence of degradation in the screen without (left) and with (right) normalization to their initial abundance.

Degradation rates were generally slower in stationary phase, consistent with reported differential instability measurements across a limited panel of model substrates.[60] We again observed a strong enrichment of established protease substrates (Fisher's exact test; p-value = 1.1×10^{-8}) among 55 high-confidence substrate candidates (fold change decrease > 2; FDR-adjusted p-value < 0.05) and gradual abundance decays in many substrate candidates exhibiting statistically significant degradation rates (Fig. 2.6C).

Stationary phase substrates are implicated in stress response

Among proteins bearing a high-confidence degradation rate estimate in either growth phase, we observed moderate overlap in the identities of annotated substrates between the two growth phases, whereas unannotated substrate candidates were distinct

to each growth phase (Fig. 2.7A). Degradation rates for annotated substrates across these physiological states were correlated (Pearson's $R = 0.63$), suggesting that both labeling strategies produce comparable identifications and that many established substrates are not conditionally unstable (Fig. 2.7B). Candidate substrates identified exclusively in the stationary phase dataset were often implicated in stress response, including the oxidative stress regulator SoxS and the canonical N-end rule substrate PatA, whereas those identified exclusively in exponential phase were associated with motility and flagellar assembly, which terminate in stationary phase.[4,45,61]

One of the most destabilized stationary phase candidates we observed was RpoZ, the nonessential ω subunit of RNA polymerase implicated in biofilm production and sigma factor selectivity (Fig. 2.7C).[62] In addition, we observed instability of subunits β and β', which were previously identified as unstable in stationary phase.[63] Stationary phase immunoblotting stability analysis of RpoZ fused at either terminus to the 3xFLAG epitope demonstrated a gradual decline in abundance (Fig. 2.7D). We also selected stationary phase–exclusive candidate substrates for immunoblotting stability analysis and identified instability in N-terminal 3xFLAG fusions of BolA, which mediates stationary phase morphology and was previously seen to be proteolytically unstable, and BssR, a regulator of biofilm production linked to quorum sensing and indole transport (Fig. S2.10).[64–66] Protease mapping identified ClpAP as the cognate protease for these regulators (Fig. 2.7E).

Machine learning–assisted substrate identification uncovers proteolysis

To address the abundance bias and physicochemical limitations that preclude full coverage in mass spectrometry–based proteomics, we sought to build and validate a predictive classification model for determining which proteins among those not identified in our proteomic profiling may also be subject to proteolytic regulation. Our training data consisted of 2023 proteins quantified in our proteomic analyses, and the 286 of these proteins bearing statistically significant evidence of instability served as positive examples. A total of 9795 nonredundant features were sourced from the UniProt, EcoCyc, and AlphaFold databases or computed from primary amino acid sequences.

Four models were considered for this classification task: logistic regression with L1 regularization, random forest, artificial neural network (ANN), and gradient-boosted logistic regression with XGBoost.[67] Models were evaluated by subjecting training data to repeated nested cross validation with stratification and sample weighting

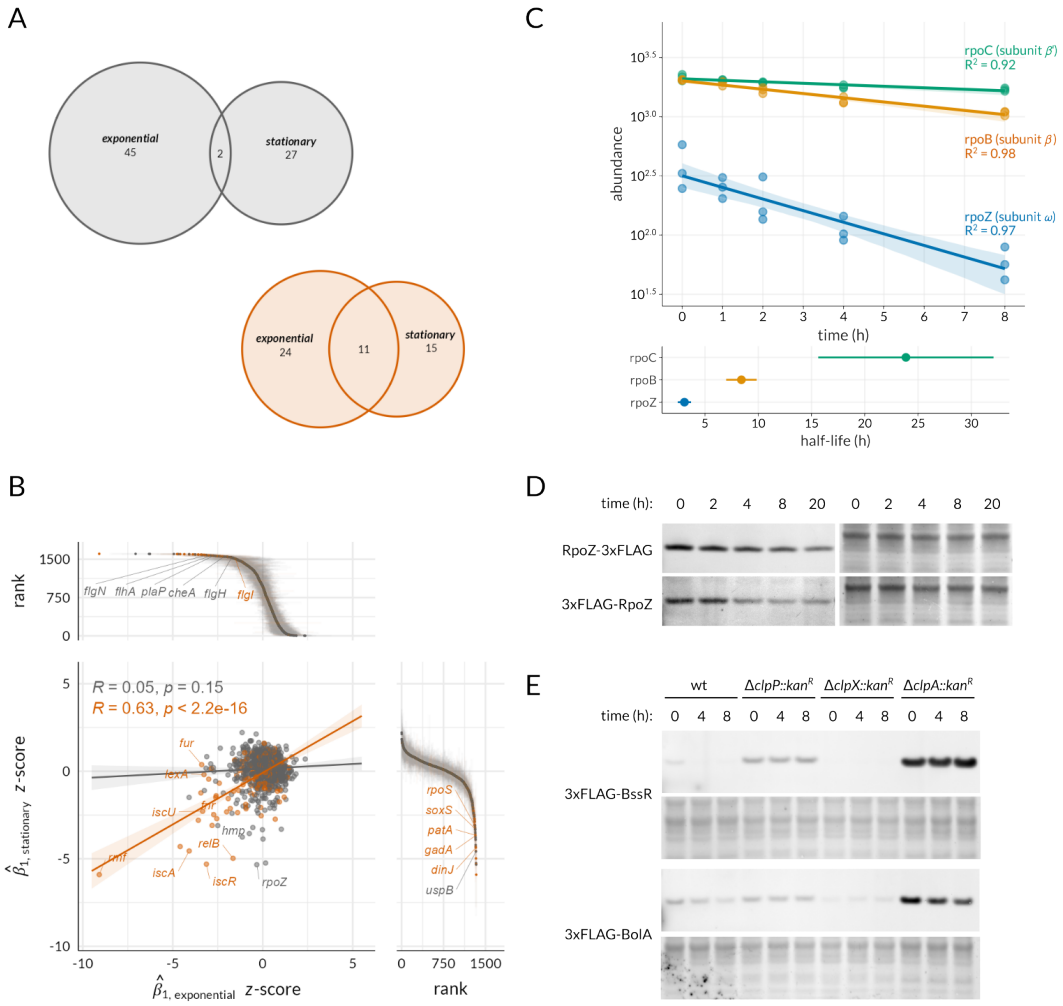

Figure 2.7: **Comparison of exponential and stationary phase degradation.** (A) Venn diagrams of high-confidence substrate candidates (fold change decrease > 2; FDR-adjusted p-value < 0.05) in each substrate annotation class. Orange shading represents proteins bearing direct or inferential evidence of instability in the literature as assessed by manual curation or text mining, and grey shading indicates a lack of this evidence. (B) Joint plot comparing degradation rates across mutually identified proteins in the scatter plot panel. The rate constant distributions in each condition contain both mutually and exclusively identified proteins. Select exclusively identified proteins are labeled by gene name. (C) Model fits and half-life estimates for the subunits of RNAP exhibiting instability in the stationary phase screen. The shaded regions indicates a 95% confidence interval for the predicted values. (D) Immunoblotting stability analysis of RpoZ bearing a 3xFLAG tag at the N or C terminus expressed from a vector (pBbA2c) in stationary phase. (E) Protease mapping of 3xFLAG-BssR and 3xFLAG-BolA expressed from a vector (pBbA2c) in stationary phase across wild-type and protease deletion mutant backgrounds.

to address class imbalance, optimizing for mean average precision. The XGBoost model achieved the highest mean average precision (0.371), representing a 2.75-fold increase over the baseline precision for this dataset (0.139), and comparable area under the curve (AUC) for the receiver operating characteristic (ROC) curve (0.723) to that of the ANN model class (Fig. 2.8), providing an improvement over an existing model in discriminating unstable from stable proteins.[16] We therefore selected the XGBoost model for its interpretability and leading precision.

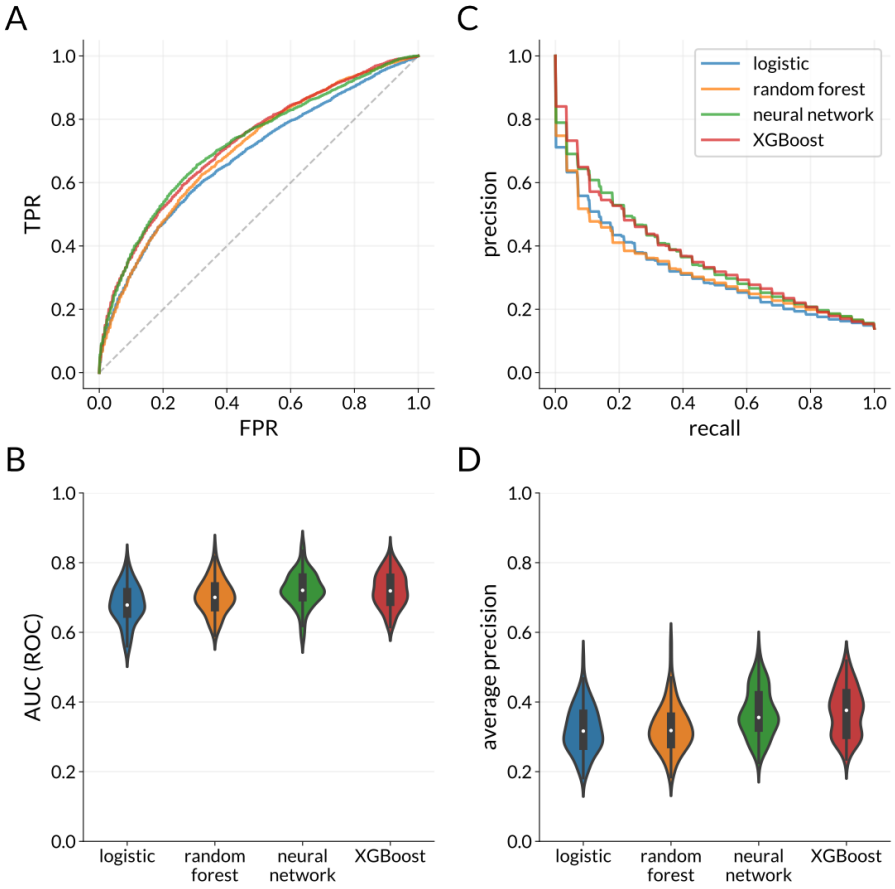

Figure 2.8: **An XGBoost classifier improves precision three-fold over baseline** (A) Receiver operating characteristic (ROC) curves for models. (B) Violin plots of area under the curve (AUC) calculations for ROC curves in each test set from repeated nested cross validation. (C) Precision-recall (PR) curves for models. (D) Violin plots of average precision in each test set from repeated nested cross validation.

Competitive enrichment of probability-ranked proteins conducted on the STRING database server yielded functionalities associated with proteolysis in the literature and in our proteomic profiling (Fig. 2.9A). Our model tended to assign higher probabilities to unseen proteins identified as protease substrates in the literature

(p-value = 4.6×10^{-10}; two-sided Mann–Whitney U test) than to proteins lacking documented evidence of instability in our curation, which were deemed on average as unstable or stable, respectively (Fig. 2.9B). Prototypical protease substrates such as the heat shock sigma factor RpoH, antibiotic resistance regulator MarA, and FlhC and FlhD subunits of the FlhDC complex were assigned to the proteolytically unstable class by the model.[42,61,68]

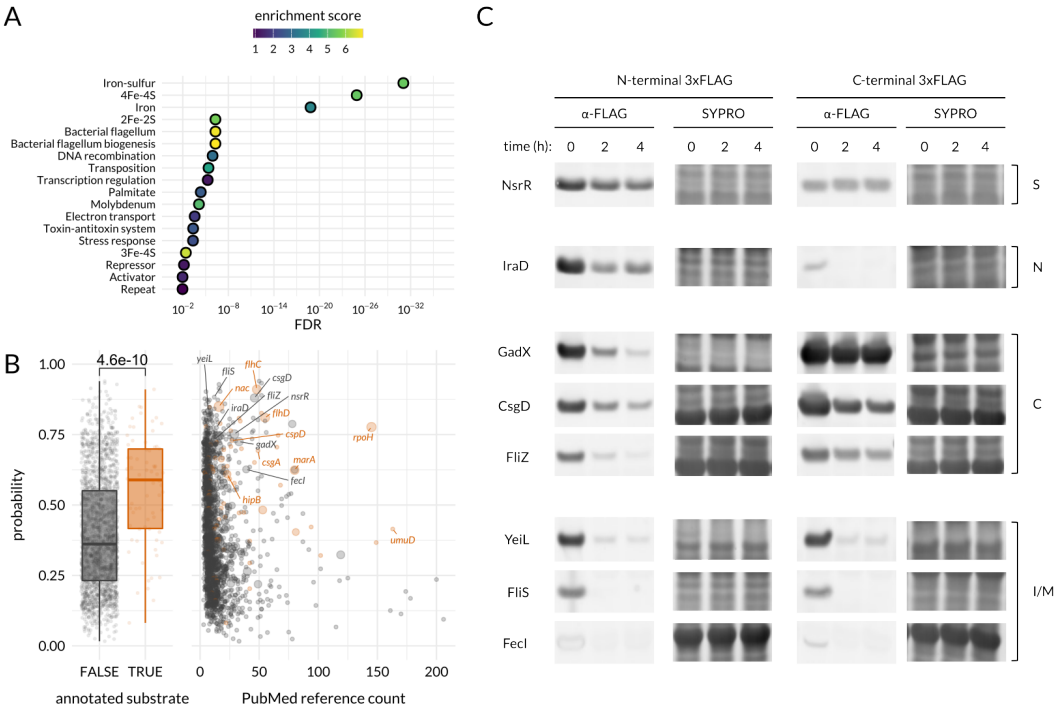

Figure 2.9: **Prediction and validation of instability in unseen proteins.** (A) Enriched UniProt keywords from competitive functional enrichment analysis of unseen proteins ranked by their assigned probability of instability. (B) Instability probabilities stratified by annotation status and PubMed reference count for proteins unseen by the model. (C) Immunoblotting stability analysis of selected candidate substrates predicted as unstable by the model. Substrates were fused with the 3xFLAG epitope at either terminus and expressed from a medium-copy plasmid (pBAD33). Candidates were assigned as stable (S) or unstable based on the presence of an N-degron (N), C-degron (C), or an internal degron or multiple degrons (I/M).

We then investigated the stability of unannotated high-confidence predicted substrates chosen for their representation across predictive features and for their known biological functions. Immunoblotting stability analysis in exponential phase cells yielded time-series reduction in abundance of two anti-adaptor proteins, IraD and FliS, which neutralize the RpoS-binding RssB and FliA-binding FlgM adaptors, respectively; the iron-responsive extracytoplasmic sigma factor FecI; FNR homolog

YeiL; a transcriptional activator of the acid resistance system, GadX; the RpoS antagonist and motility regulator FliZ; and the transcriptional activator CsgD, which is required for curli fiber expression (Fig. 2.9C).

Lon-mediated degradation of motility and biofilm regulators FliZ and CsgD

Given the high representation of the flagellar regulatory cascade in our experimental and computational identifications, we selected CsgD and FliZ for protease mapping and found complete stabilization of these two substrates in a *lon* background (Fig. 2.10A). Notably, CsgD bears a C-terminal LuxR-type helix-turn-helix domain also found in Lon substrates RcsA and GadE (Fig. 2.10B) and is regulated by the greatest number of transcriptional regulators across the *E. coli* proteome, both being features deemed predictive by our model (Fig. S2.11).[29]

A genetic analysis of the flagellar and curli expression cascades revealed the class I master regulator FlhDC indirectly inhibits curli expression and maintains motility by driving expression of FliZ and PdeH as class II and class III gene products, thereby inhibiting expression of CsgD.[45] Our results indicate all three of these regulators are actively degraded in vivo (Fig. 2.10C). Thus, we conclude that proteolytic degradation moderates the abundance of key regulators at every level of the flagellar gene cascade hierarchy.

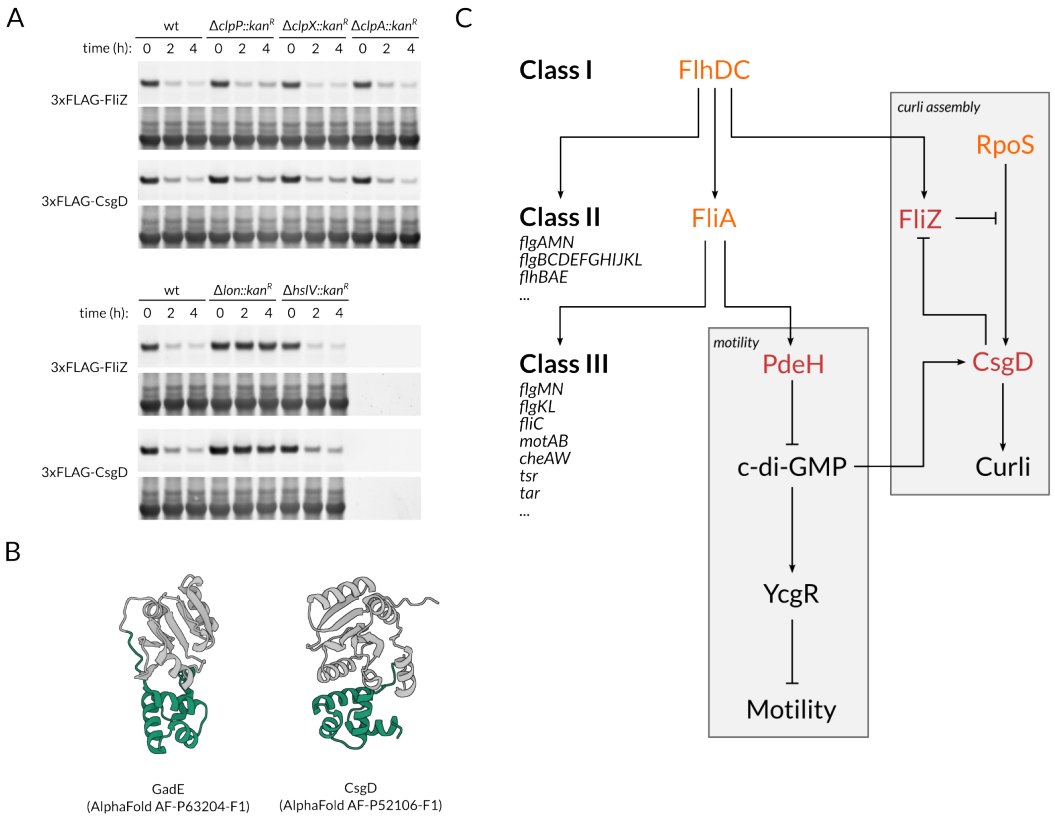

Figure 2.10: **Lon-mediated proteolysis of FliZ and CsgD** (A) Protease mapping of 3xFLAG-FliZ and 3xFLAG-CsgD expressed from a vector (pBAD33) in exponential phase across wild-type and protease deletion mutant backgrounds. (B) AlphaFold approximations of Lon substrate GadE and CsgD. The C-terminal LuxR-type helix-turn-helix domains are highlighted in green. (C) Simplified genetic circuit illustrating regulation of motility and surface adhesion, with newly identified protease substrates (red) and known substrates (orange) indicated.

2.4 Discussion

Much of the proteolytic degradation established here pertains to motility and the regulation of c-di-GMP abundance. As the timed destruction of PdeA drives cell cycle progression in *Caulobacter crescentus*, the regulated proteolysis of a c-di-GMP PDE is not unprecedented. A stabilizing point mutation in the N-terminal Per-ARNT-Sim (PAS) domain of PdeA abrogated recognition by protease chaperone CpdR.[69] The PdeH N-terminal extension has neither structural information nor assigned functions, and our results do not exclude the possibility of an unidentified chaperone recognizing this portion of the protein for subsequent ClpXP degradation. In support of this prospect, the stabilizing L19A mutation is exceptionally far into the primary sequence relative to many other established ClpXP degrons.[70] A comparative analysis identified absolute conservation of PdeH across 61 *E. coli* strains including both commensals and pathogens.[38] Whether this proteolytic regulation exists and influences c-di-GMP in these and other *Enterobacteriaceae* members warrants further investigation. We also note that the PdeH-antagonizing DGC, DgcE, was reported to degrade in vivo, though the protease responsible for the observed instability could not be identified.[71,72] As c-di-GMP lies at the heart of many lifestyle changes and irreversible cell fate commitments, regulated proteolysis of PDEs and DGCs may prove to be a continuing theme in further study of microbial physiology.

The *clpS* promoter is downregulated by low cytoplasmic Mg^{2+}, and elevated ClpA abundance in stationary phase decreases the ClpS:ClpAP ratio, suggesting ClpS abundance and relative stoichiometry are tuned to address stress.[73,74] Degradation of ClpS may contribute to this end, and its occurrence at both physiological and excess stoichiometries suggests proteolysis may proceed constitutively, though further study should probe any potential stabilizing effects of N-end rule substrate and ClpA binding. How ClpS may escape degradation by ClpAP has not been established in contemporary models across numerous biochemical, biophysical, and structural studies.[51] The prospect of adaptor-substrate co-degradation has been recognized as compatible with other mechanistic aspects of the ClpS-ClpAP system and is seen in the competence-regulating ClpCP adaptor MecA in *C. crescentus*.[75,76] The conflicting outcomes of published in vitro and in vivo ClpS stability analyses may further suggest an unconsidered variable or unidentified factor absent in reconstituted in vitro systems is involved in ClpS substrate exchange with ClpAP. Indeed, an unexplained in vivo stabilization of ssrA-tagged substrates against ClpAP was

eventually attributed to the SspB and ClpS adaptors upon their discovery and characterization.[2,49,77,78]

ISC to SUF switching drives antibiotic tolerance to aminoglycosides and fluoroquinones through impaired assembly of respiratory complexes and enhanced assembly of oxidative stress response regulators, respectively.[79,80] As our exponential phase analysis reflects proteome remodeling upon aminoglycoside treatment, the differential instability of essential Fe-S scaffold complex subunits may contribute to killing dynamics during time-series viability measurements. More broadly, much remains to be learned about the relative activities of ATCs across physiological states and the role of proteolysis therein, as both the origin of iron in Fe-S assembly and its fate when bound to a proteolytically degraded carrier are not fully understood.[54] While ErpA is important for aerobic growth and thought to be a gatekeeping carrier by receiving assembled clusters from IscA, SufA, and NfuA, some notable apoproteins rely specifically on a single ATC.[56,81,82] For example, the anaerobic transition regulator FNR relies exclusively on IscA for its Fe-S cluster.[82] The role of proteolysis in determining the activities and relative ratios of the ISC or SUF assembly pathways, and thus the coordination status of their apoproteins, merits further analysis.

While our classifier achieved actionable performance, the 9795 ontological, regulatory, and physicochemical protein features did not strongly capture all proteins with measured instability in our proteomics analysis. Regardless, our experimental and computational analyses still presented some notable themes in substrate attributes. The intrinsically disordered region of FtsZ contributes to its recognition and degradation by ClpXP, suggesting that an absence of defined structure may serve as a recognition determinant as observed in eukaryotes.[83,84] Some substrates in our analysis possess termini that are poorly predicted by AlphaFold (e.g., PdeH and BolA), which is predictive of intrinsic disorder in the AlphaFold2 model, or alternatively have explicitly described intrinsic disorder or conformational flexibility in the literature (e.g., ClpS and RpoZ).[50,85–87] Whether proteolytic degradation mediated by intrinsic disorder is adventitious or instead leveraged for regulatory purposes may be contextual. The regulatory network also provided predictive power for several of the validated substrates. In particular, the number of transcriptional regulators mediating a given protein's abundance served as a predictive feature, suggesting a relationship exists between the degree of transcriptional and post-translational regulation governing protein abundance. Furthermore, several proteins subject to active

degradation in our proteomic profiling are elevated in abundance due to transposition events in their respective promoters: differential expression of the catabolic galactitol operon genes (*gatYZABCD*) and FlhDC-regulated genes such as PdeH is well documented in comparative genetics and proteomics studies of *E. coli* K-12 lineages.[88,89] Proteolysis may in part attenuate or render dynamic the protein overexpression elicited by mobile genetic elements.

We have demonstrated that a bacteriostatic aminoglycoside pulse chase is a relatively simple method sufficient for global identification of known and novel protease substrates. While Aha labeling has proven effective in globally profiling protein degradation across multiple studies, this report also details the first application of Anl for time-series degradation profiling. We anticipate this approach would similarly enable degradation profiling in other biological contexts distinguished by low translational activity. As expression of NLL-MetRS may be tied to pre-determined transcriptional states, degradation profiling in defined cellular subpopulations could further elucidate differential instability of proteins across physiologies.

2.5 Acknowledgements

We thank Dr. Megan Bergkessel and Dr. Alexander Varshavsky for their perspectives on this work. E.J.M. was supported in part by a National Science Graduate Research Fellowship. This work was supported by the Institute for Collaborative Biotechnologies through cooperative agreement W911NF-19-2-0026 from the U.S. Army Research Office. The content of the information does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

2.6 Materials and Methods

Strain and plasmid construction

Gene deletion mutants were constructed with the Datsenko and Wanner method using template plasmid pKD13 and recombinase plasmid pKD46 as previously described.[90]

Genes encoding all substrates were amplified directly from the MG1655 genome. Genes encoding exponential phase substrates were inserted into a pBAD33 vector encoding the FLAG or 3xFLAG tags for arabinose-inducible gene expression using Gibson assembly.[40,91] Vectors for alanine scanning mutagenesis were constructed by site-directed mutagenesis with phosphorylated primers. Genes encoding stationary phase substrates were inserted into a pBbA2c vector encoding the 3xFLAG tag for anhydrotetracycline-inducible gene expression.[92] Carbenicillin, kanamycin, spectinomycin, and chloramphenicol were used at 100 μg/mL, 50 μg/mL, 50 μg/mL, and 25 μg/mL, respectively, for plasmid maintenance where appropriate.

For chromosomal modification of the *pdeH* locus, mutation and epitope tag insertion with the Cas9 system from Jiang et al. was performed as described.[93] The N20 sequence for the single guide RNA targeting *pdeH* was selected from a genome-scale library thresholded for strict off-target quality control with optimal CRISPR/Cas9 activity.[94,95] The donor DNA for constructing pdeH(L19A)-3xFLAG was supplied in pTarget, whereas the donor DNA for the alanine stretch mutant pdeH(X(15-19)A)-3xFLAG was supplied as a fragment. Donor DNA was constructed by Gibson assembly to incorporate 500 bp of homology flanking the mutation and a PAM-silencing silent mutation. Insertion and sequence fidelity were verified by PCR and Sanger sequencing. The curing of pTarget and pCas was verified by recovery of antibiotic sensitivity to spectinomycin and kanamycin.

Media and growth conditions

M9 glycerol medium consisted of 0.2% glycerol, 1 mM $MgCl_2$, 0.1 mM $CaCl_2$, 3 g/L KH_2PO_4, 0.5 g/L NaCl, 6.78 g/L Na_2HPO_4, and 1 g/L NH_4Cl supplemented with 5 mg/L L-methionine and 40 mg/L of the remaining canonical amino acids. LB consisted of 10 g/L tryptone, 5 g/L yeast extract, and 10 g/L sodium chloride. Growth conditions were 37 °C at 250 RPM in an orbital shaker unless otherwise noted.

Text mining of PubMed and EcoCyc databases

The Entrez Direct utilities were used to query PubMed for terms related to ATP-dependent protein degradation in *E. coli*. Fetched PubMed abstracts or EcoCyc gene summaries were tokenized to sentences and words with the *tidytext* R package. Sentences containing query terms and an identified *E. coli* gene name were inspected in a specialized Shiny application.[30] Publications with sentences describing proteolytic regulation of an unannotated gene were manually inspected for final approval. Protein-protein interaction and functional enrichment analysis of the 375 annotated substrates was performed on the STRING database server.[96]

In-gel fluorescence measurements

M9 glycerol was inoculated from a glycerol stock of *E. coli* MG1655 for overnight growth. Cells were then diluted to an OD_{600} of 0.025 in fresh M9 glycerol and grown to to an OD_{600} of 0.75. Cells were treated with 1 mM Aha for 30 min and then treated with 100 µg/mL chloramphenicol. Cells were removed from the culture at the indicated time points, lysed in 0.5% SDS in PBS at 95°C for 15 min, and stored at -20°C for further processing.

For click reaction conjugation, lysates were thawed, clarified for 20 min at 18,000 RCF, and transferred to new microcentrifuge tubes. Protein concentrations were assessed with bicinchoninic assay (Thermo Scientific) and equalized with the addition of lysis buffer. Click reactions were performed as outlined in Hong et al. with 10 µM TAMRA-alkyne, 250 µM $CuSO_4$, 1.25 mM tris(3-hydroxypropyltriazolylmethyl)amine (THPTA), 5 mM (+)-sodium L-ascorbate, and 5 mM aminoguanidine hydrochloride.[97] Reactions proceeded for 1 h at 37 °C and were then mixed with 10X Reducing Agent and 4X LDS Fluorescence-Compatible Loading Dye (Thermo Fisher), heat treated at 70°C for 10 min, and resolved by SDS-PAGE on 4-12% Bis-Tris NuPAGE gels (Invitrogen). Gels were destained in 10% acetic acid, 40% methanol, and 50% deionized water overnight. The next day, gels were rehydrated in deionized water and imaged on a Typhoon Trio (GE). Gels were then incubated in InstantBlue Coomassie Protein Stain (Expedeon) for at least 2 h and imaged again on a Typhoon Trio (GE).

Immunoblotting stability analysis

For screening exponential phase substrates, overnight cultures of cells carrying a plasmid for expression of an epitope-tagged substrate inoculated from glycerol stocks

were diluted to OD_{600} 0.025 in fresh M9 glycerol medium. For screening the stability of Fe-S cluster biogenesis pathway substrates, the medium was supplemented with 20 μM $FeSO_4$ for the iron replete condition. Cultures were grown to an OD_{600} of 0.5 and then induced for substrate expression with the addition of 0.2% arabinose for 30 min. Spectinomycin was added to 200 μg/mL in the culture medium, and aliquots of cells were withdrawn from the culture at the indicated time points. Cells were lysed in 0.5% SDS in PBS, heat-treated at 95°C for 15 min, and stored at -20°C for further processing.

For screening the stability of PdeH produced from the endogenous locus, M9 glycerol medium was inoculated from a glycerol stock for overnight growth. The next day, cells were diluted to an OD_{600} of 0.025 in fresh medium and grown to an OD_{600} of 0.5. Cells were then treated with 200 μg/mL spectinomycin, and aliquots of cells were withdrawn from the culture at the indicated time points. Cells were lysed in 0.5% SDS in PBS, heat-treated at 95°C for 15 min, and stored at -20°C for further processing.

For screening the abundance of PdeH produced from the endogenous locus in stationary phase, M9 glycerol medium was inoculated from a glycerol stock for overnight growth. The next day, cells were diluted to an OD_{600} of 0.025 in fresh medium. Cells were then collected after 24 h of incubation and were lysed in 0.5% SDS in PBS, heat-treated at 95°C for 15 min, and stored at -20°C for further processing.

For screening stationary phase substrates, M9 glycerol medium was inoculated from a glycerol stock for overnight growth. The next day, cells were diluted 1:200 for 24 h growth. Cultures were then treated with 200 μM anhydrotetracycline for 4 h to induce protein expression, after which spectinomycin was added to 200 μg/mL to terminate protein synthesis. Cells were removed at the indicated time points, collected at 18,000 RCF for 1 min, and lysed in 0.5% SDS in PBS by heat treatment at 95°C for 15 min. Lysates were stored at -20°C for further processing.

Protein lysates were thawed, clarified for 20 min at 18,000 RCF, and transferred to new microcentrifuge tubes. Protein concentrations were assessed with bicinchoninic assay (Thermo Scientific) and equalized with the addition of lysis buffer. Lysates were then mixed with 10X Reducing Agent and 4X LDS Fluorescence-Compatible Loading Dye (Thermo Fisher), heat treated at 70°C for 10 min, and resolved by SDS-PAGE on 4-12% Bis-Tris NuPAGE gels (Invitrogen). Proteins were transferred to nitrocellulose membranes (0.2-μm pore size, Thermo Fisher Scientific) with the

iBlot 2 Dry Blotting System (Invitrogen). Protein loading was assessed by staining with SYPRO Ruby (Thermo Fisher) per manufacturer instructions and imaged on a Typhoon Trio (GE). Membranes were then blocked with 5% milk in PBST for 1 h and treated with anti-FLAG M2 antibody (Millipore Sigma) diluted 1:5000 overnight at 4°C. Membranes were washed in four consecutive rounds with PBST for 5 min each, treated with Goat anti-Mouse Alexa Fluor 647 antibody (Invitrogen #A32728) diluted 1:10000 for 1.5 h, washed again in four consecutive rounds with PBST for 5 min each, and then imaged on a Typhoon Trio (GE).

Chemoproteomic enrichment

For the exponential phase enrichment, three replicate cultures for each time point were inoculated from an overnight culture to an OD_{600} of 0.01 and grown to OD_{600} 0.75. Then, cultures were treated with 1 mM Aha for 30 min and subsequently chased with 100 μg/mL chloramphenicol. Cultures were centrifuged at 5000 RCF for 10 min and then resuspended in 4% SDS in 100 mM Tris (pH 8.0) supplemented with cOmplete EDTA-free protease inhibitor for lysis. Lysates were heat treated at 95°C for 10 min and frozen at -80°C for further processing.

For the stationary phase enrichment, six cultures were inoculated from individual frozen glycerol stocks to and OD_{600} of 0.01 and grown for 24 h. Then, three cultures were treated with 1 mM Anl for 4 h. The remaining three cultures were centrifuged at 5000 RCF for 10 min, and supernatants were collected, pooled, and sterile filtered to provide spent medium for the chase period. Anl-treated cultures were centrifuged at 5000 RCF for 10 min and resuspended in spent medium supplemented with 1 mM Met. Cells were then removed from each culture at the specified chase time points (0, 1, 2, 4, 8 h), centrifuged at 5000 RCF for 5 min, and resuspended 4% SDS in 100 mM Tris (pH 8.0) supplemented with cOmplete EDTA-free protease inhibitor for lysis. Lysates were heat treated at 95°C for 10 min and frozen at -80°C for further processing.

Thawed lysates were probe sonicated for 30 s at 20% amplitude (QSonica) with 2 s pulses and 1 s rests to shear chromosomal DNA. Lysates were then clarified by microcentrifugation at 18000 RCF for 20 min and transferred to new microcentrifuge tubes. Protein concentrations were assessed with bicinchoninic assay (Thermo Scientific) and equalized across samples with the addition of lysis buffer.

Chemoproteomic enrichment was performed as described in Glenn et al.[98] Lysates were brought to 500 μL by the addition of 1% SDS in PBS and then treated with 100

µL of 600 mM chloroacetamide in 0.8% SDS in PBS at 65°C for 30 min in the dark at 1200 RPM on a thermoshaker. For each replicate, 80 µL of dibenzocyclooctyne-agarose (DBCO-agarose) beads (Click Chemistry Tools) was washed thrice in 0.8% SDS in PBS and collected at 1500 RCF for 1 min. Lysates were then mixed with 600 µL of a solution of 8M urea and 0.85 M NaCl in PBS. Beads resuspended in 80 µL 0.8% SDS in PBS were then added to the lysates and incubated at room temperature on a rotary wheel in the dark for 24 h.

Beads were then collected at 1500 RCF for 1 min, washed with 1 mL deionized water, and collected again at 1500 RCF for 1 min. Supernatants were then discarded, and beads were resuspended in 500 µL of 5 mM dithiothreitol (DTT) (Millipore Sigma) and incubated for 15 min at 70°C shaking at 1200 RPM on a thermoshaker in the dark. Supernatants were then removed after collecting beads at 1500 RCF for 1 min, and beads were resuspended in 500 µL of 40 mM chloroacetamide (Millipore Sigma) and incubated for 30 min in the dark at room temperature on a rotary wheel.

For washing, beads were transferred to a PolyPrep gravity column (Bio-Rad) and washed ten times with 5 mL of 0.8% SDS in PBS (wash solution A), ten times with 8M urea in 100 mM Tris-Base (pH 8) (wash solution B), and ten times with 20% acetonitrile in deionized water (wash solution C). Columns were capped before addition of the second washes, and beads were incubated in the wash solution for 10 min, 30 min, and 10 min for solutions A, B, and C, respectively.

Beads were resuspended in 1.5 mL of 50 mM ammonium bicarbonate (Millipore Sigma) in a 10% acetonitrile and 90% deionized water solution, collected at 1.5k RCF for 1 min, and concentrated to 100 µL. Lys-C was added to 0.5 ng/µL for 4 h of digestion at 37 °C in the dark on a thermoshaker set to 1200 RPM. Trypsin was then added to 1 ng/µL for digestion overnight.

The next day, samples were mixed with 100 µL of 20% acetonitrile in deionized water, and beads were collected at 1500 RCF for 1 min. Supernatants were then transferred to centrifugation columns (Pierce). This process was repeated twice with an increased volume of 150 µL of 20% acetonitrile in deionized water. Peptide solutions were vacuum concentrated (SpeedVac) to dryness and resuspended in 20 µL of 0.2% formic acid in mass spectrometry–grade water. Samples were then desalted with C18 ZipTips (EMD Millipore) per manufacturers instructions and submitted for TMT labeling and LC-MS/MS analysis.

LC-MS/MS analysis

Enriched peptide samples were labelled with TMTpro (Thermo Fisher Scientific) according to the manufacturer's protocol, quenched with 5% hydroxylamine for 15min, pooled, and lyophilized to dryness. Samples were resuspended in 5% acetonitrile/0.5% TFA, desalted on Pierce C18 spin columns (Thermo Scientific #89870), and the eluted peptides dried by lyophilization. Peptides were then resuspended in 2% acetonitrile/0.2% formic acid for LC-MS analysis.

Liquid chromatography-mass spectrometry (LC-MS) analysis was carried out on an EASY-nLC 1200 (Thermo Fisher Scientific) coupled to an Orbitrap Eclipse mass spectrometer (Thermo Fisher Scientific) equipped with a Nanospray Flex ion source. 1 µg peptides were directly loaded onto an Aurora 25cm x 75µm ID, 1.6µm C18 column (Ion Opticks, Victoria, Australia) heated to 50°C. The peptides were separated with a 180 min gradient at a flow rate of 350 nL/min as follows: 2–6% Solvent B (11 min), 6-25% B (124 min), 25-40% B (45 min), 40-98% B (1 min), and held at 98% B (15 min). 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.

For FAIMS analysis, FAIMS was used alternating between 3 cycles at -35V, -50V, and -65V with cycle times of 1.2, 1.0, and 0.8 seconds respectively. MS1 spectra were acquired in the Orbitrap at 120K resolution with a scan range from 350-1600 m/z, an AGC target of 1e6, and a maximum injection time of 50 ms in Profile mode. Features were filtered for monoisotopic peaks with a charge state of 2-7 and minimum intensity of 2.5e4, with dynamic exclusion set to exclude features after 1 time for 45 seconds with a 5-ppm mass tolerance. HCD fragmentation was performed with fixed collision energy of 32% after quadrupole isolation of features using an isolation window of 0.5 m/z, an AGC target of 5e4, and maximum injection time of 86 ms. MS2 scans were then acquired in the Orbitrap at 50k resolution in Centroid mode with first mass fixed at 110.

For RTS-SPS-MS3 analysis, MS1 spectra were acquired in the Orbitrap at 120K resolution with a scan range from 350-2000 m/z, an AGC target of 1e6, and a maximum injection time set to Auto in Profile mode. Features were filtered for monoisotopic peaks with a charge state of 2-7 and a minimum intensity of 1e4, with dynamic exclusion set to exclude features after 1 time for 45 seconds with a 5-ppm mass tolerance. CID fragmentation was performed with collision energy of 35%, activation time of 10ms, and activation Q of 0.25 after quadrupole isolation

of features using an isolation window of 0.7 m/z, an AGC target of 1e4, and a maximum injection time of 45 ms. MS2 scans were then acquired in the ion trap at rapid rate in Centroid mode with normal mass range and Auto scan range mode. RTS using the *E. coli* K-12 MG1655 UniProt reference proteome was performed with carbamidomethylation of cysteines and TMTpro modification of lysines and peptide N-termini set as static modifications while methionine oxidation and deamidation of asparagine and glutamine were set as dynamic modifications. The maximum number of missed cleavages was set to 1, maximum variable mods was set two 2, and FDR filtering was not enabled. TMT SPS MS3 mode was used, and maximum search time was set to 35ms. Top-20 SPS matches were used with a precursor selection range of 300-2000m/z, precursor ion exclusion of -50 to +5 m/z, and isobaric tag loss exclusion reagent set to TMTpro. MS3 scans were then performed in the Orbitrap at 50k resolution in Centroid mode with a scan range of 100-500 m/z, using an MS isolation window of 1.2 m/z and an MS2 isolation window of 2 m/z, followed by HCD fragmentation with a fixed collision energy of 55%. The MS3 AGC target was set at 2.5e5 and maximum injection mode set to Auto.

Bioinformatic analysis

Analysis of LC-MS data was performed in Proteome Discoverer 2.4 (Thermo Scientific). Raw files were searched using the SequestHT search algorithm and the UniProt *E. coli* K-12 MG1655 reference proteome (UP000000625). Percolator FDR thresholds were 0.05 (relaxed) and 0.01 (strict). Trypsin was set as the cleavage enzyme, and the missed cleavage tolerance was set to 2. Dynamic modifications were set as oxidation of methionine and N-terminal acetylation. Static modifications were set as N-terminal TMTpro modification, TMTpro modification of lysine, and carbamidomethylation of cysteine. The minimum peptide length was set to 6, and peptide FDRs were set to 0.01 (relaxed) and 0.001 (strict). The co-isolation threshold for reporter ion quantification was set to 50 with a minimum average signal to noise threshold of 10. For SPS MS3 quantification, the SPS mass match threshold was set to 65%.

All normalization and differential expression analysis was conducted with the R package *limma*.[99] Protein abundances were log2-transformed and normalized with the cyclic Loess method. Normalized abundances were fit to a linear model for empirical Bayes–moderated *t*-testing of each coefficient with Benjamini-Hochberg adjustment of *p*-values for FDR control. Competitive functional enrichment analysis of ranked protein lists was performed on the STRING database server.[96]

Machine learning

Protein features were sourced from the UniProt, EcoCyc, and AlphaFold databases. All proteins quantified across the exponential and stationary phase proteomics datasets were included as training examples, with proteins exhibiting statistically significant degradation rates (FDR-adjusted p-value < 0.05) in either dataset labeled as positive examples. Hyperparameter tuning and model evaluation were performed with 10 distinct rounds of nested cross validation with 10 inner and outer folds, each stratified to maintain class imbalance. Parameter values for hyperparameter selection are provided in Table 2.1. Average precision was used as the performance metric for model selection. Training for the L1-regularized logistic regression and random forest model classes was conducted with the scikit-learn Python module with balanced class weights to address class imbalance.[100] Neural networks were constructed with PyTorch and evaluated with the scikit-learn API provided by the skorch Python module.[101] As two or more hidden layers provided no performance improvement at the cost of computation time, only networks with one hidden layer were considered. The BCEWithLogitsLoss loss function was used with positive weighting given by the class ratio to address class imbalance, and the loss function was optimized with the Adadelta optimizer using a batch size of 128.[102] Max epochs were set to 20 with early stopping set to a patience window of 3 monitoring the validation loss function. XGBoost was operated in linear mode with positive instances weighted according to the class ratio to address class imbalance, and the model was evaluated with the scikit-learn API of the XGBoost Python module. Performance metrics were assessed with the scikit-learn model selection routines. ROC and PR curves were averaged across all evaluated test sets, and PR curves were interpolated. PubMed reference counts for each protein were acquired from the Full Report entries on NCBI Gene with the Entrez Direct utilities. SHAP values for feature importance were calculated as previously described for linear models.[103]

2.7 Supplemental Information

Supplementary Figures

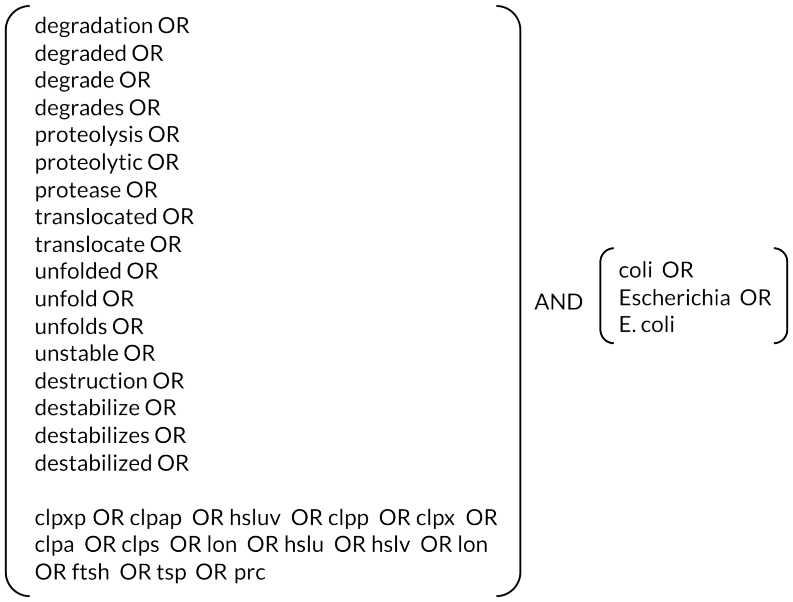

Figure S2.1: **Query submitted to PubMed via Entrez Direct.** Illustration of the Boolean query submitted to NCBI PubMed via the command line Entrez Direct utilities. The query yielded 25551 PubMed IDs that were then submitted to the PubMed2XL utility to retrieve article titles, abstracts, and metadata. This same query with the species specification removed was used to reduce the EcoCyc corpus of protein summaries.

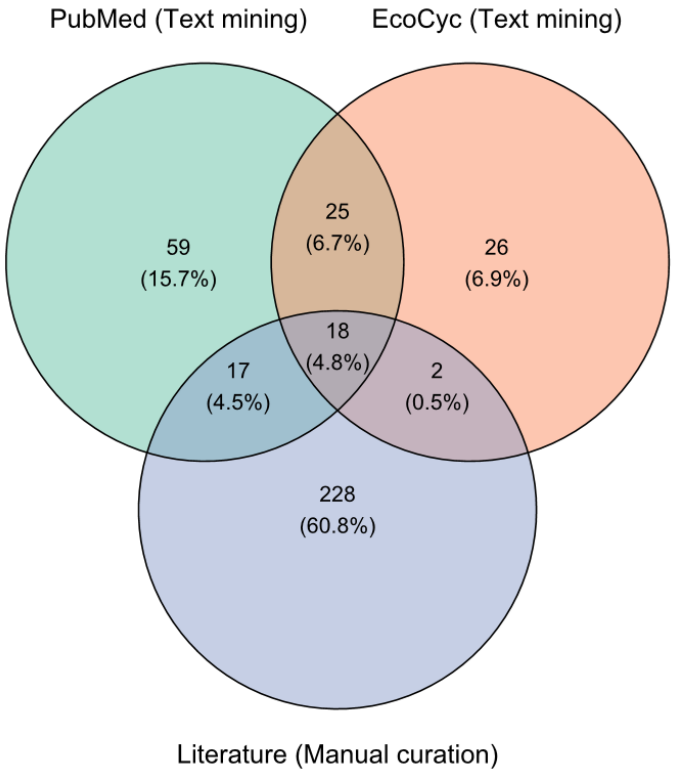

Figure S2.2: **Comparison of annotation mappings across knowledge bases.** Venn diagram depicting positive annotations retrieved from the PubMed or EcoCyc databases via text mining or from manual curation of the research literature. Text mining grew the set of protease substrate annotations by 41%, with the mined knowledge bases providing distinct identifications.

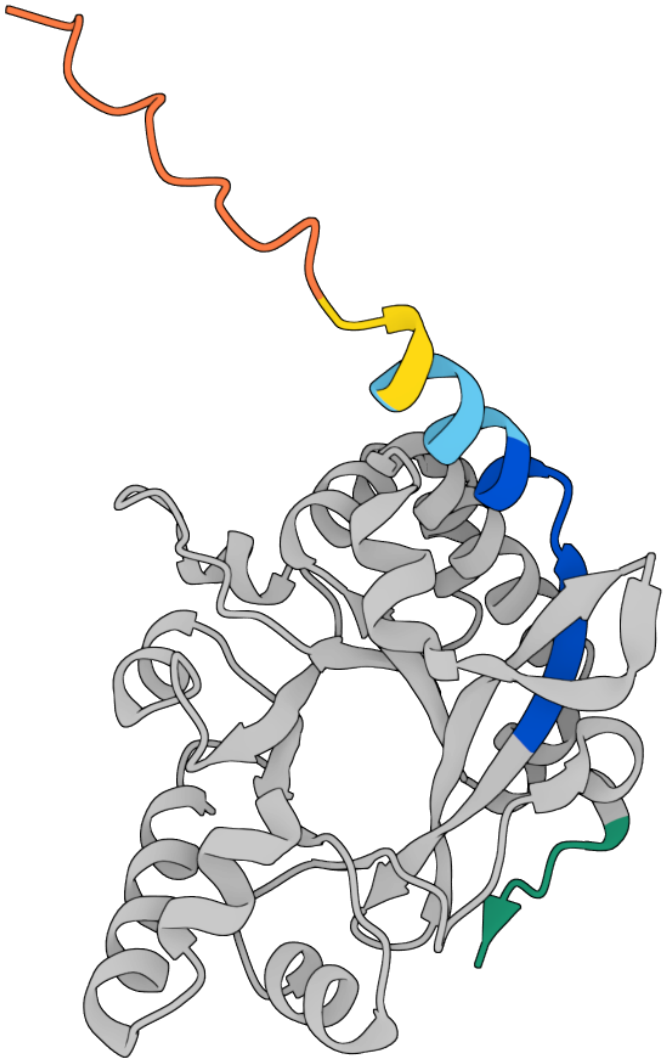

Figure S2.3: **AlphaFold prediction of PdeH structure.** The conserved c-di-GMP EAL domain is shaded in grey, and the N-terminal extension is colored by the prediction's per-residue model confidence score (pLDDT). The yellow and orange regions indicate low and very low model confidence scores, respectively.

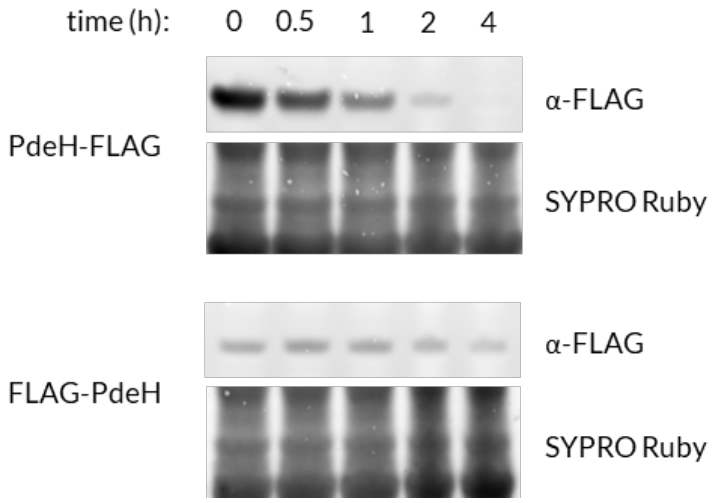

Figure S2.4: **Immunoblotting stability analysis of N- and C-terminally FLAG-tagged PdeH.** While the C-terminal FLAG tag did not interfere with PdeH instability, blocking of the N terminus with the FLAG sequence stabilized PdeH. Expression of the N-terminal fusion may be lower due to altered translation initiation efficiency from sequence perturbation near the Shine-Dalgarno sequence.

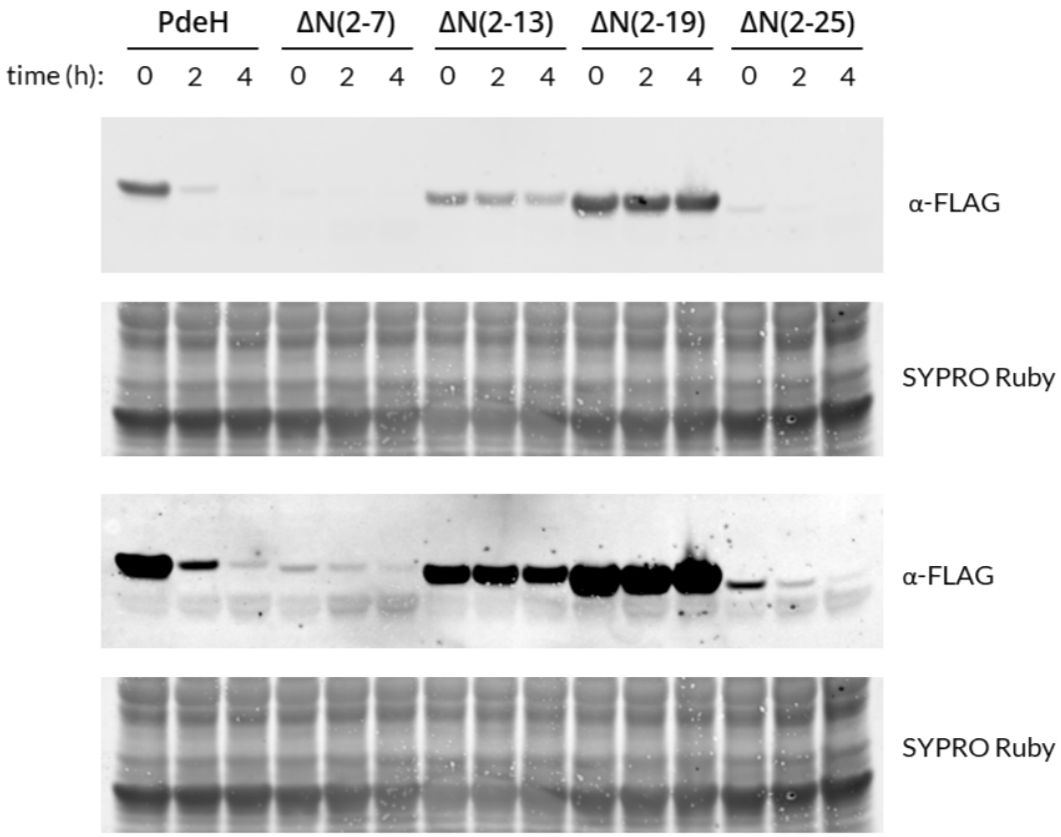

Figure S2.5: **Immunoblotting stability analysis of N-terminal truncations for PdeH-FLAG fusion.** Successive deletions of six-residue stretches in the PdeH N terminus yield a stabilized PdeH-FLAG mutant with the removal of 12 or 18 amino acids. The bottom panel is the same blot with narrowed contrast for visualizing bands with reduced signal.

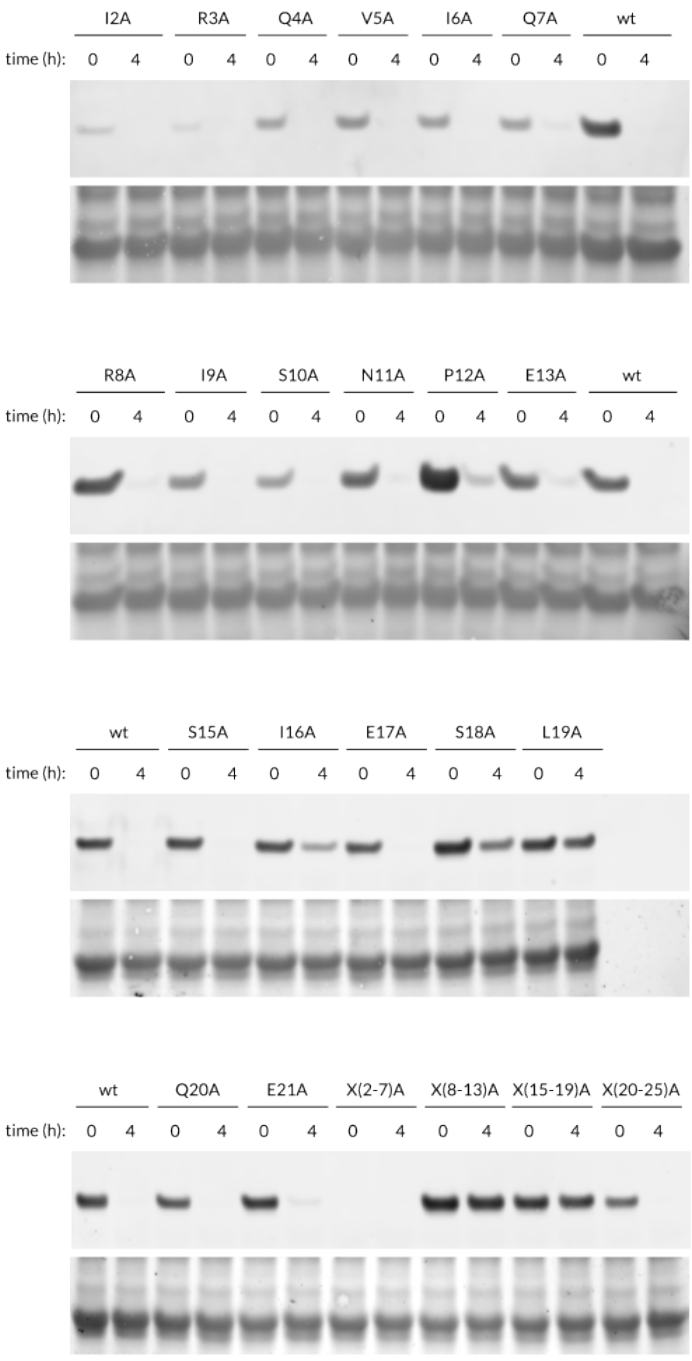

Figure S2.6: **Immunoblotting stability analysis of N-terminal alanine scanning and stretch mutants of PdeH-FLAG.** Alanine point or stretch mutation across the N-terminal extension of PdeH yields the stable L19A, X(8-13)A, and X(15-19)A PdeH-FLAG mutants.

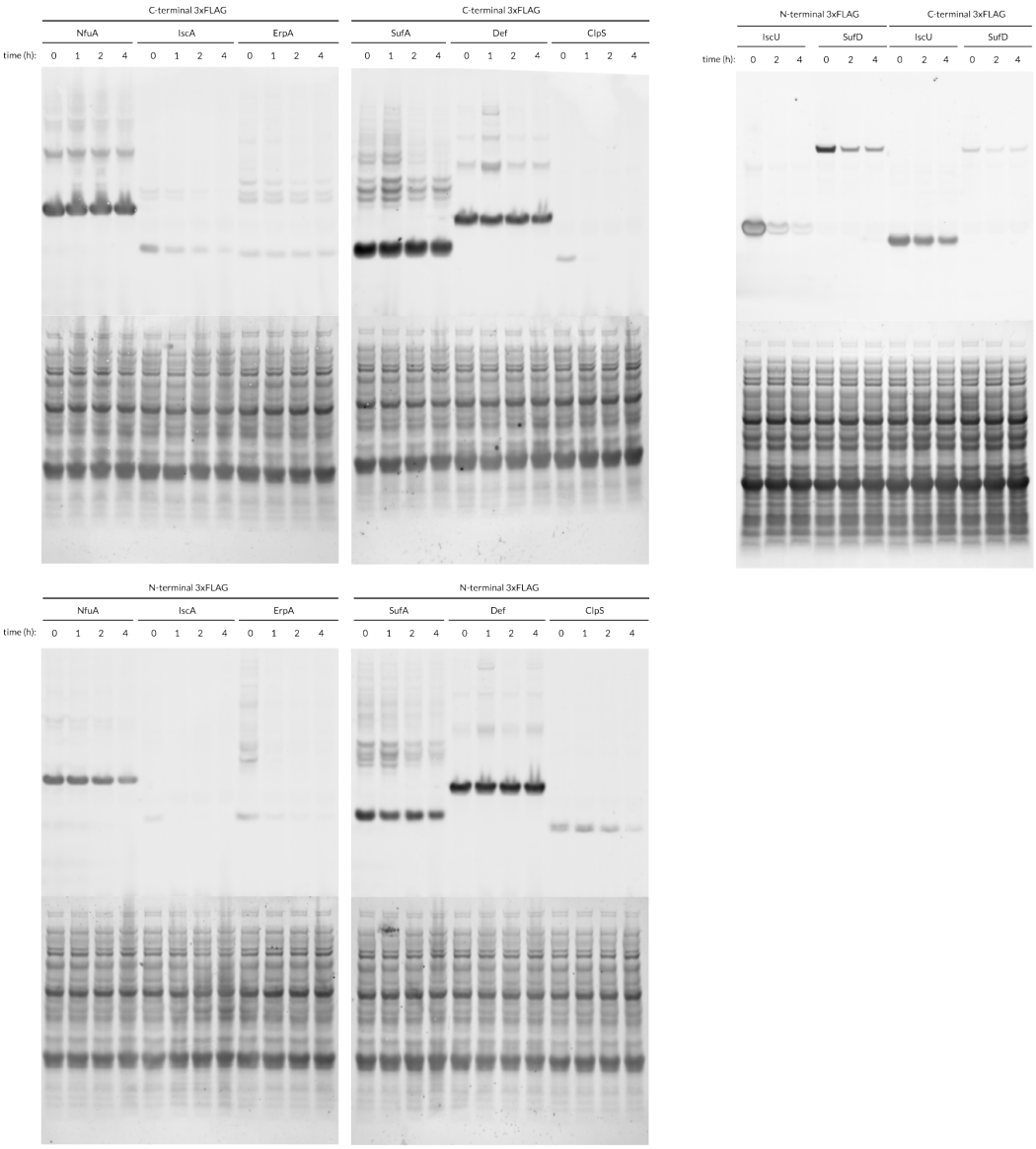

Figure S2.7: **Immunoblotting stability analysis of N- and C-terminally tagged candidate exponential phase substrates.** N-terminal 3xFLAG fusions recapitulate the proteolysis and rank order of expression for the A-type carriers IscA, ErpA, NfuA, and SufA and scaffold components SufD and IscU. The candidate substrate Def did not validate in this immunoblotting stability analysis. The ClpS-3xFLAG fusion displays the expected degradation profile, whereas perturbation of the N-terminus of ClpS, which mediates ClpS-ClpAP binding, stabilized the protein.

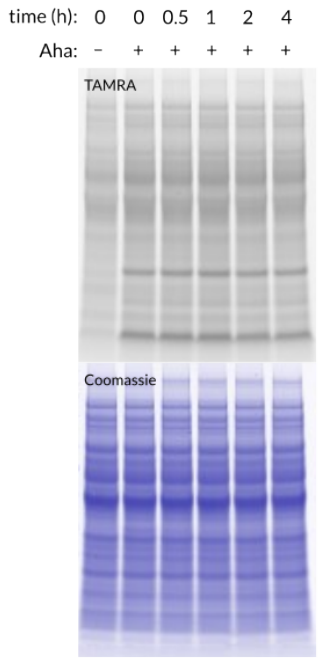

Figure S2.8: **Aha labeling is reduced in stationary phase cells.** Cells from an overnight culture in M9 glycerol were diluted 1:200 in fresh medium, grown 24 h, and treated with 1 mM Aha. Aliquots were collected at the indicated time points, and cells were collected and lysed. Lysates were used in click conjugation to TAMRA-alkyne, and proteins were separated by mass via SDS-PAGE for in-gel fluorescence detection.

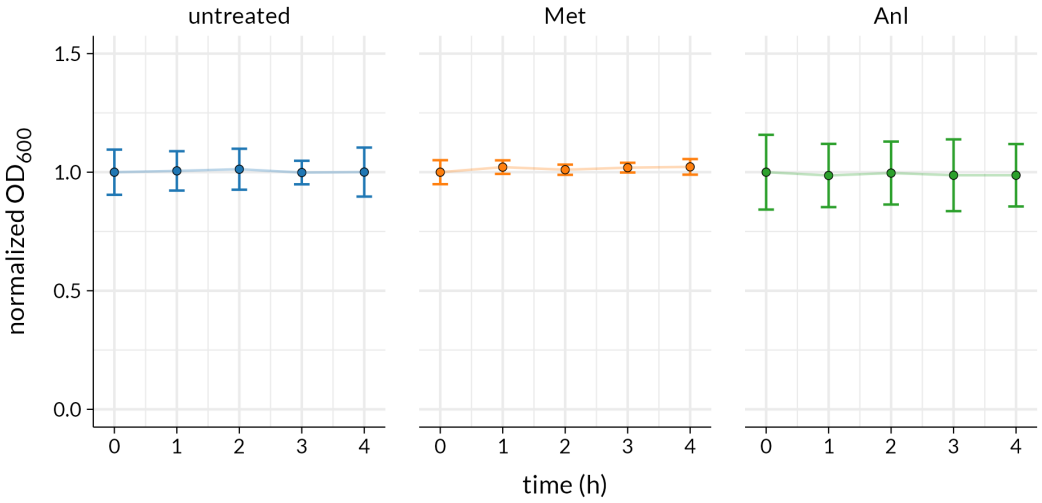

Figure S2.9: **Met and Anl do not elicit growth resumption in stationary phase cells.** Cells from an overnight culture in M9 glycerol were diluted 1:200 in fresh medium, grown 24 h, and treated with 1 mM Met or 1 mM Anl. Cells were removed from cultures at the indicated timepoints for optical density measurements. Measurements were made in triplicate and normalized to the mean optical density at the initial timepoint for each series. Error bars indicate 95% confidence intervals.

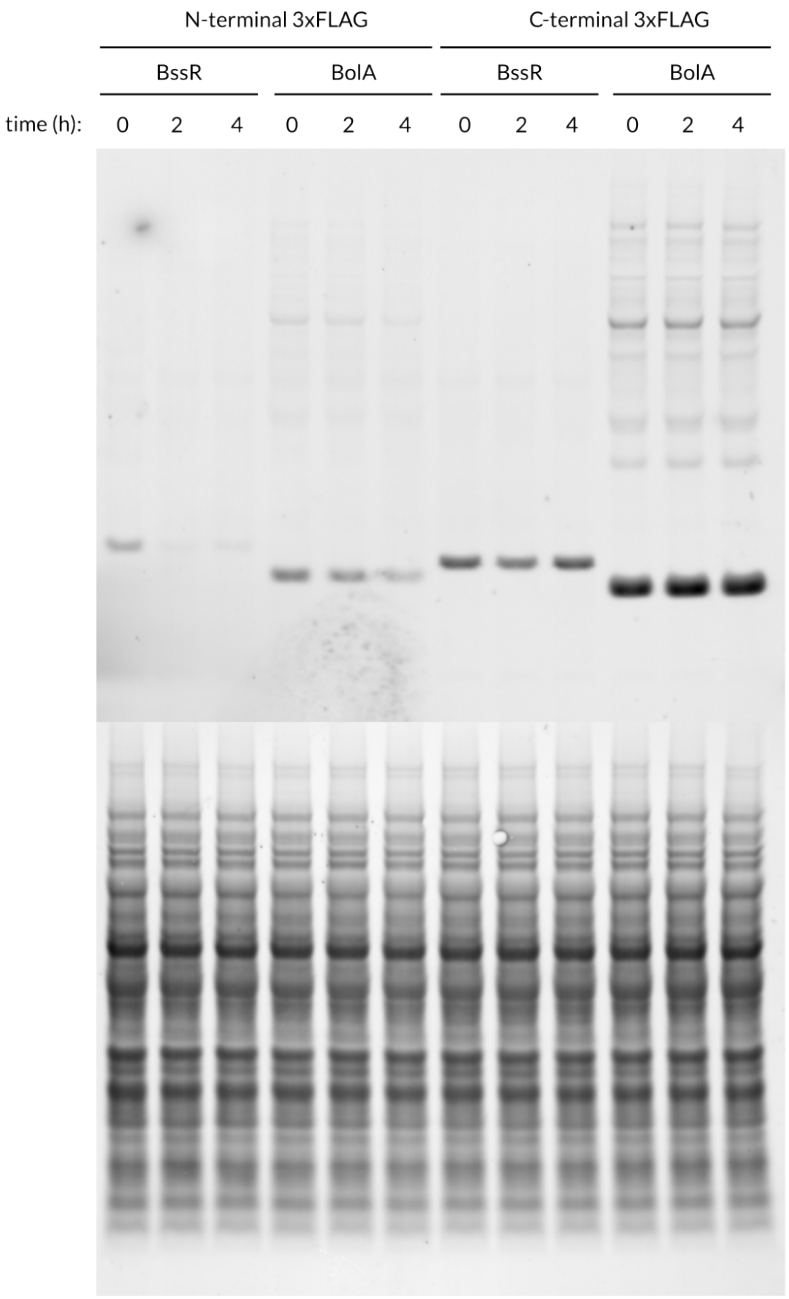

Figure S2.10: **Immunoblotting stability analysis for candidate stationary phase substrates.** N-terminal 3xFLAG fusions of BssR and BolA exhibit instability.

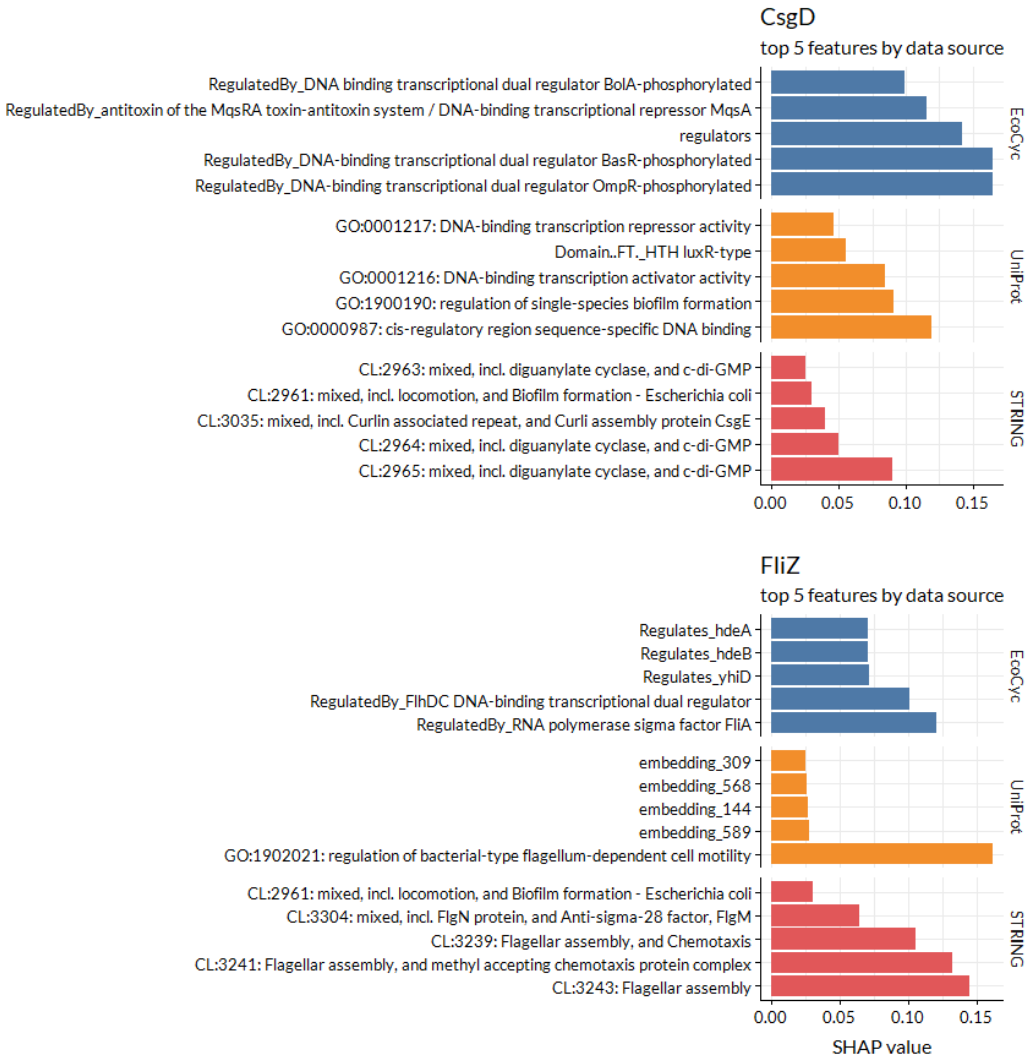

Figure S2.11: **Top predictive features across data sources for CsgD and FliZ.** The top 5 features across annotation databases for predicted substrates CsgD and FliZ based on SHAP value.[103]

Table 2.1: Parameters used in grid searching during nested cross validation

Model	Parameter	Values
Logistic regression	C	1e3, 1e4, 1e5
Random forest	max_depth	5, 10, 15, None
	min_samples_leaf	50, 100
	min_samples_split	2, 5, 10
	n_estimators	50, 100
Artificial neural network	Weight decay	1e-2, 1e-3
	Dropout	0.2, 0.3, 0.4, 0.5
	Hidden units	20, 50, 100, 200
XGBoost	learning_rate	1e-1, 1e-2
	reg_lambda	1e-2, 1e-3, 1e-4
	reg_alpha	1e-3, 5e-4, 1e-4
	n_estimators	25, 10

Strains

Table 2.2: Strains used in this study

Strain	Description	Reference
MG1655	*Escherichia coli* K-12 derivative	[104]
MG1655 Δ*clpP::kan*^{R}	ClpP knockout	This work
MG1655 Δ*clpX::kan*^{R}	ClpX knockout	This work
MG1655 Δ*clpA::kan*^{R}	ClpA knockout	This work
MG1655 Δ*lon::kan*^{R}	Lon knockout	This work
MG1655 Δ*hslV::kan*^{R}	HslV knockout	This work
MG1655 *pdeH-3xFLAG*	C-terminal 3xFLAG insertion at *pdeH* locus	This work
MG1655 *pdeH(L19A)-3xFLAG*	N-terminal L19A mutation and C-terminal 3xFLAG insertion	This work
MG1655 *pdeH(X(15-19)A)-3xFLAG*	N-terminal alanine stretch spanning residues 15-19 and C-terminal 3xFLAG insertion	This work

Plasmids

Table 2.3: Plasmids used in this study

Plasmid	Description	Reference
pKD46	Recombinase expression for chromosomal deletion	[90]
pKD13	Template plasmid for Kan cassette insertion	[90]
pBAD33	Empty vector for arabinose-inducible expression	[40]
pBAD33-pdeH-FLAG	Arabinose-inducible expression of PdeH-FLAG	This work
pBAD33-FLAG-pdeH	Arabinose-inducible expression of FLAG-pdeH	This work

Table 2.3 continued from previous page

Plasmid	Description	Reference
pBAD33-ΔN(2-7)-pdeH-FLAG	Arabinose-inducible expression of N-terminal truncation of PdeH-FLAG	This work
pBAD33-ΔN(2-13)-pdeH-FLAG	Arabinose-inducible expression of N-terminal truncation of PdeH-FLAG	This work
pBAD33-ΔN(2-19)-pdeH-FLAG	Arabinose-inducible expression of N-terminal truncation of PdeH-FLAG	This work
pBAD33-ΔN(2-25)-pdeH-FLAG	Arabinose-inducible expression of N-terminal truncation of PdeH-FLAG	This work
pBAD33-pdeH(I2A)-FLAG	Arabinose-inducible expression of alanine point mutant of PdeH-FLAG	This work
pBAD33-pdeH(R3A)-FLAG	Arabinose-inducible expression of alanine point mutant of PdeH-FLAG	This work
pBAD33-pdeH(Q4A)-FLAG	Arabinose-inducible expression of alanine point mutant of PdeH-FLAG	This work
pBAD33-pdeH(V5A)-FLAG	Arabinose-inducible expression of alanine point mutant of PdeH-FLAG	This work
pBAD33-pdeH(I6A)-FLAG	Arabinose-inducible expression of alanine point mutant of PdeH-FLAG	This work
pBAD33-pdeH(Q7A)-FLAG	Arabinose-inducible expression of alanine point mutant of PdeH-FLAG	This work
pBAD33-pdeH(R8A)-FLAG	Arabinose-inducible expression of alanine point mutant of PdeH-FLAG	This work
pBAD33-pdeH(I9A)-FLAG	Arabinose-inducible expression of alanine point mutant of PdeH-FLAG	This work

Table 2.3 continued from previous page

Plasmid	Description	Reference
pBAD33-pdeH(S10A)-FLAG	Arabinose-inducible expression of alanine point mutant of PdeH-FLAG	This work
pBAD33-pdeH(N11A)-FLAG	Arabinose-inducible expression of alanine point mutant of PdeH-FLAG	This work
pBAD33-pdeH(P12A)-FLAG	Arabinose-inducible expression of alanine point mutant of PdeH-FLAG	This work
pBAD33-pdeH(E13A)-FLAG	Arabinose-inducible expression of alanine point mutant of PdeH-FLAG	This work
pBAD33-pdeH(S15A)-FLAG	Arabinose-inducible expression of alanine point mutant of PdeH-FLAG	This work
pBAD33-pdeH(I16A)-FLAG	Arabinose-inducible expression of alanine point mutant of PdeH-FLAG	This work
pBAD33-pdeH(E17A)-FLAG	Arabinose-inducible expression of alanine point mutant of PdeH-FLAG	This work
pBAD33-pdeH(S18A)-FLAG	Arabinose-inducible expression of alanine point mutant of PdeH-FLAG	This work
pBAD33-pdeH(L19A)-FLAG	Arabinose-inducible expression of alanine point mutant of PdeH-FLAG	This work
pBAD33-pdeH(Q20A)-FLAG	Arabinose-inducible expression of alanine point mutant of PdeH-FLAG	This work
pBAD33-pdeH(E21A)-FLAG	Arabinose-inducible expression of alanine point mutant of PdeH-FLAG	This work
pBAD33-pdeH-X(2-7)A-FLAG	Arabinose-inducible expression of alanine point mutant of PdeH-FLAG	This work

Table 2.3 continued from previous page

Plasmid	Description	Reference
pBAD33-pdeH-X(8-13)A-FLAG	Arabinose-inducible expression of alanine point mutant of PdeH-FLAG	This work
pBAD33-pdeH-X(15-19)A-FLAG	Arabinose-inducible expression of alanine point mutant of PdeH-FLAG	This work
pBAD33-pdeH-X(20-25)A-FLAG	Arabinose-inducible expression of alanine point mutant of PdeH-FLAG	This work
pTarget	sgRNA expression vector	[93]
pCas	Cas9 expression vector	[93]
pTarget-pdeH(N20)	sgRNA expression targeting Cas9 to pdeH locus	This work
pTarget-pdeH-3xFLAG	sgRNA expression targeting Cas9 to pdeH locus with pdeH-3xFLAG donor DNA	This work
pTarget-pdeH(L19A)-3xFLAG	sgRNA expression targeting Cas9 to pdeH locus with pdeH(L19A)-3xFLAG donor DNA	This work
pBAD33-clpS-3xFLAG	Arabinose-inducible expression of clpS-3xFLAG	This work
pBAD33-ΔNTE-clpS-3xFLAG	Arabinose-inducible expression of ΔNTE-clpS-3xFLAG	This work
pBAD33-3xFLAG-iscA	Arabinose-inducible expression of alanine point mutant of 3xFLAG-iscA	This work
pBAD33-3xFLAG-erpA	Arabinose-inducible expression of alanine point mutant of 3xFLAG-erpA	This work
pBAD33-3xFLAG-sufA	Arabinose-inducible expression of alanine point mutant of 3xFLAG-sufA	This work
pBAD33-3xFLAG-nfuA	Arabinose-inducible expression of alanine point mutant of 3xFLAG-nfuA	This work

Table 2.3 continued from previous page

Plasmid	Description	Reference
pBAD33-iscA-3xFLAG	Arabinose-inducible expression of alanine point mutant of iscA-3xFLAG	This work
pBAD33-erpA-3xFLAG	Arabinose-inducible expression of alanine point mutant of erpA-3xFLAG	This work
pBAD33-sufA-3xFLAG	Arabinose-inducible expression of alanine point mutant of sufA-3xFLAG	This work
pBAD33-nfuA-3xFLAG	Arabinose-inducible expression of alanine point mutant of nfuA-3xFLAG	This work
pBAD33-3xFLAG-gadX	Arabinose-inducible expression of alanine point mutant of 3xFLAG-gadX	This work
pBAD33-3xFLAG-nsrR	Arabinose-inducible expression of alanine point mutant of 3xFLAG-nsrR	This work
pBAD33-3xFLAG-fliZ	Arabinose-inducible expression of alanine point mutant of 3xFLAG-fliZ	This work
pBAD33-3xFLAG-csgD	Arabinose-inducible expression of alanine point mutant of 3xFLAG-csgD	This work
pBAD33-3xFLAG-fecI	Arabinose-inducible expression of alanine point mutant of 3xFLAG-fecI	This work
pBAD33-3xFLAG-iraD	Arabinose-inducible expression of alanine point mutant of 3xFLAG-iraD	This work
pBAD33-3xFLAG-yeiL	Arabinose-inducible expression of alanine point mutant of 3xFLAG-yeiL	This work
pBAD33-3xFLAG-fliS	Arabinose-inducible expression of alanine point mutant of 3xFLAG-fliS	This work

Table 2.3 continued from previous page

Plasmid	Description	Reference
pBAD33-gadX-3xFLAG	Arabinose-inducible expression of alanine point mutant of gadX-3xFLAG	This work
pBAD33-nsrR-3xFLAG	Arabinose-inducible expression of alanine point mutant of nsrR-3xFLAG	This work
pBAD33-fliZ-3xFLAG	Arabinose-inducible expression of alanine point mutant of fliZ-3xFLAG	This work
pBAD33-csgD-3xFLAG	Arabinose-inducible expression of alanine point mutant of csgD-3xFLAG	This work
pBAD33-fecI-3xFLAG	Arabinose-inducible expression of alanine point mutant of fecI-3xFLAG	This work
pBAD33-iraD-3xFLAG	Arabinose-inducible expression of alanine point mutant of iraD-3xFLAG	This work
pBAD33-yeiL-3xFLAG	Arabinose-inducible expression of alanine point mutant of yeiL-3xFLAG	This work
pBAD33-fliS-3xFLAG	Arabinose-inducible expression of alanine point mutant of fliS-3xFLAG	This work
pRrnB-GFP-NLL-MetRS	Stringently controlled expression of NLL-MetRS fused to GFP	This work
pBbA2c-RFP	Anhydrotetracycline-inducible expression of RFP	[92]
pBbA2c-rpoZ-3xFLAG	Anhydrotetracycline-inducible expression of rpoZ-3xFLAG	This work
pBbA2c-bolA-3xFLAG	Anhydrotetracycline-inducible expression of bolA-3xFLAG	This work
pBbA2c-gadE-3xFLAG	Anhydrotetracycline-inducible expression of gadE-3xFLAG	This work
pBbA2c-bssR-3xFLAG	Anhydrotetracycline-inducible expression of bssR-3xFLAG	This work

Table 2.3 continued from previous page

Plasmid	Description	Reference
pBbA2c-narP-3xFLAG	Anhydrotetracycline-inducible expression of narP-3xFLAG	This work
pBbA2c-3xFLAG-rpoZ	Anhydrotetracycline-inducible expression of 3xFLAG-rpoZ	This work
pBbA2c-3xFLAG-bolA	Anhydrotetracycline-inducible expression of 3xFLAG-bolA	This work
pBbA2c-3xFLAG-gadE	Anhydrotetracycline-inducible expression of 3xFLAG-gadE	This work
pBbA2c-3xFLAG-bssR	Anhydrotetracycline-inducible expression of 3xFLAG-bssR	This work
pBbA2c-3xFLAG-narP	Anhydrotetracycline-inducible expression of 3xFLAG-narP	This work

Primers

Table 2.4: Primers used in this study

Primer	Sequence (5' to 3')	Purpose
pKD13-lon-fwd	TTTTTATTAGTGCATTTTGC GCGAGGTCACTATTTTGCAG TCACAACCTGTGTAGGCTGG AGCTGCTTCG	Amplification of pKD13 resistance cassette for Lon deletion
pKD13-lon-rev	TCGTGTCATCTGATTACCTG GCGGAAATTAAACTAAGAGA GAGCTCTATGATTCCGGGGA TCCGTCGACC	Amplification of pKD13 resistance cassette for Lon deletion
pKD13-clpP-fwd	AGTATAGCGGCACAGTTGCG CCTCTGGCATCAATTACGAT GGGTCAGAATTGTAGGCTGG AGCTGCTTCG	Amplification of pKD13 resistance cassette for ClpP deletion
pKD13-clpP-rev	TTACAATCGGTACAGCAGGT TTTTTCAATTTTATCCAGGA GACGGAAATGATTCCGGGGA TCCGTCGACC	Amplification of pKD13 resistance cassette for ClpP deletion
pKD13-hslV-fwd	CGTGGGGTCATTTCAGACAT GGGAGATCCTTACGCTTTGT AGCTTAATTCTGTAGGCTGG AGCTGCTTCG	Amplification of pKD13 resistance cassette for HslV deletion
pKD13-hslV-rev	TTTTGTACGGGGTTTGTACT CTGTATTCGTAACCAAGGGG TCAGCTCGTGATTCCGGGGA TCCGTCGACC	Amplification of pKD13 resistance cassette for HslV deletion
pKD13-clpA-fwd	GGACTTGACCAACCTACCTA ACAATCAGATTAATGCGCTG CTTCCGCCTTTGTAGGCTGG AGCTGCTTCG	Amplification of pKD13 resistance cassette for ClpA deletion

Table 2.4 continued from previous page

Primer	Sequence (5' to 3')	Purpose
pKD13-clpA-rev	CGCTAGAAAAAGCCTGAATG CAGGCATAAAAATTGGGGGA GGTGCCTATGATTCCGGGGA TCCGTCGACC	Amplification of pKD13 resistance cassette for ClpA deletion
pKD13-clpX-fwd	TGGTTAACTAATTGTATGGG AATGGTTAATTATTCACCAG ATGCCTGTTGTGTAGGCTGG AGCTGCTTCG	Amplification of pKD13 resistance cassette for ClpX deletion
pKD13-clpX-rev	TTGCGTCGTCGTGTGCGGCA CAAAGAACAAAGAAGAGGTT TTGACCCATGATTCCGGGGA TCCGTCGACC	Amplification of pKD13 resistance cassette for ClpX deletion
clpX-seq-fwd	TTCGATTCTGACCCATCGTA ATTGATG	Deletion screening for ClpX
clpX-seq-rev	GATGACACGACTGTGCTTCA CG	Deletion screening for ClpX
clpP-seq-fwd	AAGCCTCTTTCGGTGTTAGC G	Deletion screening for ClpP
clpP-seq-rev	GCCATCTTTGCGTTTATCTG TCATGG	Deletion screening for ClpP
lon-seq-fwd	CTCTATTCTCGGCGTTGAAT GTGG	Deletion screening for Lon
lon-seq-rev	GCAGTTATATCAGGCCAGCC ATC	Deletion screening for Lon
hslV-seq-fwd	CAATCGGTTGAAGATGGCGG G	Deletion screening for HslV
hslV-seq-rev	CATGGCGCAGCTCTTCGTTG	Deletion screening for HslV
clpA-seq-fwd	GCGGAGTCTTTACCGCCGAG	Deletion screening for ClpA
clpA-seq-rev	TCCTTACCTTCCCGCAGCAC	Deletion screening for ClpA

Table 2.4 continued from previous page

Primer	Sequence (5' to 3')	Purpose
pdeH-B-FLAG-C-fwd	atacggcggttctggcgcta GACTACAAGGACGACGACGAC	Backbone amplification of pBAD33-FLAG for PdeH-FLAG expression vector
pdeH-B-FLAG-C-rev	cgctggataacctgccttat CATcgtttcactcctggcct tc	Backbone amplification of pBAD33-FLAG for PdeH-FLAG expression vector
pdeH-FLAG-C-fwd	aggccaggagtgaaacgATG ataaggcaggttatccagcg aataag	Chromosomal amplification of PdeH for PdeH-FLAG insert
pdeH-FLAG-C-rev	TCGTCGTCGTCCTTGTAGTC tagcgccagaaccgcc	Chromosomal amplification of PdeH for PdeH-FLAG insert
pdeH-B-FLAG-N-fwd	ATACGGCGGTTCTGGCGCTA TGAgtcgacctgcaggc	Backbone amplification of pBAD33-FLAG for FLAG-PdeH expression vector
pdeH-B-FLAG-N-rev	CGCTGGATAACCTGCCTTAT CTTGTCGTCGTCGTCCTTGT AG	Backbone amplification of pBAD33-FLAG for FLAG-PdeH expression vector
pdeH-FLAG-N-fwd	ACAAGGACGACGACGACAAG ataaggcaggttatccagcg aataag	Chromosomal amplification of PdeH for FLAG-PdeH insert
pdeH-FLAG-N-rev	catgcctgcaggtcgacTCA tagcgccagaaccgcc	Chromosomal amplification of PdeH for FLAG-PdeH insert
ΔN(1-6)-pdeH-fwd	/5Phos/cgaataagcaacc ctgaagcaag	Truncation of residues 2-6 in pBAD33-PdeH-FLAG
ΔN(1-12)-pdeH-fwd	/5Phos/gcaagcatcgaga gcttgc	Truncation of residues 2-12 in pBAD33-PdeH-FLAG
ΔN(1-18)-pdeH-fwd	/5Phos/caggaacggcgtt tttggttg	Truncation of residues 2-18 in pBAD33-PdeH-FLAG
ΔN(1-24)-pdeH-fwd	/5Phos/ttgcagtgtgagc gtgc	Truncation of residues 2-24 in pBAD33-PdeH-FLAG

Table 2.4 continued from previous page

Primer	Sequence (5' to 3')	Purpose
ΔN-pdeH-rev	/5Phos/CATcgtttcactc ctggcc	Backbone amplification for N-terminal truncation in pBAD33-PdeH-FLAG
pdeH-I2A-fwd	/5Phos/GCGaggcaggtta tccagcgaataag	Site-directed mutagenesis of pBAD33-pdeH-FLAG
pdeH-I2A-rev	/5Phos/CATcgtttcactc ctggcct	Site-directed mutagenesis of pBAD33-pdeH-FLAG
pdeH-R3A-fwd	/5Phos/GCGcaggttatcc agcgaataagcaac	Site-directed mutagenesis of pBAD33-pdeH-FLAG
pdeH-R3A-rev	/5Phos/tatCATcgtttca ctcctggcc	Site-directed mutagenesis of pBAD33-pdeH-FLAG
pdeH-Q4A-fwd	/5Phos/GCGgttatccagc gaataagcaaccc	Site-directed mutagenesis of pBAD33-pdeH-FLAG
pdeH-Q4A-rev	/5Phos/ccttatCATcgtt tcactcctgg	Site-directed mutagenesis of pBAD33-pdeH-FLAG
pdeH-V5A-fwd	/5Phos/GCGatccagcgaa taagcaaccctg	Site-directed mutagenesis of pBAD33-pdeH-FLAG
pdeH-V5A-rev	/5Phos/ctgccttatCATc gtttcactcc	Site-directed mutagenesis of pBAD33-pdeH-FLAG
pdeH-I6A-fwd	/5Phos/GCGcagcgaataa gcaaccctgaag	Site-directed mutagenesis of pBAD33-pdeH-FLAG
pdeH-I6A-rev	/5Phos/aacctgccttatC ATcgtttcac	Site-directed mutagenesis of pBAD33-pdeH-FLAG
pdeH-Q7A-fwd	/5Phos/GCGcgaataagca accctgaagcaag	Site-directed mutagenesis of pBAD33-pdeH-FLAG
pdeH-Q7A-rev	/5Phos/gataacctgcctt atCATcgtttcac	Site-directed mutagenesis of pBAD33-pdeH-FLAG
pdeH-R8A-fwd	/5Phos/GCGataagcaacc ctgaagcaagc	Site-directed mutagenesis of pBAD33-pdeH-FLAG

Table 2.4 continued from previous page

Primer	Sequence (5' to 3')	Purpose
pdeH-R8A-rev	/5Phos/ctggataacctgc cttatCATcgtttc	Site-directed mutagenesis of pBAD33-pdeH-FLAG
pdeH-I9A-fwd	/5Phos/GCGagcaaccctg aagcaagc	Site-directed mutagenesis of pBAD33-pdeH-FLAG
pdeH-I9A-rev	/5Phos/tcgctggataacc tgccttatC	Site-directed mutagenesis of pBAD33-pdeH-FLAG
pdeH-S10A-fwd	/5Phos/GCGaaccctgaag caagcatcgag	Site-directed mutagenesis of pBAD33-pdeH-FLAG
pdeH-S10A-rev	/5Phos/tattcgctggata acctgccttatC	Site-directed mutagenesis of pBAD33-pdeH-FLAG
pdeH-N11A-fwd	/5Phos/GCGcctgaagcaa gcatcgagagc	Site-directed mutagenesis of pBAD33-pdeH-FLAG
pdeH-N11A-rev	/5Phos/gcttattcgctgg ataacctgcc	Site-directed mutagenesis of pBAD33-pdeH-FLAG
pdeH-P12A-fwd	/5Phos/GCGgaagcaagca tcgagagcttg	Site-directed mutagenesis of pBAD33-pdeH-FLAG
pdeH-P12A-rev	/5Phos/gttgcttattcgc tggataacctg	Site-directed mutagenesis of pBAD33-pdeH-FLAG
pdeH-E13A-fwd	/5Phos/GCGgcaagcatcg agagcttgc	Site-directed mutagenesis of pBAD33-pdeH-FLAG
pdeH-E13A-rev	/5Phos/agggttgcttatt cgctggataac	Site-directed mutagenesis of pBAD33-pdeH-FLAG
pdeH-S15A-fwd	/5Phos/GCGatcgagagct tgcaggaacg	Site-directed mutagenesis of pBAD33-pdeH-FLAG
pdeH-S15A-rev	/5Phos/tgcttcagggttg cttattcgc	Site-directed mutagenesis of pBAD33-pdeH-FLAG
pdeH-I16A-fwd	/5Phos/GCGgagagcttgc aggaacggc	Site-directed mutagenesis of pBAD33-pdeH-FLAG

Table 2.4 continued from previous page

Primer	Sequence (5' to 3')	Purpose
pdeH-I16A-rev	/5Phos/gcttgcttcaggg ttgcttattcg	Site-directed mutagenesis of pBAD33-pdeH-FLAG
pdeH-E17A-fwd	/5Phos/GCGagcttgcagg aacggcg	Site-directed mutagenesis of pBAD33-pdeH-FLAG
pdeH-E17A-rev	/5Phos/gatgcttgcttca gggttgct	Site-directed mutagenesis of pBAD33-pdeH-FLAG
pdeH-S18A-fwd	/5Phos/GCGttgcaggaac ggcgtttttg	Site-directed mutagenesis of pBAD33-pdeH-FLAG
pdeH-S18A-rev	/5Phos/ctcgatgcttgct tcagggttg	Site-directed mutagenesis of pBAD33-pdeH-FLAG
pdeH-L19A-fwd	/5Phos/GCGcaggaacggc gtttttggttg	Site-directed mutagenesis of pBAD33-pdeH-FLAG
pdeH-L19A-rev	/5Phos/gctctcgatgctt gcttcagg	Site-directed mutagenesis of pBAD33-pdeH-FLAG
pdeH-Q20A-fwd	/5Phos/GCGgaacggcgtt tttggttgcag	Site-directed mutagenesis of pBAD33-pdeH-FLAG
pdeH-Q20A-rev	/5Phos/caagctctcgatg cttgcttcag	Site-directed mutagenesis of pBAD33-pdeH-FLAG
pdeH-E21A-fwd	/5Phos/GCGcggcgttttt ggttgcagtg	Site-directed mutagenesis of pBAD33-pdeH-FLAG
pdeH-E21A-rev	/5Phos/ctgcaagctctcg atgcttgc	Site-directed mutagenesis of pBAD33-pdeH-FLAG
pdeH-R22A-fwd	/5Phos/GCGcgtttttggt tgcagtgtgagc	Site-directed mutagenesis of pBAD33-pdeH-FLAG
pdeH-R22A-rev	/5Phos/ttcctgcaagctc tcgatgc	Site-directed mutagenesis of pBAD33-pdeH-FLAG
pdeH-(2-7)XA-fwd	/5Phos/GCCGCAGCGcgaa taagcaaccctgaagcaagc	Site-directed mutagenesis of pBAD33-pdeH-FLAG

Table 2.4 continued from previous page

Primer	Sequence (5' to 3')	Purpose
pdeH-(2-7)XA-rev	/5Phos/GGCGGCTGCCATc gtttcactcctggccttc	Site-directed mutagenesis of pBAD33-pdeH-FLAG
pdeH-(8-13)XA-fwd	/5Phos/GCCGCAGCGgcaa gcatcgagagcttgca	Site-directed mutagenesis of pBAD33-pdeH-FLAG
pdeH-(8-13)XA-rev	/5Phos/GGCGGCTGCctgg ataacctgccttatCATcgt ttc	Site-directed mutagenesis of pBAD33-pdeH-FLAG
pdeH-(15-19)XA-fwd	/5Phos/GCCGCAGCGcagg aacggcgtttttggttg	Site-directed mutagenesis of pBAD33-pdeH-FLAG
pdeH-(15-19)XA-rev	/5Phos/GGCGGCTGCttca gggttgcttattcgctgga	Site-directed mutagenesis of pBAD33-pdeH-FLAG
pdeH-(20-25)XA-fwd	/5Phos/GCCGCAGCGttgc agtgtgagcgtgcttac	Site-directed mutagenesis of pBAD33-pdeH-FLAG
pdeH-(20-25)XA-rev	/5Phos/GGCGGCTGCcaag ctctcgatgcttgcttcag	Site-directed mutagenesis of pBAD33-pdeH-FLAG
pdeH-F1-fwd	ATGTAGTAAAGGCCCGGCAG AC	Donor DNA construction of pdeH-3xFLAG flanked by 500 bp homology
pdeH-F1-rev	GCTCAATCTGACGCAGGATT TTTGG	Donor DNA construction of pdeH-3xFLAG flanked by 500 bp homology
pdeH-F2-fwd	CCAAAAATCCTGCGTCAGAT TGAGC	Donor DNA construction of pdeH-3xFLAG flanked by 500 bp homology
pdeH-F2-rev	TAGCGCCAGAACCGCCG	Donor DNA construction of pdeH-3xFLAG flanked by 500 bp homology
pdeH-F3-fwd	ATACGGCGGTTCTGGCGCTA AGCGGTGGCGGCG	Donor DNA construction of pdeH-3xFLAG flanked by 500 bp homology

Table 2.4 continued from previous page

Primer	Sequence (5' to 3')	Purpose
pdeH-F3-rev	GCGGAAAATGAGGCAGCtta CTTATATAACTCGTCCATGC CATGAGTG	Donor DNA construction of pdeH-3xFLAG flanked by 500 bp homology
pdeH-F4-fwd	taaGCTGCCTCATTTTCCGC CTG	Donor DNA construction of pdeH-3xFLAG flanked by 500 bp homology
pdeH-F4-rev	AGATTAAACGTCCAGTTATT GTTCTTTTCAGAGAG	Donor DNA construction of pdeH-3xFLAG flanked by 500 bp homology
B-pdeH-frag-fwd	AATAACTGGACGTTTAATCT tcgagttcatgtgcagctcc	Donor DNA construction of pdeH-3xFLAG flanked by 500 bp homology
B-pdeH-frag-rev	CTGCCGGGCCTTTACTACAT gtagggataacagggtaata gatctaagc	Donor DNA construction of pdeH-3xFLAG flanked by 500 bp homology
pdeH-val-1-fwd	GACTGCTCACTCTCCAGCCA G	Validation of pdeH-3xFLAG insertion at pdeH locus
pdeH-val-1-rev	GTCATCAATGGCGATCCGCC	Validation of pdeH-3xFLAG insertion at pdeH locus
yhjHb3525-366 N20 fwd	/5Phos/GCCAGGGAAGACG CTCAATCgttttagagctag aaatagcaagttaaaataag g	Replacement of N20 for targeting pdeH locus in pTarget
pTarget-N20-ins-rev	/5Phos/actagtattatac ctaggactgagctag	Replacement of N20 for targeting pdeH locus in pTarget
pdeH-pTarget-ins-fwd	ATGTAGTAAAGGCCCGGCAG AC	Amplification of assembled pdeH-3xFLAG donor DNA for insertion into pTarget-pdeH(N20)

Table 2.4 continued from previous page

Primer	Sequence (5' to 3')	Purpose
pdeH-pTarget-ins-rev	TCAGAGAGGCGAATCAGGCG	Amplification of assembled pdeH-3xFLAG donor DNA for insertion into pTarget-pdeH(N20)
B-pdeH-pTarget fwd	CGCCTGATTCGCCTCTCTGA tcgagttcatgtgcagctcc	Amplification of pTarget-pdeH(N20) backbone for insertion of pdeH-3xFLAG donor DNA
B-pdeH-pTarget rev	CTGCCGGGCCTTTACTACAT gtagggataacagggtaata gatctaagc	Amplification of pTarget-pdeH(N20) backbone for insertion of pdeH-3xFLAG donor DNA
B-R-ins-3xFLAG-rev	catcgtttcactcctggcct tc	Backbone amplification of pBAD33 to create pBAD33-3xFLAG
B-R-ins-3xFLAG-fwd	GACTATAAGGATCATGATGG GGATTACAAGGATCATGATA TCGATTACAAAGACGACGAC GATAAAtgagtcgacctgca ggcatg	Backbone amplification of pBAD33 to create pBAD33-3xFLAG
B-R-3xFLAG-ins-rev	TTTATCGTCGTCGTCTTTGT AATCGATATCATGATCCTTG TAATCCCCATCATGATCCTT ATAGTCcatcgtttcactcc tggccttc	Backbone amplification of pBAD33 to create pBAD33-3xFLAG
B-R-3xFLAG-ins-fwd	tgagtcgacctgcaggcatg	Backbone amplification of pBAD33 to create pBAD33-3xFLAG
3xFLAG-clpS-fwd	acaaagacgacgacgataaa GGTAAAACGAACGACTGGCT GG	3xFLAG-ClpS insertion into pBAD33-3xFLAG

Table 2.4 continued from previous page

Primer	Sequence (5' to 3')	Purpose
3xFLAG-clpS-rev	catgcctgcaggtcgactca GGCTTTTTCTAGCGTACACA GCAATG	3xFLAG-ClpS insertion into pBAD33-3xFLAG
clpS-3xFLAG-fwd	aggccaggagtgaaacgatg GGTAAAACGAACGACTGGCT GG	ClpS-3xFLAG insertion into pBAD33-3xFLAG
clpS-3xFLAG-rev	ccatcatgatccttatagtc GGCTTTTTCTAGCGTACACA GCAATG	ClpS-3xFLAG insertion into pBAD33-3xFLAG
ClpS-NTD-rev	/5Phos/catcgtttcactc ctggccttcg	Site-directed mutagenesis of pBAD33-clpS-3xFLAG for NTE removal
NTE-fwd	/5Phos/TCTATGTATAAAG TGATATTAGTCAATGATGAT TACACTCCG	Site-directed mutagenesis of pBAD33-clpS-3xFLAG for NTE removal
3xFLAG-iscA-fwd	acaaagacgacgacgataaa TCGATTACACTGAGCGACAG TGC	3xFLAG-IscA insertion into pBAD33-3xFLAG
3xFLAG-iscA-rev	catgcctgcaggtcgactca AACGTGGAAGCTTTCGCCGC	3xFLAG-IscA insertion into pBAD33-3xFLAG
3xFLAG-erpA-fwd	acaaagacgacgacgataaa AGTGATGACGTAGCACTGCC G	3xFLAG-ErpA insertion into pBAD33-3xFLAG
3xFLAG-erpA-rev	catgcctgcaggtcgactca GATACTAAAGGAAGAACCGC AACCGC	3xFLAG-ErpA insertion into pBAD33-3xFLAG
3xFLAG-sufA-fwd	acaaagacgacgacgataaa GACATGCATTCAGGAACCTT TAACCC	3xFLAG-SufA insertion into pBAD33-3xFLAG

Table 2.4 continued from previous page

Primer	Sequence (5' to 3')	Purpose
3xFLAG-sufA-rev	catgcctgcaggtcgactca TACCCCAAAGCTTTCGCCAC AG	3xFLAG-SufA insertion into pBAD33-3xFLAG
3xFLAG-nfuA-fwd	acaaagacgacgacgataaa ATCCGTATTTCCGATGCTGC ACAAG	3xFLAG-NfuA insertion into pBAD33-3xFLAG
3xFLAG-nfuA-rev	catgcctgcaggtcgactca GTAGTAGGAGTGTTCGCCGC G	3xFLAG-NfuA insertion into pBAD33-3xFLAG
iscA-3xFLAG-fwd	aggccaggagtgaaacgatg TCGATTACACTGAGCGACAG TGC	IscA-3xFLAG insertion into pBAD33-3xFLAG
iscA-3xFLAG-rev	ccatcatgatccttatagtc AACGTGGAAGCTTTCGCCGC	IscA-3xFLAG insertion into pBAD33-3xFLAG
erpA-3xFLAG-fwd	aggccaggagtgaaacgatg AGTGATGACGTAGCACTGCC G	ErpA-3xFLAG insertion into pBAD33-3xFLAG
erpA-3xFLAG-rev	ccatcatgatccttatagtc GATACTAAAGGAAGAACCGC AACCGC	ErpA-3xFLAG insertion into pBAD33-3xFLAG
sufA-3xFLAG-fwd	aggccaggagtgaaacgatg GACATGCATTCAGGAACCTT TAACCC	SufA-3xFLAG insertion into pBAD33-3xFLAG
sufA-3xFLAG-rev	ccatcatgatccttatagtc TACCCCAAAGCTTTCGCCAC AG	SufA-3xFLAG insertion into pBAD33-3xFLAG
nfuA-3xFLAG-fwd	aggccaggagtgaaacgatg ATCCGTATTTCCGATGCTGC ACAAG	NfuA-3xFLAG insertion into pBAD33-3xFLAG

Table 2.4 continued from previous page

Primer	Sequence (5' to 3')	Purpose
nfuA-3xFLAG-rev	ccatcatgatccttatagtc GTAGTAGGAGTGTTCGCCGC G	NfuA-3xFLAG insertion into pBAD33-3xFLAG
3xFLAG-gadX-fwd	acaaagacgacgacgataaa CAATCACTACATGGGAATTG TCTAATTGCGTATG	3xFLAG-GadX insertion into pBAD33-3xFLAG
3xFLAG-gadX-rev	catgcctgcaggtcgactca TAATCTTATTCCTTCCGCAG AACGGTC	3xFLAG-GadX insertion into pBAD33-3xFLAG
3xFLAG-nsrR-fwd	acaaagacgacgacgataaa CAGTTAACGAGTTTCACTGA TTACGGATTACG	3xFLAG-NsrR insertion into pBAD33-3xFLAG
3xFLAG-nsrR-rev	catgcctgcaggtcgactca CTCCACCAGCAATAATTTAT AAAGCGGTTG	3xFLAG-NsrR insertion into pBAD33-3xFLAG
3xFLAG-fliZ-fwd	acaaagacgacgacgataaa ATGGTGCAGCACCTGAAAAG ACG	3xFLAG-FliZ insertion into pBAD33-3xFLAG
3xFLAG-fliZ-rev	catgcctgcaggtcgactca ATATATATCAGAAGAAGGCA GGCTGGAGG	3xFLAG-FliZ insertion into pBAD33-3xFLAG
3xFLAG-csgD-fwd	acaaagacgacgacgataaa TTTAATGAAGTCCATAGTAT TCATGGTCATACATTATTGT TG	3xFLAG-CsgD insertion into pBAD33-3xFLAG
3xFLAG-csgD-rev	catgcctgcaggtcgactca TCGCCTGAGGTTATCGTTTG CC	3xFLAG-CsgD insertion into pBAD33-3xFLAG
3xFLAG-fecI-fwd	acaaagacgacgacgataaa TCTGACCGCGCCACTACC	3xFLAG-FecI insertion into pBAD33-3xFLAG

Table 2.4 continued from previous page

Primer	Sequence (5' to 3')	Purpose
3xFLAG-fecI-rev	catgcctgcaggtcgactca TAACCCATACTCCAGACGGA ACAGC	3xFLAG-FecI insertion into pBAD33-3xFLAG
3xFLAG-iraD-fwd	acaaagacgacgacgataaa ATGCGACAATCACTTCAGGC TGTTTTAC	3xFLAG-IraD insertion into pBAD33-3xFLAG
3xFLAG-iraD-rev	catgcctgcaggtcgactca GCTGACATTCTCCAGCGTCG C	3xFLAG-IraD insertion into pBAD33-3xFLAG
3xFLAG-yeiL-fwd	acaaagacgacgacgataaa AGTGAATCCGCGTTTAAGGA TTGC	3xFLAG-YeiL insertion into pBAD33-3xFLAG
3xFLAG-yeiL-rev	catgcctgcaggtcgactca CTGCATCATCCCGGAGAATT TATTCTCC	3xFLAG-YeiL insertion into pBAD33-3xFLAG
3xFLAG-fliS-fwd	acaaagacgacgacgataaa TACGCGGCAAAAGGCACCC	3xFLAG-FliS insertion into pBAD33-3xFLAG
3xFLAG-fliS-rev	catgcctgcaggtcgactca GACTGGGTCCTGAATCAAAG AAGGG	3xFLAG-FliS insertion into pBAD33-3xFLAG
3xFLAG-gadX-fwd	aggccaggagtgaaacgatg CAATCACTACATGGGAATTG TCTAATTGCGTATG	3xFLAG-GadX insertion into pBAD33-3xFLAG
3xFLAG-gadX-rev	ccatcatgatccttatagtc TAATCTTATTCCTTCCGCAG AACGGTC	3xFLAG-GadX insertion into pBAD33-3xFLAG
3xFLAG-nsrR-fwd	aggccaggagtgaaacgatg CAGTTAACGAGTTTCACTGA TTACGGATTACG	3xFLAG-NsrR insertion into pBAD33-3xFLAG

Table 2.4 continued from previous page

Primer	Sequence (5' to 3')	Purpose
3xFLAG-nsrR-rev	ccatcatgatccttatagtc CTCCACCAGCAATAATTTAT AAAGCGGTTG	3xFLAG-NsrR insertion into pBAD33-3xFLAG
3xFLAG-fliZ-fwd	aggccaggagtgaaacgatg ATGGTGCAGCACCTGAAAAG ACG	3xFLAG-FliZ insertion into pBAD33-3xFLAG
3xFLAG-fliZ-rev	ccatcatgatccttatagtc ATATATATCAGAAGAAGGCA GGCTGGAGG	3xFLAG-FliZ insertion into pBAD33-3xFLAG
3xFLAG-csgD-fwd	aggccaggagtgaaacgatg TTTAATGAAGTCCATAGTAT TCATGGTCATACATTATTGT TG	3xFLAG-CsgD insertion into pBAD33-3xFLAG
3xFLAG-csgD-rev	ccatcatgatccttatagtc TCGCCTGAGGTTATCGTTTG CC	3xFLAG-CsgD insertion into pBAD33-3xFLAG
3xFLAG-fecI-fwd	aggccaggagtgaaacgatg TCTGACCGCGCCACTACC	3xFLAG-FecI insertion into pBAD33-3xFLAG
3xFLAG-fecI-rev	ccatcatgatccttatagtc TAACCCATACTCCAGACGGA ACAGC	3xFLAG-FecI insertion into pBAD33-3xFLAG
3xFLAG-iraD-fwd	aggccaggagtgaaacgatg ATGCGACAATCACTTCAGGC TGTTTTAC	3xFLAG-IraD insertion into pBAD33-3xFLAG
3xFLAG-iraD-rev	ccatcatgatccttatagtc GCTGACATTCTCCAGCGTCG C	3xFLAG-IraD insertion into pBAD33-3xFLAG
3xFLAG-yeiL-fwd	aggccaggagtgaaacgatg AGTGAATCCGCGTTTAAGGA TTGC	3xFLAG-YeiL insertion into pBAD33-3xFLAG

Table 2.4 continued from previous page

Primer	Sequence (5' to 3')	Purpose
3xFLAG-yeiL-rev	ccatcatgatccttatagtc CTGCATCATCCCGGAGAATT TATTCTCC	3xFLAG-YeiL insertion into pBAD33-3xFLAG
3xFLAG-fliS-fwd	aggccaggagtgaaacgatg TACGCGGCAAAAGGCACCC	3xFLAG-FliS insertion into pBAD33-3xFLAG
3xFLAG-fliS-rev	ccatcatgatccttatagtc GACTGGGTCCTGAATCAAAG AAGGG	3xFLAG-FliS insertion into pBAD33-3xFLAG
B-A2c-ins-3xFLAG-fwd	GACTACAAGGACCACGACGG T	Backbone amplification of pBbA2c for C-terminal 3xFLAG-tagged substrate insertion
B-A2c-ins-3xFLAG-rev	catatgtatatctccttctt aaaagatcttttgaattctt ttctc	Backbone amplification of pBbA2c for C-terminal 3xFLAG-tagged substrate insertion
bolA-3xFLAG-fwd	aagaaggagatatacatatg ATGATACGTGAGCGGATAGA AGAAAAATTAAGG	BolA-3xFLAG insertion into pBbA2c
bolA-3xFLAG-rev	CCGTCGTGGTCCTTGTAGTC CGCGATGCTTCCTGCTCCAC	BolA-3xFLAG insertion into pBbA2c
gadE-3xFLAG-fwd	aagaaggagatatacatatg ATTTTTCTCATGACGAAAGA TTCTTTTCTTTTACAGG	GadE-3xFLAG insertion into pBbA2c
gadE-3xFLAG-rev	CCGTCGTGGTCCTTGTAGTC AAAATAAGATGTGATACCCA GGGTGACG	GadE-3xFLAG insertion into pBbA2c
bssR-3xFLAG-fwd	aagaaggagatatacatatg TTCGTTGACAGACAGCGAAT CGAT	BssR-3xFLAG insertion into pBbA2c

Table 2.4 continued from previous page

Primer	Sequence (5' to 3')	Purpose
bssR-3xFLAG-rev	CCGTCGTGGTCCTTGTAGTC GGCCTTCTCAAGCATGGCG	BssR-3xFLAG insertion into pBbA2c
narP-3xFLAG-fwd	aagaaggagatatacatatg CCTGAAGCAACACCTTTTCA GGTG	NarP-3xFLAG insertion into pBbA2c
narP-3xFLAG-rev	CCGTCGTGGTCCTTGTAGTC TTGTGCCCCGCGTTGTTG	NarP-3xFLAG insertion into pBbA2c
B-A2c-3xFLAG-ins-fwd	TAAGGATCCAAACTCGAGTA AGGATCTCCAGG	Backbone amplification of pBbA2c for N-terminal 3xFLAG-tagged substrate insertion
B-A2c-3xFLAG-ins-rev	CTTGTCGTCGTCGTCCTTGT AGTCGATG	Backbone amplification of pBbA2c for N-terminal 3xFLAG-tagged substrate insertion
3xFLAG-gadE-fwd	ACAAGGACGACGACGACAAG ATTTTTCTCATGACGAAAGA TTCTTTTCTTTTACAGG	3xFLAG-GadE insertion into pBbA2c
3xFLAG-gadE-rev	tactcgagtttggatcctta AAAATAAGATGTGATACCCA GGGTGACG	3xFLAG-GadE insertion into pBbA2c
3xFLAG-bssR-fwd	ACAAGGACGACGACGACAAG TTCGTTGACAGACAGCGAAT CGAT	3xFLAG-BssR insertion into pBbA2c
3xFLAG-bssR-rev	tactcgagtttggatcctta GGCCTTCTCAAGCATGGCG	3xFLAG-BssR insertion into pBbA2c
3xFLAG-narP-fwd	ACAAGGACGACGACGACAAG CCTGAAGCAACACCTTTTCA GGTG	3xFLAG-NarP insertion into pBbA2c

Table 2.4 continued from previous page

Primer	Sequence (5' to 3')	Purpose
3xFLAG-narP-rev	tactcgagtttggatcctta TTGTGCCCCGCGTTGTTG	3xFLAG-NarP insertion into pBbA2c
3xFLAG-bolA-fwd	ACAAGGACGACGACGACAAG ATGATACGTGAGCGGATAGA AGAAAAATTAAGG	3xFLAG-BolA insertion into pBbA2c
3xFLAG-bolA-rev	tactcgagtttggatcctta CGCGATGCTTCCTGCTCCAC	3xFLAG-BolA insertion into pBbA2c

2.8 Supplementary Datasets

Dataset 2.1. Proteomic results for exponential phase degradation profiling.
This dataset describes all proteins identified through LC-MS/MS analysis of BONCAT-enriched proteins from the exponential phase stability screen. Column 1 provides the corresponding NCBI Gene Symbol for the protein. Columns 2-16 provide the TMT reporter ion abundances for each sample. Column 17 provides the estimated degradation rate. Columns 18 and 19 provide the p-value and FDR-adjusted p-value for the estimated degradation rate, respectively.

Dataset 2.2. Proteomic results for stationary phase degradation profiling.
This dataset describes all proteins identified through LC-MS/MS analysis of BONCAT-enriched proteins from the stationary phase stability screen. Column 1 provides the corresponding NCBI Gene Symbol for the protein. Columns 2-16 provide the TMT reporter ion abundances for each sample. Column 17 provides the estimated degradation rate. Columns 18 and 19 provide the p-value and FDR-adjusted p-value for the estimated degradation rate, respectively.

Dataset 2.3. Machine learning analysis results.
This dataset provides the machine learning training set and the feature values for all proteins in the analysis. Column 1 provides the UniProt ID. Column 2 provides the official NCBI Gene Symbol for the protein. Column 3 provides the feature name. Column 4 provides the feature value. Column 5 provides the coefficient for the feature from the XGBoost model. Column 6 reports the class label used during training if the indicated protein was present in the training data. Column 7 reports the probability estimate for instability.

References

(1) Sauer, R. T.; Baker, T. A. AAA+ proteases: ATP-fueled machines of protein destruction. *Annu Rev Biochem* **2011**, *80*, 587–612.

(2) Levchenko, I.; Seidel, M.; Sauer, R. T.; Baker, T. A. A specificity-enhancing factor for the ClpXP degradation machine. *Science* **2000**, *289*, 2354–2356.

(3) Erbse, A.; Schmidt, R.; Bornemann, T.; Schneider-Mergener, J.; Mogk, A.; Zahn, R.; Dougan, D. A.; Bukau, B. ClpS is an essential component of the N-end rule pathway in Escherichia coli. *Nature* **2006**, *439*, 753–756.

(4) Ninnis, R. L.; Spall, S. K.; Talbo, G. H.; Truscott, K. N.; Dougan, D. A. Modification of PATase by L/F-transferase generates a ClpS-dependent N-end rule substrate in Escherichia coli. *EMBO J* **2009**, *28*, 1732–1744.

(5) Nath, K.; Koch, A. L. Protein degradation in Escherichia coli. I. Measurement of rapidly and slowly decaying components. *J Biol Chem* **1970**, *245*, 2889–2900.

(6) Willetts, N. S. Intracellular protein breakdown in growing cells of Escherichia coli. *Biochem J* **1967**, *103*, 462–466.

(7) Gur, E.; Biran, D.; Ron, E. Z. Regulated proteolysis in Gram-negative bacteria–how and when? *Nat Rev Microbiol* **2011**, *9*, 839–848.

(8) Keiler, K. C.; Waller, P. R.; Sauer, R. T. Role of a peptide tagging system in degradation of proteins synthesized from damaged messenger RNA. *Science* **1996**, *271*, 990–993.

(9) Domian, I. J.; Quon, K. C.; Shapiro, L. Cell type-specific phosphorylation and proteolysis of a transcriptional regulator controls the G1-to-S transition in a bacterial cell cycle. *Cell* **1997**, *90*, 415–424.

(10) Cahn, J. K. B.; Baumschlager, A.; Brinkmann-Chen, S.; Arnold, F. H. Mutations in adenine-binding pockets enhance catalytic properties of NAD (P) H-dependent enzymes. *Protein Engineering Design and Selection* **2016**, *19*, 31–38,

(11) Hengge, R. Proteolysis of sigmaS (RpoS) and the general stress response in Escherichia coli. *Res Microbiol* **2009**, *160*, 667–676.

(12) Westphal, K.; Langklotz, S.; Thomanek, N.; Narberhaus, F. A trapping approach reveals novel substrates and physiological functions of the essential protease FtsH in Escherichia coli. *J Biol Chem* **2012**, *287*, 42962–42971.

(13) Arends, J.; Griego, M.; Thomanek, N.; Lindemann, C.; Kutscher, B.; Meyer, H. E.; Narberhaus, F. An Integrated Proteomic Approach Uncovers Novel Substrates and Functions of the Lon Protease in Escherichia coli. *Proteomics* **2018**, *18*, e1800080.

(14) Neher, S. B.; n, J.; Oakes, E. C.; Bakalarski, C. E.; Sauer, R. T.; Gygi, S. P.; Baker, T. A. Proteomic profiling of ClpXP substrates after DNA damage reveals extensive instability within SOS regulon. *Mol Cell* **2006**, *22*, 193–204.

(15) Rei Liao, J. Y.; van Wijk, K. J. Discovery of AAA+ Protease Substrates through Trapping Approaches. *Trends Biochem Sci* **2019**, *44*, 528–545.

(16) Nagar, N.; Ecker, N.; Loewenthal, G.; Avram, O.; Ben-Meir, D.; Biran, D.; Ron, E.; Pupko, T. Harnessing machine learning to unravel protein degradation in Escherichia coli. *mSystems* **2021**, *6*.

(17) Bergkessel, M.; Basta, D. W.; Newman, D. K. The physiology of growth arrest: uniting molecular and environmental microbiology. *Nat Rev Microbiol* **2016**, *14*, 549–562.

(18) Mendes, M. L.; Dittmar, G. Targeted proteomics on its way to discovery. *Proteomics* **2022**, *22*, e2100330.

(19) Dieterich, D. C.; Link, A. J.; Graumann, J.; Tirrell, D. A.; Schuman, E. M. Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT). *Proc Natl Acad Sci U S A* **2006**, *103*, 9482–9487.

(20) Babin, B. M.; Bergkessel, M.; Sweredoski, M. J.; Moradian, A.; Hess, S.; Newman, D. K.; Tirrell, D. A. SutA is a bacterial transcription factor expressed during slow growth in Pseudomonas aeruginosa. *Proc Natl Acad Sci U S A* **2016**, *113*, 597–605.

(21) Babin, B. M.; Atangcho, L.; van Eldijk, M. B.; Sweredoski, M. J.; Moradian, A.; Hess, S.; Tolker-Nielsen, T.; Newman, D. K.; Tirrell, D. A. Selective Proteomic Analysis of Antibiotic-Tolerant Cellular Subpopulations in Pseudomonas aeruginosa Biofilms. *mBio* **2017**, *8*, 10.1128/mbio.01593–17.

(22) Hatzenpichler, R.; Connon, S. A.; Goudeau, D.; Malmstrom, R. R.; Woyke, T.; Orphan, V. J. Visualizing in situ translational activity for identifying and sorting slow-growing archaeal-bacterial consortia. *Proc Natl Acad Sci U S A* **2016**, *113*, E4069–4078.

(23) McShane, E.; Sin, C.; Zauber, H.; Wells, J. N.; Donnelly, N.; Wang, X.; Hou, J.; Chen, W.; Storchova, Z.; Marsh, J. A.; Valleriani, A.; Selbach, M. Kinetic Analysis of Protein Stability Reveals Age-Dependent Degradation. *Cell* **2016**, *167*, 803–815.

(24) Ngo, J. T.; Champion, J. A.; Mahdavi, A.; Tanrikulu, I. C.; Beatty, K. E.; Connor, R. E.; Yoo, T. H.; Dieterich, D. C.; Schuman, E. M.; Tirrell, D. A. Cell-selective metabolic labeling of proteins. *Nat Chem Biol* **2009**, *5*, 715–717.

(25) Tanrikulu, I. C.; Schmitt, E.; Mechulam, Y.; Goddard, W. A.; Tirrell, D. A. Discovery of Escherichia coli methionyl-tRNA synthetase mutants for efficient labeling of proteins with azidonorleucine in vivo. *Proc Natl Acad Sci U S A* **2009**, *106*, 15285–15290.

(26) Ngo, J. T.; Babin, B. M.; Champion, J. A.; Schuman, E. M.; Tirrell, D. A. State-selective metabolic labeling of cellular proteins. *ACS Chem Biol* **2012**, *7*, 1326–1330.

(27) Li, J. et al. TMTpro reagents: a set of isobaric labeling mass tags enables simultaneous proteome-wide measurements across 16 samples. *Nat Methods* **2020**, *17*, 399–404.

(28) Thompson, A.; fer, J.; Kuhn, K.; Kienle, S.; Schwarz, J.; Schmidt, G.; Neumann, T.; Johnstone, R.; Mohammed, A. K.; Hamon, C. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem* **2003**, *75*, 1895–1904.

(29) Karp, P. D. et al. The EcoCyc Database (2023). *EcoSal Plus* **2023**, *11*, eesp00022023.

(30) Chang, W.; Cheng, J.; Allaire, J.; Sievert, C.; Schloerke, B.; Xie, Y.; Allen, J.; McPherson, J.; Dipert, A.; Borges, B. shiny: Web Application Framework for R, R package version 1.8.1.9001, https://github.com/rstudio/shiny, 2024.

(31) Bateman, A. et al. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res* **2023**, *51*, D523–D531.

(32) Kanehisa, M.; Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **2000**, *28*, 27–30.

(33) Pruteanu, M.; Baker, T. A. Proteolysis in the SOS response and metal homeostasis in Escherichia coli. *Res Microbiol* **2009**, *160*, 677–683.

(34) Mahmoud, S. A.; Chien, P. Regulated Proteolysis in Bacteria. *Annu Rev Biochem* **2018**, *87*, 677–696.

(35) Li, J.; Cai, Z.; Vaites, L. P.; Shen, N.; Mitchell, D. C.; Huttlin, E. L.; Paulo, J. A.; Harry, B. L.; Gygi, S. P. Proteome-wide mapping of short-lived proteins in human cells. *Mol Cell* **2021**, *81*, 4722–4735.

(36) Larance, M.; Ahmad, Y.; Kirkwood, K. J.; Ly, T.; Lamond, A. I. Global subcellular characterization of protein degradation using quantitative proteomics. *Mol Cell Proteomics* **2013**, *12*, 638–650.

(37) Jenal, U.; Reinders, A.; Lori, C. Cyclic di-GMP: second messenger extraordinaire. *Nat Rev Microbiol* **2017**, *15*, 271–284.

(38) Povolotsky, T. L.; Hengge, R. Genome-Based Comparison of Cyclic Di-GMP Signaling in Pathogenic and Commensal Escherichia coli Strains. *J Bacteriol* **2016**, *198*, 111–126.

(39) Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589.

(40) Guzman, L. M.; Belin, D.; Carson, M. J.; Beckwith, J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J Bacteriol* **1995**, *177*, 4121–4130.

(41) Ko, M.; Park, C. Two novel flagellar components and H-NS are involved in the motor function of Escherichia coli. *J Mol Biol* **2000**, *303*, 371–382.

(42) Kitagawa, R.; Takaya, A.; Yamamoto, T. Dual regulatory pathways of flagellar gene expression by ClpXP protease in enterohaemorrhagic Escherichia coli. *Microbiology (Reading)* **2011**, *157*, 3094–3103.

(43) Tomoyasu, T.; Takaya, A.; Isogai, E.; Yamamoto, T. Turnover of FlhD and FlhC, master regulator proteins for Salmonella flagellum biogenesis, by the ATP-dependent ClpXP protease. *Mol Microbiol* **2003**, *48*, 443–452.

(44) Takaya, A.; Erhardt, M.; Karata, K.; Winterberg, K.; Yamamoto, T.; Hughes, K. T. YdiV: a dual function protein that targets FlhDC for ClpXP-dependent degradation by promoting release of DNA-bound FlhDC complex. *Mol Microbiol* **2012**, *83*, 1268–1284.

(45) Pesavento, C.; Becker, G.; Sommerfeldt, N.; Possling, A.; Tschowri, N.; Mehlis, A.; Hengge, R. Inverse regulatory coordination of motility and curli-mediated adhesion in Escherichia coli. *Genes Dev* **2008**, *22*, 2434–2446.

(46) Bachmair, A.; Finley, D.; Varshavsky, A. In vivo half-life of a protein is a function of its amino-terminal residue. *Science* **1986**, *234*, 179–186.

(47) Tobias, J. W.; Shrader, T. E.; Rocap, G.; Varshavsky, A. The N-end rule in bacteria. *Science* **1991**, *254*, 1374–1377.

(48) Hou, J. Y.; Sauer, R. T.; Baker, T. A. Distinct structural elements of the adaptor ClpS are required for regulating degradation by ClpAP. *Nat Struct Mol Biol* **2008**, *15*, 288–294.

(49) Dougan, D. A.; Reid, B. G.; Horwich, A. L.; Bukau, B. ClpS, a substrate modulator of the ClpAP machine. *Mol Cell* **2002**, *9*, 673–683.

(50) Torres-Delgado, A.; Kotamarthi, H. C.; Sauer, R. T.; Baker, T. A. The Intrinsically Disordered N-terminal Extension of the ClpS Adaptor Reprograms Its Partner AAA+ ClpAP Protease. *J Mol Biol* **2020**, *432*, 4908–4921.

(51) Kim, S.; Fei, X.; Sauer, R. T.; Baker, T. A. AAA+ protease-adaptor structures reveal altered conformations and ring specialization. *Nat Struct Mol Biol* **2022**, *29*, 1068–1079.

(52) Zeth, K.; Ravelli, R. B.; Paal, K.; Cusack, S.; Bukau, B.; Dougan, D. A. Structural analysis of the adaptor protein ClpS in complex with the N-terminal domain of ClpA. *Nat Struct Biol* **2002**, *9*, 906–911.

(53) Wofford, J. D.; Bolaji, N.; Dziuba, N.; Outten, F. W.; Lindahl, P. A. Evidence that a respiratory shield in Escherichia coli protects a low-molecular-mass FeII pool from O2-dependent oxidation. *Journal of Biological Chemistry* **2019**, *294*, 50–62.

(54) Esquilin-Lebron, K.; Dubrac, S.; Barras, F.; Boyd, J. M. Bacterial Approaches for Assembling Iron-Sulfur Proteins. *mBio* **2021**, *12*, e0242521.

(55) Hartmann, A.; Braun, V. Iron uptake and iron limited growth of Escherichia coli K-12. *Arch Microbiol* **1981**, *130*, 353–356.

(56) Oppermann, S.; fflin, S.; Friedrich, T. ErpA is important but not essential for the Fe/S cluster biogenesis of Escherichia coli NADH:ubiquinone oxidoreductase (complex I). *Biochim Biophys Acta Bioenerg* **2020**, *1861*, 148286.

(57) Beloin, C.; Valle, J.; Latour-Lambert, P.; Faure, P.; Kzreminski, M.; Balestrino, D.; Haagensen, J. A.; Molin, S.; Prensier, G.; Arbeille, B.; Ghigo, J. M. Global impact of mature biofilm lifestyle on Escherichia coli K-12 gene expression. *Mol Microbiol* **2004**, *51*, 659–674.

(58) Schneider, D. A.; Gaal, T.; Gourse, R. L. NTP-sensing by rRNA promoters in Escherichia coli is direct. *Proc Natl Acad Sci U S A* **2002**, *99*, 8602–8607.

(59) Schweppe, D. K.; Eng, J. K.; Yu, Q.; Bailey, D.; Rad, R.; Navarrete-Perea, J.; Huttlin, E. L.; Erickson, B. K.; Paulo, J. A.; Gygi, S. P. Full-Featured, Real-Time Database Searching Platform Enables Fast and Accurate Multiplexed Quantitative Proteomics. *J Proteome Res* **2020**, *19*, 2026–2034.

(60) Yeom, J.; Groisman, E. A. Reduced ATP-dependent proteolysis of functional proteins during nutrient limitation speeds the return of microbes to a growth state. *Sci Signal* **2021**, *14*.

(61) Griffith, K. L.; Shah, I. M.; Wolf, R. E. Proteolytic degradation of Escherichia coli transcription activators SoxS and MarA as the mechanism for reversing the induction of the superoxide (SoxRS) and multiple antibiotic resistance (Mar) regulons. *Mol Microbiol* **2004**, *51*, 1801–1816.

(62) Bhardwaj, N.; Syal, K.; Chatterji, D. The role of -subunit of Escherichia coli RNA polymerase in stress response. *Genes Cells* **2018**, *23*, 357–369.

(63) Ishihama, A.; Fujita, N.; Glass, R. E. Subunit assembly and metabolic stability of E. coli RNA polymerase. *Proteins* **1987**, *2*, 42–53.

(64) Santos, J. M.; Lobo, M.; Matos, A. P.; De Pedro, M. A.; Arraiano, C. M. The gene bolA regulates dacA (PBP5), dacC (PBP6) and ampC (AmpC), promoting normal morphology in Escherichia coli. *Mol Microbiol* **2002**, *45*, 1729–1740.

(65) Galego, L.; Barahona, S.; o, C. V.; Arraiano, C. M. Phosphorylation status of BolA affects its role in transcription and biofilm development. *FEBS J* **2021**, *288*, 961–979.

(66) Domka, J.; Lee, J.; Wood, T. K. YliH (BssR) and YceP (BssS) regulate Escherichia coli K-12 biofilm formation by influencing cell signaling. *Appl Environ Microbiol* **2006**, *72*, 2449–2459.

(67) Chen, T.; Guestrin, C. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Association for Computing Machinery: San Francisco, California, USA, 2016, pp 785–794.

(68) Herman, C.; venet, D.; D'Ari, R.; Bouloc, P. Degradation of sigma 32, the heat shock regulator in Escherichia coli, is governed by HflB. *Proc Natl Acad Sci U S A* **1995**, *92*, 3516–3520.

(69) Rood, K. L.; Clark, N. E.; Stoddard, P. R.; Garman, S. C.; Chien, P. Adaptor-dependent degradation of a cell-cycle regulator uses a unique substrate architecture. *Structure* **2012**, *20*, 1223–1232.

(70) Flynn, J. M.; Neher, S. B.; Kim, Y. I.; Sauer, R. T.; Baker, T. A. Proteomic discovery of cellular substrates of the ClpXP protease reveals five classes of ClpX-recognition signals. *Mol Cell* **2003**, *11*, 671–683.

(71) Sarenko, O.; Klauck, G.; Wilke, F. M.; Pfiffer, V.; Richter, A. M.; Herbst, S.; Kaever, V.; Hengge, R. More than Enzymes That Make or Break Cyclic Di-GMP–Local Signaling in the Interactome of GGDEF/EAL Domain Proteins of Escherichia coli. *mBio* **2017**, *8*, 10.1128/mbio.01639–17.

(72) Pfiffer, V.; Sarenko, O.; Possling, A.; Hengge, R. Genetic dissection of Escherichia coli's master diguanylate cyclase DgcE: Role of the N-terminal MASE1 domain and direct signal input from a GTPase partner system. *PLoS Genet* **2019**, *15*, e1008059.

(73) Yeom, J.; Gao, X.; Groisman, E. A. Reduction in adaptor amounts establishes degradation hierarchy among protease substrates. *Proc Natl Acad Sci U S A* **2018**, *115*, E4483–E4492.

(74) Farrell, C. M.; Grossman, A. D.; Sauer, R. T. Cytoplasmic degradation of ssrA-tagged proteins. *Mol Microbiol* **2005**, *57*, 1750–1761.

(75) ndez, G.; Hou, J. Y.; Grant, R. A.; Sauer, R. T.; Baker, T. A. The ClpS adaptor mediates staged delivery of N-end rule substrates to the AAA+ ClpAP protease. *Mol Cell* **2011**, *43*, 217–228.

(76) Turgay, K.; Hahn, J.; Burghoorn, J.; Dubnau, D. Competence in Bacillus subtilis is controlled by regulated proteolysis of a transcription factor. *EMBO J* **1998**, *17*, 6730–6738.

(77) Gottesman, S.; Roche, E.; Zhou, Y.; Sauer, R. T. The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. *Genes Dev* **1998**, *12*, 1338–1347.

(78) Flynn, J. M.; Levchenko, I.; Seidel, M.; Wickner, S. H.; Sauer, R. T.; Baker, T. A. Overlapping recognition determinants within the ssrA degradation tag allow modulation of proteolysis. *Proc Natl Acad Sci U S A* **2001**, *98*, 10584–10589.

(79) Ezraty, B.; Vergnes, A.; Banzhaf, M.; Duverger, Y.; Huguenot, A.; Brochado, A. R.; Su, S. Y.; Espinosa, L.; Loiseau, L.; Py, B.; Typas, A.; Barras, F. Fe-S cluster biosynthesis controls uptake of aminoglycosides in a ROS-less death pathway. *Science* **2013**, *340*, 1583–1587.

(80) Gerstel, A.; o Beas, J.; Duverger, Y.; Bouveret, E.; Barras, F.; Py, B. Oxidative stress antagonizes fluoroquinolone drug sensitivity via the SoxR-SUF Fe-S cluster homeostatic axis. *PLoS Genet* **2020**, *16*, e1009198.

(81) Py, B.; Gerez, C.; Huguenot, A.; Vidaud, C.; Fontecave, M.; Ollagnier de Choudens, S.; Barras, F. The ErpA/NfuA complex builds an oxidation-resistant Fe-S cluster delivery pathway. *J Biol Chem* **2018**, *293*, 7689–7702.

(82) Mettert, E. L.; Outten, F. W.; Wanta, B.; Kiley, P. J. The impact of O(2) on the Fe-S cluster biogenesis requirements of Escherichia coli FNR. *J Mol Biol* **2008**, *384*, 798–811.

(83) Viola, M. G.; Perdikari, T. M.; Trebino, C. E.; Rahmani, N.; Mathews, K. L.; Pena, C. M.; Chua, X. Y.; Xuan, B.; LaBreck, C. J.; Fawzi, N. L.; Camberg, J. L. An enhancer sequence in the intrinsically disordered region of FtsZ promotes polymer-guided substrate processing by ClpXP protease. *Protein Sci* **2022**, *31*, e4306.

(84) Van der Lee, R.; Lang, B.; Kruse, K.; Gsponer, J.; nchez de Groot, N.; Huynen, M. A.; Matouschek, A.; Fuxreiter, M.; Babu, M. M. Intrinsically disordered segments affect protein half-life in the cell and during evolution. *Cell Rep* **2014**, *8*, 1832–1844.

(85) Tunyasuvunakool, K. et al. Highly accurate protein structure prediction for the human proteome. *Nature* **2021**, *596*, 590–596.

(86) Bhowmik, D.; Bhardwaj, N.; Chatterji, D. Influence of Flexible '' on the Activity of E. coli RNA Polymerase: A Thermodynamic Analysis. *Biophys J* **2017**, *112*, 901–910.

(87) Sarkar, P.; Sardesai, A. A.; Murakami, K. S.; Chatterji, D. Inactivation of the bacterial RNA polymerase due to acquisition of secondary structure by the subunit. *J Biol Chem* **2013**, *288*, 25076–25087.

(88) Mori, M.; Zhang, Z.; Banaei-Esfahani, A.; Lalanne, J. B.; Okano, H.; Collins, B. C.; Schmidt, A.; Schubert, O. T.; Lee, D. S.; Li, G. W.; Aebersold, R.; Hwa, T.; Ludwig, C. From coarse to fine: the absolute Escherichia coli proteome under diverse growth conditions. *Mol Syst Biol* **2021**, *17*, e9536.

(89) Browning, D. F.; Hobman, J. L.; Busby, S. J. W. K-12: things are seldom what they seem. *Microb Genom* **2023**, *9*.

(90) Datsenko, K. A.; Wanner, B. L. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc Natl Acad Sci U S A* **2000**, *97*, 6640–6645.

(91) Gibson, D. G.; Young, L.; Chuang, R. Y.; Venter, J. C.; Hutchison, C. A.; Smith, H. O. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* **2009**, *6*, 343–345.

(92) Lee, T. S.; Krupa, R. A.; Zhang, F.; Hajimorad, M.; Holtz, W. J.; Prasad, N.; Lee, S. K.; Keasling, J. D. BglBrick vectors and datasheets: A synthetic biology platform for gene expression. *J Biol Eng* **2011**, *5*, 12.

(93) Jiang, Y.; Chen, B.; Duan, C.; Sun, B.; Yang, J.; Yang, S. Multigene editing in the Escherichia coli genome via the CRISPR-Cas9 system. *Appl Environ Microbiol* **2015**, *81*, 2506–2514.

(94) Guo, J.; Wang, T.; Guan, C.; Liu, B.; Luo, C.; Xie, Z.; Zhang, C.; Xing, X. H. Improved sgRNA design in bacteria via genome-wide activity profiling. *Nucleic Acids Res* **2018**, *46*, 7052–7069.

(95) Wang, T.; Guan, C.; Guo, J.; Liu, B.; Wu, Y.; Xie, Z.; Zhang, C.; Xing, X. H. Pooled CRISPR interference screening enables genome-scale functional genomics study in bacteria with superior performance. *Nat Commun* **2018**, *9*, 2475.

(96) Szklarczyk, D.; Kirsch, R.; Koutrouli, M.; Nastou, K.; Mehryary, F.; Hachilif, R.; Gable, A. L.; Fang, T.; Doncheva, N. T.; Pyysalo, S.; Bork, P.; Jensen, L. J.; von Mering, C. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res* **2023**, *51*, D638–D646.

(97) Hong, V.; Presolski, S. I.; Ma, C.; Finn, M. G. Analysis and optimization of copper-catalyzed azide-alkyne cycloaddition for bioconjugation. *Angew Chem Int Ed Engl* **2009**, *48*, 9879–9883.

(98) Glenn, W. S.; Stone, S. E.; Ho, S. H.; Sweredoski, M. J.; Moradian, A.; Hess, S.; Bailey-Serres, J.; Tirrell, D. A. Bioorthogonal Noncanonical Amino Acid Tagging (BONCAT) Enables Time-Resolved Analysis of Protein Synthesis in Native Plant Tissue. *Plant Physiol* **2017**, *173*, 1543–1553.

(99) Ritchie, M. E.; Phipson, B.; Wu, D.; Hu, Y.; Law, C. W.; Shi, W.; Smyth, G. K. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **2015**, *43*, e47.

(100) Pedregosa, F. et al. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **2011**, *12*, 2825–2830.

(101) Paszke, A. et al. In *Advances in Neural Information Processing Systems*, ed. by Wallach, H.; Larochelle, H.; Beygelzimer, A.; d'Alché-Buc, F.; Fox, E.; Garnett, R., Curran Associates, Inc.: 2019; Vol. 32.

(102) Zeiler, M. D. ADADELTA: An Adaptive Learning Rate Method, 2012.

(103) Lundberg, S. M.; Lee, S.-I. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Curran Associates Inc.: Long Beach, California, USA, 2017, pp 4768–4777.

(104) Bagert, J. D.; Xie, Y. J.; Sweredoski, M. J.; Qi, Y.; Hess, S.; Schuman, E. M.; Tirrell, D. A. Quantitative, time-resolved proteomic analysis by combining bioorthogonal noncanonical amino acid tagging and pulsed stable isotope labeling by amino acids in cell culture. *Mol Cell Proteomics* **2014**, *13*, 1352–1358.

Chapter 3

ENGINEERED DESTABILIZATION OF NLL-METRS FOR TIME-RESOLVED ANALYSIS OF TRANSIENT PHYSIOLOGICAL STATES

3.1 Abstract

The state-selective metabolic labeling of proteins may be compromised when the duration of label incorporation exceeds that of the physiological state of interest. To address this deficiency in temporal resolution, we leveraged the widely conserved ssrA proteolytic pathway to destabilize the mutant methionyl-tRNA synthetase NLL-MetRS. We identified several ssrA tags bearing mutated ClpX recognition motifs that afforded a spectrum of destabilization. Destabilization enabled a concomitant reduction in NLL-MetRS abundance and label incorporation relative to the stable variants after cessation of expression. Fusion of NLL-MetRS with these tags and the fluorescent protein GFPmut3 enabled live cell stability and expression analyses. Finally, the transcriptional fusion of a destabilized NLL-MetRS mutant with a stringently regulated rRNA promoter demonstrated the advantage of engineered degradation in coupling protein labeling to transient promoter activity. Together, our results suggest that active degradation can improve the temporal resolution of state-selective metabolic labeling.

3.2 Introduction

Bacteria have evolved a multitude of gene expression programs to drive homeostasis and physiological adaptation. Among these, researchers have identified rich dynamics and heterogeneity in transcriptional activity across isogenic populations of bacterial cells, a strategy thought to enable survival of a small subpopulation of cells in the face of rare but highly adverse conditions on the basis of theoretical and experimental investigations.[1] For example, stochastic expression of the trimethylamine oxide respiratory system permitted growth of only a subpopulation of cells upon oxygen loss.[2] Such phenotypic heterogeneity in bacteria may also be relevant to human health: A landmark study of bacterial persistence, a phenomenon in which a subpopulation of bacteria survive exposure to a bactericidal drug concentration, attributed survival to a pre-existing phenotype not heritable by the progeny of surviving cells.[3,4] Taken together, these and similar studies highlight a need for methods that enable the characterization of cell subpopulations distinguished by stochasticity and transience.

Investigators have approached the proteomic analysis of cell subpopulations by coupling the bioorthogonal noncanonical amino acid tagging (BONCAT) method to pre-determined transcriptional states for state-selective analysis.[5] In this approach, the mutant methionyl-tRNA synthetase NLL-MetRS is transcriptionally fused to a promoter that is activated specifically in a physiological state of interest, thereby permitting the incorporation of azidonorleucine for the metabolic labeling of cell populations of interest. However, the proteolytic stability of NLL-MetRS, a mutant of the endogenous *E. coli* MetRS, may preclude accurate state-selective metabolic labeling of cells in transient physiological states because the duration of label incorporation may exceed that of the physiological state.

To address this challenge, we sought to engineer the instability of NLL-MetRS such that its abundance accurately reflects the activity of transiently activated promoters. To that end, we turned to the most characterized and efficient mechanism of active degradation in bacteria, the ssrA pathway. This pathway rescues ribosomes arrested on truncated mRNAs by releasing the truncated nascent polypeptide upon addition of an eleven residue C-terminal peptide (AANDENYALAA), which invokes rapid recognition and degradation of the tagged protein by the ClpXP protease.[6,7] The implementation of this active degradation pathway in synthetic biological circuits has underpinned the achievement of expression architectures that require dynamics

faster than the rate of cell division for protein depletion, ranging from the time-resolved visualization of promoter activity to the engineering of oscillatory circuits with reporter synchrony spanning 14 cell generations.[8,9] We found that these C-terminal ssrA tags can confer a wide range of proteolytic instability to NLL-MetRS. Time-series protein labeling experiments demonstrated that the response timescales of both NLL-MetRS abundance and enzymatic output can be shaped by active degradation. We then constructed translational fusions of destabilized NLL-MetRS variants to the fluorescent protein GFPmut3, enabling the time-resolved visualization of promoter and NLL-MetRS activities. Transcriptional fusion of a destabilized NLL-MetRS to the stringently regulated ribosomal RNA promoter demonstrated the utility of instability in accurately relating promoter activity to protein synthesis following a phenotypic switch.

3.3 Results

Mutant ssrA tags destabilize NLL-MetRS

To determine whether the abundance of NLL-MetRS could be controlled with active degradation, we fused previously identified variants of the C-terminal ssrA degradation signal to a 3xFLAG-tagged NLL-MetRS.[10,11] The final three amino acids in the ssrA tag (LAA) mediate substrate recognition by the AAA+ protease ClpXP, and we will refer to each mutant ssrA tag by its ClpX recognition motif hereafter.[12] Whereas mutation of the wildtype LAA recognition motif attenuates the interaction of a tagged substrate with ClpX, as is the case for the LDD control tag that abolishes ClpXP-mediated degradation, tandem duplication of the NY residue pair accelerates substrate recognition and degradation.[13] Degradation tags containing this duplication will contain an additional '+2' in their name.

We inserted the 3xFLAG-NLL-MetRS-ssrA coding sequence into the pBAD33 multiple cloning site for transcriptional fusion to the L-arabinose–responsive *araBAD* promoter, which enables inducible control over the expression cassette with fast kinetics of activation and repression (Figure 3.1A).[14] *E. coli* DH10B cells bearing these expression plasmids were grown to exponential phase in LB and treated with 0.2% L-arabinose for 1 h to induce expression, which was then terminated by treatment with 100 μg/mL gentamicin. Lysates harvested at each chase time were probed by immunoblotting with anti-FLAG antibody to screen for ssrA-mediated destabilization (Figure 3.1B). Several mutant ssrA tags reduced the apparent 3xFLAG-NLL-MetRS abundance at later timepoints.

Destabilization accelerates dynamics of NLL-MetRS abundance and activity

We then evaluated the dynamics of Anl incorporation by a NLL-MetRS variant bearing the NAA tag through a time-series labeling experiment. Cells were grown to early exponential phase in M9 glycerol medium, induced for expression with arabinose, and treated with 1 mM Anl to initiate labeling. After 15 min of expression, we centrifuged the cell cultures, decanted culture supernatants, and resuspended the cell pellets in M9 glycerol medium containing 1 mM Anl and 0.3% D-fucose, a nonmetabolizable sugar that facilitates *araBAD* promoter repression.[15] Cells were then removed from the culture every 30 min and lysed. Labeled proteins were conjugated to TAMRA-alkyne and resolved by mass with SDS-PAGE for in-gel fluorescence detection (Figure 3.1C). As an additional control, Aha was substituted for Anl in a culture of cells haboring an empty vector. Separately, lysates were also

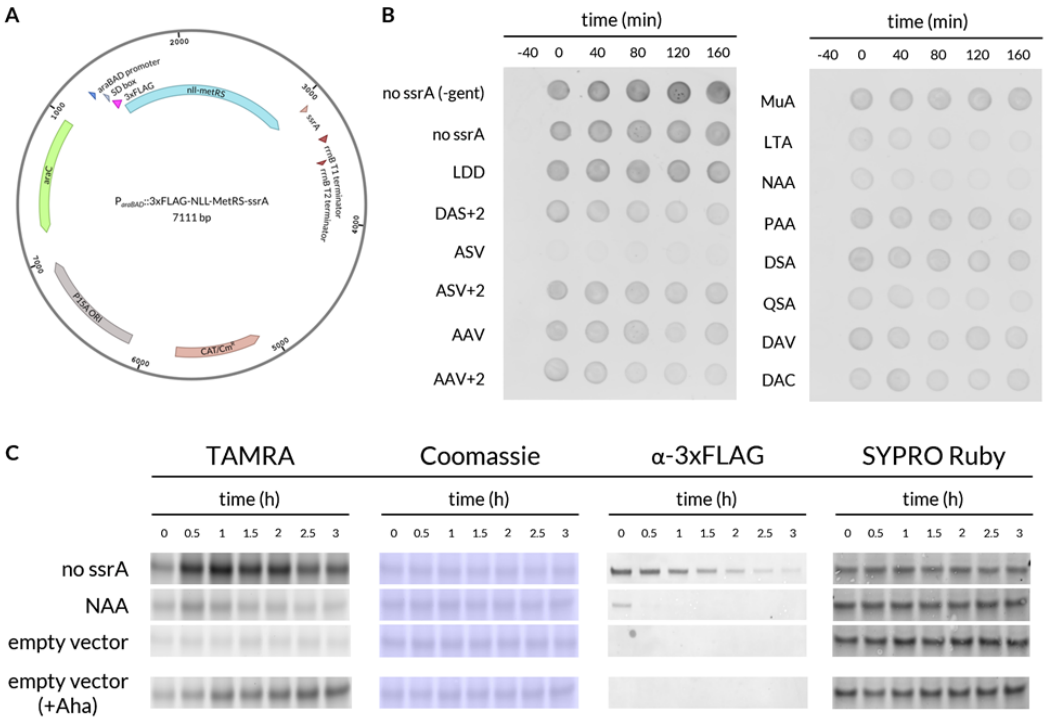

Figure 3.1: **Screening of mutant ssrA tags for destabilization of NLL-MetRS activity and abundance.** (A) Plasmid map for 3xFLAG-NLL-MetRS-ssrA fusions inserted into the pBAD33 multiple cloning site. (B) Spot blot for 3xFLAG-NLL-MetRS-ssrA variant abundance during a gentamicin chase following 1 h of arabinose-induced overexpression. (C) In-gel fluorescence detection and immunoblotting for Anl incorporation and NLL-MetRS variant abundances in time-series labeling. Coomassie and SYPRO Ruby serve as loading controls for the TAMRA and 3xFLAG signals, respectively.

evaluated for NLL-MetRS variant abundance by immunoblotting. Chronological decreases in both TAMRA intensity and NLL-MetRS abundance were accelerated by ssrA destabilization relative to the stable NLL-MetRS variant. These results suggest active degradation of NLL-MetRS could be used to accelerate the decay timescale of Anl incorporation beyond that provided by cell division.

Fluorescent fusions of destabilized NLL-MetRS enable live cell analysis

To assess expression levels and degradation kinetics of ssrA-tagged NLL-MetRS variants in single cells, we fused GFPmut3, a rapidly maturing mutant of the *Aequorea victoria* GFP, to the N terminus of each destabilized NLL-MetRS variant.[16] We also introduced a forward-designed ribosome binding site upstream of the GFP-NLL-MetRS-ssrA start codon to enhance translation of the fusion.[17] Exponential-

phase *E. coli* MG1655 cells transformed with these plasmids were induced for expression of the fusion proteins and then subjected to a protein synthesis inhibition treatment with spectinomycin. Cells were spotted on LB agarose pads immediately after spectinomycin treatment to monitor for loss and heterogeneity of fluorescence with time-lapse fluorescence microscopy (Figure 3.2A).

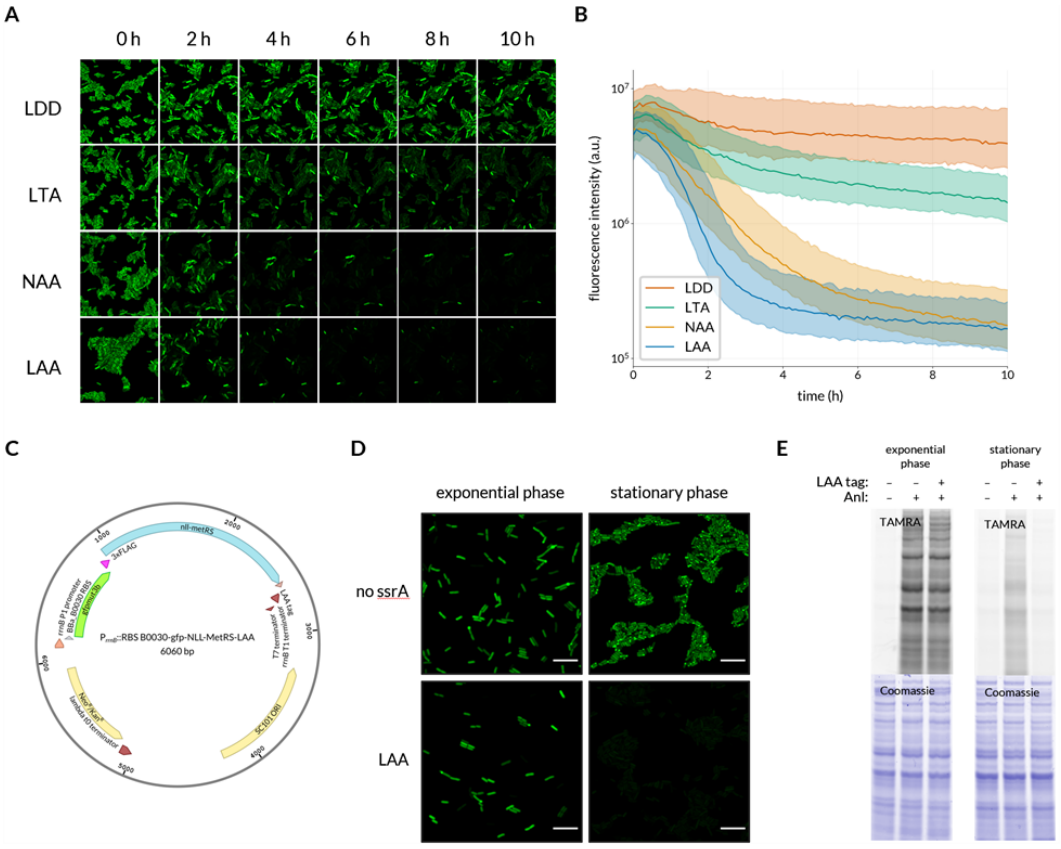

Figure 3.2: **GFP-NLL-MetRS-ssrA degradation profiles and fusion of GFP-NLL-MetRS-LAA with the *rrnB* P1 promoter.** (A) Film strip of GFP-NLL-MetRS-ssrA degradation in single cells. (B) Fluorescence distribution traces of single-cell fluorescence intensities. Shaded areas depict the interquartile range, and solid lines represent the mean. (C) Genetic circuit for *rrnB* P1 promoter fusions with destabilized fusion NLL-metRS variants. (D) Fluorescence micrographs of cells grown in M9 glucose medium expressing *rrnB* P1 promoter fusions in exponential or stationary phase. (E) In-gel fluorescence detection of Anl incorporation in cells expressing *rrnB* P1 promoter fusions in exponential or stationary phase.

Cell segmentation and fluorescence quantification revealed a 16- and 20-fold reduction in the mean cell fluorescence intensity by the end of the observation period for the two strongest degradation tags, NAA and LAA, respectively (Figure 3.2B).

Destabilization of NLL-MetRS improves resolution of labeling at growth arrest

To leverage the improved temporal resolution endowed by active degradation, we transcriptionally fused the GFP-NLL-MetRS-LAA coding sequence with the *rrnB* P1 promoter, which reports ribosome abundance and cellular NTP levels, and inserted the cassette into a low-copy vector encoding the SC101 origin of replication (Figure 3.2C).[18,19] Cells grown for 24 h in M9 glucose medium were diluted in fresh M9 glucose medium. The freshly diluted cells and the remaining overnight culture cells were incubated for an additional 2 h, after which cells were spotted on M9 glucose agarose pads for confocal fluorescence microscopy (Figure 3.2D). Owing to the strength of the *rrnB* P1 promoter, expression of GFP-NLL-MetRS resulted in a high stationary phase abundance of the reporter, whereas the GFP-NLL-MetRS-LAA variant was efficiently depleted from the cytosol, matching the expected activity of the *rrnB* P1 promoter. We repeated this growth protocol to obtain exponential and stationary phase cells expressing either GFP-NLL-MetRS variant, and cells were then treated with 1 mM Anl for 1 h. Lysates from these cultures were probed for Anl incorporation through conjugation of TAMRA-alkyne to labeled proteins followed by in-gel fluorescence detection (Figure 3.2E). Consistent with the fluorescence micrographs, we observed a modest ssrA-dependent decrease in TAMRA intensity for lysates harvested from exponential phase cells, while no TAMRA intensity above background is observed in stationary phase cells expressing GFP-NLL-MetRS-LAA. These results suggest that NLL-MetRS destabilization enables accurate discrimination between proliferating and quiescent cells with Anl metabolic labeling under these conditions.

3.4 Discussion

Our results indicate NLL-MetRS is amenable to ssrA-mediated destabilization, and the imposed instability can be moderated by mutation in the ClpX recognition motif of the appended ssrA tag. We found that the rank order of instability invoked by the ssrA tags evaluated in this study did not always correspond directly to their performance on other model substrates. As recognition and engagement of a substrate is the rate limiting step in AAA+ protease degradation, the local sequence context adjacent to the ssrA tag in NLL-MetRS may contribute to this discrepancy.

In our early studies with this approach, we found that the native ssrA tag was prohibitively destabilizing. An intuitive but especially important determinant of protein abundance when utilizing the ssrA system is translation efficiency of the tagged substrate, which we greatly improved by using a forward-designed ribosome binding site. Investigators using this strategy in future studies should place special consideration on this determinant of expression levels. Transcriptional fusions offer the opportunity to introduce artifical ribosome binding sites to address this problem while maintaining transcriptional regulation, a strategy that has been advised in instances of prohibitively low signal.[20]

Because the engineered degradation strategy employed in this work is active constitutively, temporal resolution and protein labeling are diametrically opposed. The identification of promoters that yielded sufficient transcriptional activity for the detection of protein labeling in transcriptional fusions with a destabilized NLL-MetRS was a challenging endeavour. We anticipate deployment of our rRNA reporter in mixed systems of proliferating and quiescent cells may improve cellular resolution of state-selective metabolic labeling. For example, glucose-gluconeogenic substrate shifts bifurcate clonal *E. coli* populations, resulting in dramatically imbalanced mixtures of growing and non-growing cells.[21]

3.5 Materials and Methods

Strain and plasmid construction

A fragment encoding the NLL-MetRS gene was produced by restriction enzyme digest of pBADP-NLL-MetRS with KpnI (New England Biolabs) and SalI (New England Biolabs) and ligated into pBAD33 to construct pBAD33-NLL-MetRS. Site-directed mutagenesis with phosphorylated primers was used to install C-terminal ssrA tags and an N-terminal 3xFLAG tag in pBAD33-NLL-MetRS. A fragment encoding GFPmut3b was amplified from P_{trc}:*gfp* and inserted into pBAD33-NLL-MetRS derivatives to construct N-terminal GFPmut3b fusions with Gibson assembly.[22] To construct PrrnB-GFP-NLL-MetRS, the GFP-NLL-MetRS coding sequence was inserted into a pBbS5k backbone containing the *rrnB* P1 promoter with Gibson assembly.

Media and growth conditions

M9 glycerol medium consisted of 0.2% glycerol, 1 mM $MgCl_2$, 0.1 mM $CaCl_2$, 3 g/L KH_2PO_4, 0.5 g/L NaCl, 6.78 g/L Na_2HPO_4, and 1 g/L NH_4Cl supplemented with 5 mg/L L-methionine and 40 mg/L of the remaining canonical amino acids. Kanamycin and chloramphenicol were used at 30 μg/mL and at 25 μg/mL for plasmid maintenance where appropriate. Growth conditions were 37 °C at 250 RPM in an orbital shaker unless otherwise noted.

Protein labeling

For screening labeling dynamics in exponential phase, M9 glycerol medium was inoculated with overnight culture of *E. coli* DH10B bearing an NLL-MetRS variant in a pBAD33 plasmid to an OD_{600} of 0.025 and grown to an OD_{600} of 0.2. Then, arabinose was added to 0.2% wt/vol to induce expression of NLL-MetRS, and cells were incubated with 1 mM Anl or 1 mM Aha in the control condition for 15 min to initiate protein labeling. Cells were collected by centrifugation at 5000 RCF for 5 min and resuspended in M9 glycerol medium supplemented with 1 mM Anl or 1 mM Aha in the control condition and 0.3% D-fucose, which represses expression from the *araBAD* promoter. Cells were incubated for continued growth and labeling. At the indicated time points, aliquots of cells were removed and lysed for in-gel fluorescence detection and immunoblotting analysis. Growth medium pre-warmed to 37 °C was added at regular intervals to the cultures to maintain exponential growth.

For screening expression in *rrnB* transcriptional fusions, MG1655 cells bearing the indicated construct were grown for 24 h in M9 glucose medium and then diluted 1:200 in fresh M9 glucose medium. The freshly diluted cells and the remaining overnight culture cells were incubated for an additional 2 h. Cells were then treated with 1 mM Anl for 1 h and lysed for further analysis.

For all lysis steps, cells were collected at 18000 RCF for 1 min, resuspended in 0.5% SDS in PBS, and heat treated at 95°C for 15 min. Lysates were stored at -20°C for further analysis.

In-gel fluorescence detection

For click reaction conjugation, lysates were thawed, clarified for 20 min at 18,000 RCF, and transferred to new microcentrifuge tubes. Protein concentrations were assessed with bicinchoninic assay (Thermo Scientific) and equalized with the addition of lysis buffer. Click reactions were performed as outlined in Hong et al. with 10 μM TAMRA-alkyne, 250 μM $CuSO_4$, 1.25 mM tris(3-hydroxypropyltriazolylmethyl)amine (THPTA), 5 mM (+)-sodium L-ascorbate, and 5 mM aminoguanidine hydrochloride.[23] Reactions proceeded for 1 h at 37 °C and were then mixed with 10X Reducing Agent and 4X LDS Fluorescence-Compatible Loading Dye (Thermo Fisher), heat treated at 70°C for 10 min, and resolved by SDS-PAGE on 4-12% Bis-Tris NuPAGE gels (Invitrogen). Gels were destained in 10% acetic acid, 40% methanol, and 50% deionized water overnight. The next day, gels were rehydrated in deionized water and imaged on a Typhoon Trio (GE). Gels were then incubated in InstantBlue Coomassie Protein Stain (Expedeon) for at least 2 h and imaged again on a Typhoon Trio (GE).

Immunoblotting

Protein lysates were thawed, clarified for 20 min at 18,000 RCF, and transferred to new microcentrifuge tubes. Protein concentrations were assessed with bicinchoninic assay (Thermo Scientific) and equalized with the addition of lysis buffer.

For spot blotting, M9 glycerol medium was inoculated with overnight culture of *E. coli* DH10B bearing an NLL-MetRS variant in a pBAD33 plasmid to an OD_{600} of 0.025 and grown to an OD_{600} of 0.2. Then, arabinose was added to 0.2% wt/vol for 1 h, and cells were treated with 100 μg/mL gentamicin to inhibit protein synthesis. Cells were collected at 5000 RCF and lysed at the indicated time points. 2 μL of lysate was pipetted directly onto a 0.2-μm nitrocellulose membrane (Amersham)

and allowed to dry for 15 min.

For western blotting, lysates were mixed with 10X Reducing Agent and 4X LDS Fluorescence-Compatible Loading Dye (Thermo Fisher), heat treated at 70°C for 10 min, and resolved by SDS-PAGE on 4-12% Bis-Tris NuPAGE gels (Invitrogen). Proteins were transferred to nitrocellulose membranes (0.2-µm pore size, Thermo Fisher Scientific) with the iBlot 2 Dry Blotting System (Invitrogen). Protein loading was assessed by staining with SYPRO Ruby (Thermo Fisher) per manufacturer instructions and imaged on a Typhoon Trio (GE).

Membranes were blocked with 5% milk in PBST for 1 h and treated with anti-FLAG M2 antibody (Millipore Sigma) diluted 1:5000 overnight at 4°C. Membranes were washed in four consecutive rounds with PBST for 5 min each, treated with Goat anti-Mouse Alexa Fluor 647 antibody (Invitrogen #A32728) diluted 1:10000 for 1.5 h, washed again in four consecutive rounds with PBST for 5 min each, and then imaged on a Typhoon Trio (GE).

Confocal fluorescence microscopy

For screening the stability of GFP-NLL-MetRS variants with time-lapse fluorescence microscopy, LB was inoculated with an overnight culture of *E. coli* MG1655 bearing an NLL-MetRS variant in a pBAD33 plasmid to an OD_{600} of 0.01 and cultured for 1 h. Then, arabinose was added to 0.2% wt/vol, and cells were incubated for 2 h to induce expression. Cells were then treated with 500 µg/mL spectinomycin to terminate expression, after which cells were spotted on LB agarose pads supplemented with 500 µg/mL spectinomycin for time-lapse confocal fluorescence microscopy. Cells were imaged on an LSM 800 with Airyscan (ZEISS) at the Beckman Imaging Facility at the Beckman Institute at Caltech using a 100×/1.4 NA Ph3 objective. The chamber temperature was set to 37 °C, and images were acquired at 5 min intervals. Fluorescence and phase contrast micrographs were segmented with SuperSegger.[24]

For screening expression in *rrnB* transcriptional fusions, MG1655 cells bearing the indicated construct were grown for 24 h in M9 glucose medium and then diluted 1:200 in fresh M9 glucose medium. The freshly diluted cells and the remaining overnight culture cells were incubated for an additional 2 h, after which cells were spotted on M9 glucose agarose pads for confocal fluorescence microscopy. Cells were imaged on an LSM 800 with Airyscan (ZEISS) at the Beckman Imaging Facility at the Beckman Institute at Caltech using a 100×/1.4 NA Ph3 objective.

Strains

Table 3.1: Strains used in this study

Strain	Description	Reference
DH10B	*Escherichia coli* K-12 derivative	[5]
MG1655	*Escherichia coli* K-12 derivative	[25]

Plasmids

Table 3.2: Plasmids used in this study

Plasmid	Description	Reference
pBADP-NLL-MetRS	Vector encoding NLL-MetRS	[26]
pBAD33	Empty vector for arabinose-inducible expression	[14]
pBAD33-NLL	Expression vector for NLL-MetRS	This work
pBAD33-NLL-LAA	Expression vector for 3xFLAG-NLL-MetRS with LAA tag	This work
pBAD33-NLL-LDD	Expression vector for 3xFLAG-NLL-MetRS with LDD tag	This work
pBAD33-NLL-DAS+2	Expression vector for 3xFLAG-NLL-MetRS with DAS+2 tag	This work
pBAD33-NLL-ASV	Expression vector for 3xFLAG-NLL-MetRS with ASV tag	This work
pBAD33-NLL-ASV+2	Expression vector for 3xFLAG-NLL-MetRS with ASV+2 tag	This work
pBAD33-NLL-AAV	Expression vector for 3xFLAG-NLL-MetRS with AAV tag	This work

Table 3.2 continued from previous page

Plasmid	Description	Reference
pBAD33-NLL-AAV+2	Expression vector for 3xFLAG-NLL-MetRS with AAV+2 tag	This work
pBAD33-NLL-MuA	Expression vector for 3xFLAG-NLL-MetRS with MuA tag	This work
pBAD33-NLL-LTA	Expression vector for 3xFLAG-NLL-MetRS with LTA tag	This work
pBAD33-NLL-NAA	Expression vector for 3xFLAG-NLL-MetRS with NAA tag	This work
pBAD33-NLL-PAA	Expression vector for 3xFLAG-NLL-MetRS with PAA tag	This work
pBAD33-NLL-DSA	Expression vector for 3xFLAG-NLL-MetRS with DSA tag	This work
pBAD33-NLL-QSA	Expression vector for 3xFLAG-NLL-MetRS with QSA tag	This work
pBAD33-NLL-DAV	Expression vector for 3xFLAG-NLL-MetRS with DAV tag	This work
pBAD33-NLL-DAC	Expression vector for 3xFLAG-NLL-MetRS with DAC tag	This work
pBAD33-dRBS-GFP-NLL	Expression vector for GFPmut3b-3xFLAG-NLL-MetRS	This work
pBAD33-dRBS-GFP-NLL-LAA	Expression vector for GFPmut3b-3xFLAG-NLL-MetRS with LAA tag	This work

Table 3.2 continued from previous page

Plasmid	Description	Reference
pBAD33-dRBS-GFP-NLL-LTA	Expression vector for GFPmut3b-3xFLAG-NLL-MetRS with LTA tag	This work
pBAD33-dRBS-GFP-NLL-NAA	Expression vector for GFPmut3b-3xFLAG-NLL-MetRS with NAA tag	This work
pBAD33-dRBS-GFP-NLL-LDD	Expression vector for GFPmut3b-3xFLAG-NLL-MetRS with LDD tag	This work
pBbS5k	Low-copy vector with SC101 origin of replication	[19]
P_{trc}:*gfp*	Vector encoding GFPmut3b	[26]
PrrnB-RBS-B0030-GFP-NLL	pBbS5k-derived expression vector for GFPmut3b-3xFLAG-NLL-MetRS transcriptionally fused to rrnB promoter and a strong ribosome binding site (BBa_B0030)	This work
PrrnB-RBS-B0030-GFP-NLL-LAA	pBbS5k-derived expression vector for GFPmut3b-3xFLAG-NLL-MetRS with LAA tag transcriptionally fused to rrnB promoter and a strong ribosome binding site (BBa_B0030)	This work

References

(1) King, O. D.; Masel, J. The evolution of bet-hedging adaptations to rare scenarios. *Theor Popul Biol* **2007**, *72*, 560–575.

(2) Carey, J. N.; Mettert, E. L.; Roggiani, M.; Myers, K. S.; Kiley, P. J.; Goulian, M. Regulated Stochasticity in a Bacterial Signaling Network Permits Tolerance to a Rapid Environmental Change. *Cell* **2018**, *173*, 196–207.

(3) Balaban, N. Q. et al. Definitions and guidelines for research on antibiotic persistence. *Nat Rev Microbiol* **2019**, *17*, 441–448.

(4) Balaban, N. Q.; Merrin, J.; Chait, R.; Kowalik, L.; Leibler, S. Bacterial persistence as a phenotypic switch. *Science* **2004**, *305*, 1622–1625.

(5) Ngo, J. T.; Babin, B. M.; Champion, J. A.; Schuman, E. M.; Tirrell, D. A. State-selective metabolic labeling of cellular proteins. *ACS Chem Biol* **2012**, *7*, 1326–1330.

(6) Keiler, K. C.; Waller, P. R.; Sauer, R. T. Role of a peptide tagging system in degradation of proteins synthesized from damaged messenger RNA. *Science* **1996**, *271*, 990–993.

(7) Gottesman, S.; Roche, E.; Zhou, Y.; Sauer, R. T. The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. *Genes Dev* **1998**, *12*, 1338–1347.

(8) Andersen, J. B.; Sternberg, C.; Poulsen, L. K.; Bjorn, S. P.; Givskov, M.; Molin, S. New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria. *Appl Environ Microbiol* **1998**, *64*, 2240–2246.

(9) Potvin-Trottier, L.; Lord, N. D.; Vinnicombe, G.; Paulsson, J. Synchronous long-term oscillations in a synthetic gene circuit. *Nature* **2016**, *538*, 514–517.

(10) Flynn, J. M.; Neher, S. B.; Kim, Y. I.; Sauer, R. T.; Baker, T. A. Proteomic discovery of cellular substrates of the ClpXP protease reveals five classes of ClpX-recognition signals. *Mol Cell* **2003**, *11*, 671–683.

(11) McGinness, K. E.; Baker, T. A.; Sauer, R. T. Engineering controllable protein degradation. *Mol Cell* **2006**, *22*, 701–707.

(12) Flynn, J. M.; Levchenko, I.; Seidel, M.; Wickner, S. H.; Sauer, R. T.; Baker, T. A. Overlapping recognition determinants within the ssrA degradation tag allow modulation of proteolysis. *Proc Natl Acad Sci U S A* **2001**, *98*, 10584–10589.

(13) Hersch, G. L.; Baker, T. A.; Sauer, R. T. SspB delivery of substrates for ClpXP proteolysis probed by the design of improved degradation tags. *Proc Natl Acad Sci U S A* **2004**, *101*, 12136–12141.

(14) Guzman, L. M.; Belin, D.; Carson, M. J.; Beckwith, J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J Bacteriol* **1995**, *177*, 4121–4130.

(15) Beverin, S.; Sheppard, D. E.; Park, S. S. D-Fucose as a gratuitous inducer of the L-arabinose operon in strains of Escherichia coli B-r mutant in gene araC. *J Bacteriol* **1971**, *107*, 79–86.

(16) Cormack, B. P.; Valdivia, R. H.; Falkow, S. FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* **1996**, *173*, 33–38.

(17) Salis, H. M.; Mirsky, E. A.; Voigt, C. A. Automated design of synthetic ribosome binding sites to control protein expression. *Nat Biotechnol* **2009**, *27*, 946–950.

(18) Schneider, D. A.; Gaal, T.; Gourse, R. L. NTP-sensing by rRNA promoters in Escherichia coli is direct. *Proc Natl Acad Sci U S A* **2002**, *99*, 8602–8607.

(19) Lee, T. S.; Krupa, R. A.; Zhang, F.; Hajimorad, M.; Holtz, W. J.; Prasad, N.; Lee, S. K.; Keasling, J. D. BglBrick vectors and datasheets: A synthetic biology platform for gene expression. *J Biol Eng* **2011**, *5*, 12.

(20) Young, J. W.; Locke, J. C.; Altinok, A.; Rosenfeld, N.; Bacarian, T.; Swain, P. S.; Mjolsness, E.; Elowitz, M. B. Measuring single-cell gene expression dynamics in bacteria using fluorescence time-lapse microscopy. *Nat Protoc* **2011**, *7*, 80–88.

(21) Kotte, O.; Volkmer, B.; Radzikowski, J. L.; Heinemann, M. Phenotypic bistability in Escherichia coli's central carbon metabolism. *Mol Syst Biol* **2014**, *10*, 736.

(22) Gibson, D. G.; Young, L.; Chuang, R. Y.; Venter, J. C.; Hutchison, C. A.; Smith, H. O. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* **2009**, *6*, 343–345.

(23) Hong, V.; Presolski, S. I.; Ma, C.; Finn, M. G. Analysis and optimization of copper-catalyzed azide-alkyne cycloaddition for bioconjugation. *Angew Chem Int Ed Engl* **2009**, *48*, 9879–9883.

(24) Stylianidou, S.; Brennan, C.; Nissen, S. B.; Kuwada, N. J.; Wiggins, P. A. SuperSegger: robust image segmentation, analysis and lineage tracking of bacterial cells. *Mol Microbiol* **2016**, *102*, 690–700.

(25) Bagert, J. D.; Xie, Y. J.; Sweredoski, M. J.; Qi, Y.; Hess, S.; Schuman, E. M.; Tirrell, D. A. Quantitative, time-resolved proteomic analysis by combining bioorthogonal noncanonical amino acid tagging and pulsed stable isotope labeling by amino acids in cell culture. *Mol Cell Proteomics* **2014**, *13*, 1352–1358.

(26) Babin, B. M.; Atangcho, L.; van Eldijk, M. B.; Sweredoski, M. J.; Moradian, A.; Hess, S.; Tolker-Nielsen, T.; Newman, D. K.; Tirrell, D. A. Selective Proteomic Analysis of Antibiotic-Tolerant Cellular Subpopulations in Pseudomonas aeruginosa Biofilms. *mBio* **2017**, *8*, 10.1128/mbio.01593–17.

Chapter 4

TIME-RESOLVED PROTEOMIC ANALYSIS OF MAZF ACTIVATION IN *ESCHERICHIA COLI*

4.1 Abstract

The endoribonuclease MazF is a widely studied toxin across the near-ubiquitous family of toxin-antitoxin systems in bacteria, yet reports of its putative functions and integration in cellular stress response networks are inconsistent and contentious. Here, we deployed time-resolved chemoproteomics to enrich and identify proteins synthesized during MazF-mediated growth arrest. Our results confirm the reported reallocation of protein synthesis capacity towards ribosome biogenesis and additionally demonstrate induction of the cold shock response, consistent with other stresses that inhibition translation. Immunoblotting validation of the endogenously produced major cold shock RNA chaperone CspA further suggested this class of proteins is upregulated upon MazF activation. These results highlight the advantage chemoproteomic enrichment provides in the analysis of protein synthesis in quiescent systems.

4.2 Introduction

In free-living bacteria, toxin-antitoxin (TA) loci are ubiquitous genetic elements typically composed of a toxin protein and a neutralizing antitoxin.[1] When encoded on a plasmid, toxins induce cell stasis or death by disrupting critical cellular processes in the absence of an antitoxin, thereby driving the post-segregational killing (PSK) of daughter cells that fail to inherit the TA system after plasmid partitioning.[2] Antitoxins regulate the activity of their cognate toxins through a variety of transcriptional, translational, or post-translational mechanisms, and recurring combinations of toxin activity and antitoxin regulation gave rise to a classification system spanning types I-VI.[3] Whereas toxins of all type I TA loci decrease the proton-motive force, those of the mechanistically diverse types II-VI can, for example, site specifically degrade RNA, inhibit FtsZ and MreB polymerization, phosphorylate glutamyl-tRNA synthetase, or poison the DNA-gyrase complex.[4–8] While the PSK model of plasmid-borne TA systems is uncontroversial, the study of chromosomal TA loci is distinctly contentious with widely varying hypothesized roles regarding their purposes and conditions of activation. Various external stresses activate transcription of *E. coli* TA loci, suggesting they may play a general role in attenuating cellular activity under adverse conditions. However, as TA systems often entail complex post-transcriptional regulation, researchers have challenged the interpretation of such induction as an indicator of toxin activity.[9] More broadly, hypothesized roles of chromosomal TA systems include the regulation of complex phenomena such as bacterial persistence, quorum sensing–mediated programmed cell death, genetic stability, and abortive phage infection, while a conservative view may describe these systems simply as self-propagating, selfish genetic elements with little integration in the cellular stress response network.[10–12]

The most well-studied type II TA locus in *E. coli* is *mazEF*, which encodes a proteolytically unstable antitoxin, MazE, and a ribosome-independent mRNA interferase, MazF, that catalyzes endoribonucleolytic hydrolysis at ACA ribonucleotide triplet motifs in single-stranded RNA (ssRNA).[13] Upon activation, MazF expression reportedly induces growth stasis or quorum sensing–mediated cell death and upregulates at least 50 proteins associated with either cell survival or death.[14,15] Primer extension analyses returned cleavage sites in several transcripts' leader regions and, most notably, a 43-nt region in the 16S rRNA.[16] The resultant leaderless mRNAs (lmRNAs), which lack a Shine-Dalgarno sequence, were preferentially translated by specialized ribosomes containing hydrolyzed 16S rRNA lacking the

anti–Shine-Dalgarno sequence. This observation led to the provocative hypothesis that a heterogeneous ribosome population reshapes the proteome through enhanced translation of native and MazF-generated lmRNAs.

Conversely, paired-end RNA-seq and ribosome profiling suggested ectopically expressed MazF curtails translation by disrupting ribosome biogenesis, challenging the specialized ribosome model.[17] 3'-end ribosome footprint counts, which should decrease after MazF-mediated transcript cleavage, increased over uninduced levels in 100 transcripts. The authors did not identify any set of stress response proteins encoded by these transcripts, suggesting that MazF does not orchestrate a specialized reprogramming of translation.

Missing from these studies is a quantitative proteomic analysis leveraging contemporary advancements in labeling strategies, multiplexing, and proteome coverage. As protein synthesis decreases by an order of magnitude after MazF activation, protein identification is strongly biased towards preexisting proteins likely unrelated to any effect MazF may elicit, which has reduced proteome coverage in published analyses of this system.[13,18,19] To address this challenge, we deployed bioorthogonal noncanonical amino acid tagging (BONCAT) for time-resolved analysis of protein synthesis during MazF activation. The BONCAT method relies on the treatment of cells with azide-labeled methionine surrogates such as azidohomoalanine (Aha), which is accepted for aminoacylation of $tRNA^{Met}$ by the native methionyl-tRNA synthetase for subsequent incorporation into nascent proteins. Labeled proteins can be conjugated to alkyne-labeled fluorophores or resins, thereby enabling visualization or the enrichment of nascent protein from the preexisting protein pool, respectively. As BONCAT has demonstrated success in the identification of proteins synthesized during such periods of quiescence,[20,21] we hypothesized that a BONCAT analysis of *mazEF* activation would complement previous reports without relying on the inference of transcriptomic and translatomic methods, which do not always correlate well with proteomic results and have had challenging interpretations in the context of widespread RNA degradation following MazF activation.[17] Our proteome-wide analysis further supports observations that MazF elicits growth stasis by disrupting ribosome biogenesis. Consistent with this inhibition of translation, we additionally identify enrichment of the cold shock response system among proteins upregulated by or insensitive to MazF activity. We then provide direct evidence at the protein level of the increased abundance of the major cold shock RNA chaperone CspA in cells expressing MazF via immunoblotting analysis.

4.3 Results

Chromosomal MazF expression gradually induces growth stasis

We selected an established system for emulating the gradual accrual of MazF under conditions which may liberate it from antitoxin neutralization.[22] This strain, MO::*mazE-mazF*, harbors an anhydrotetracycline (aTc)–inducible chromosomal insertion of MazF and an arabinose-inducible insertion of MazE at a distal chromosomal locus (Fig. 4.1A). This strain also harbors a chromosomal IPTG-inducible mCherry expression cassette for assessing protein synthesis capacity. We induced MazF expression in exponential phase cells growing in M9 glycerol medium and removed cells for growth measurement and for Aha labeling and subsequent lysis at regular intervals to assess protein synthesis by in-gel fluorescence of TAMRA alkyne–conjugated proteins. Induction of MazF elicited a gradual decline in growth that culminated in complete arrest after 2 h (Fig. 4.1B). Consistent with this decline, protein synthesis fell to a reduced steady-state level after the same duration (Fig. 4.1C). Notably, we observed a clear relationship between protein size and synthesis consistent with a previous radiolabeling analysis, with small proteins experiencing limited perturbation in synthesis along with a gradual accumulation observed in Coomassie staining.[13] Owing to their small coding sequences, such proteins are expected to have shorter transcripts with fewer MazF recognition sites on average (Fig. S4.1). We observed a 75% reduction in normalized integrated intensity of the TAMRA signal over the course of the experiment, ending with a 3-fold increase in signal relative to the Aha-free background intensity (Fig. S4.2). We conclude that BONCAT labeling provides temporal resolution in the analysis of protein synthesis under conditions of MazF activation.

BONCAT enables enrichment and identification of proteins synthesized during MazF-mediated growth arrest

We next sought to characterize the translational response to MazF activation on a proteome-wide scale. We selected two timepoints for characterization: 0.5 h and 2 h of MazF induction for the immediate translational response and for steady-state protein synthesis, respectively (Fig. 4.2A). Cells were pulse labeled with Aha and lysed for chemoproteomic enrichment of labeled proteins with or without MazF expression at these predetermined time points. Enriched peptides were isobarically labeled with tandem mass tags and analyzed by LC-MS/MS for identification and quantification, yielding 1448 quantified proteins. Clustering of samples in principal component analysis along with reduced intergroup correlation coefficients suggested

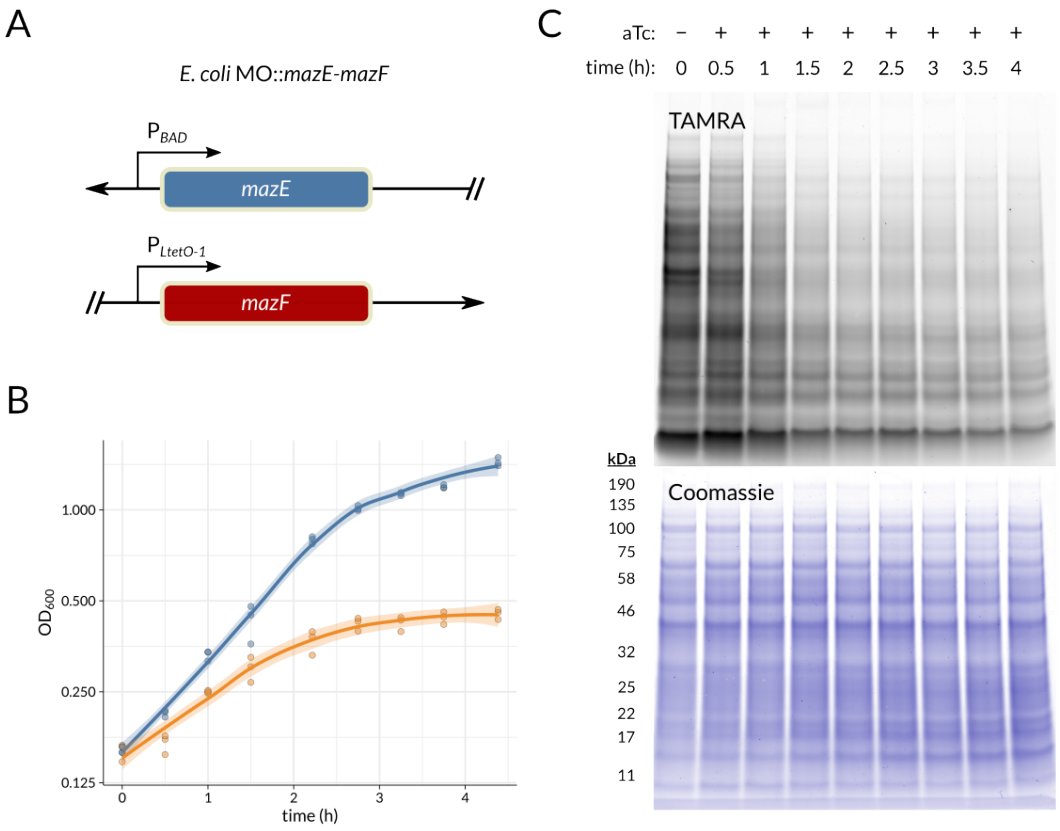

Figure 4.1: **MazF inhibition of growth and protein synthesis.** (A) Graphical summary of the expression cassettes in strain MO::*mazE-mazF*. (B) Growth of cells without (blue) or with (orange) chromosomal expression of MazF. The regression lines represent locally estimated scatterpoint smoothing (LOESS) fits with shading reporting a 95% confidence interval of the predicted values. (C) Representative in-gel fluorescence detection of Aha-labeled proteins following chromosomal expression of MazF.

MazF elicited a dramatic response in protein synthesis (Figs. 4.2B-C).

Fold change estimation yielded a median 52% and 92% decrease in enriched protein abundance relative to untreated controls at the early and late time points, respectively (Fig. 4.2D). Notably, the protein abundance fold changes in our proteomics analysis were correlated with both cleavage ratios and ribosome profiling footprint ratios reported in an integrated analysis of ectopic MazF expression, suggesting these transcriptomic, translatomic, and proteomic methods produced comparable results (Fig. 4.2D, Fig. S4.3).[17] Fold changes were most negatively correlated with protein CDS' count of ACA MazF recognition sites relative to any other nucleotide triplet (Fig. 4.2E). Further, when controlling for CDS length and using a single CDS nucleotide triplet's counts as a predictor in linear regression models of the late-

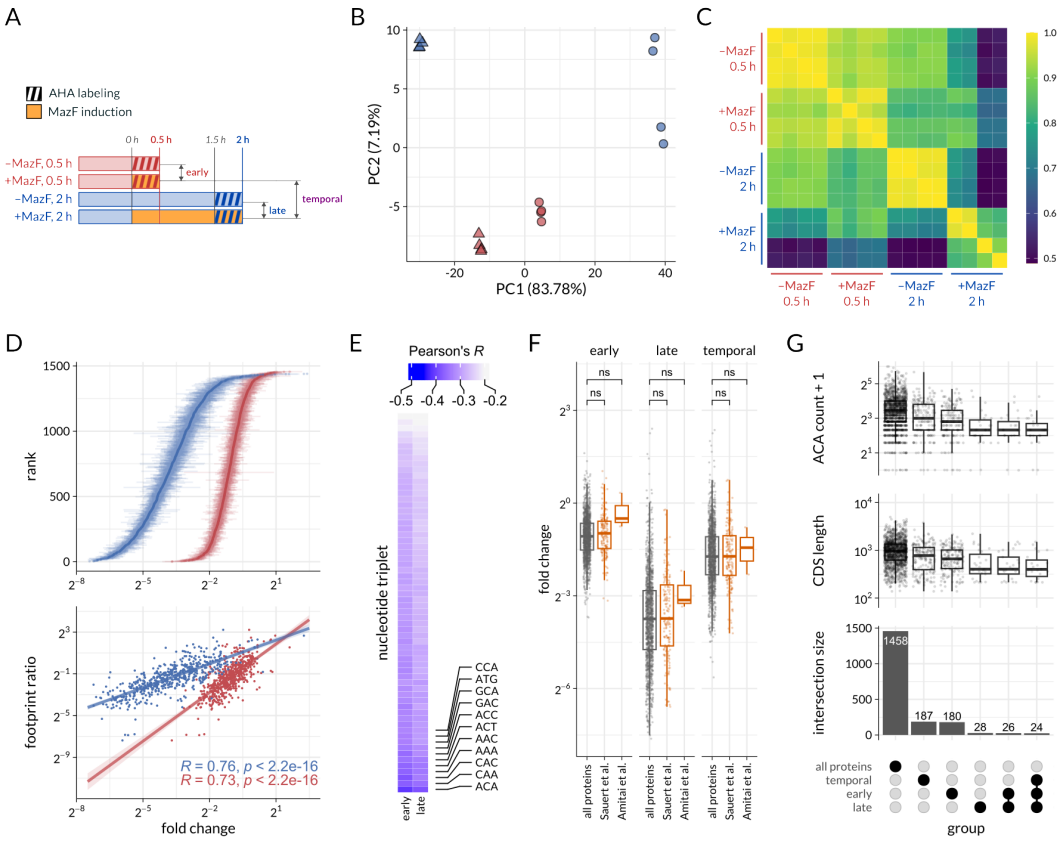

Figure 4.2: **Proteome-wide analysis of MazF activation.** (A) Experimental timeline for proteomic analysis depicting the three comaprisons in this analysis: early (0.5 h MazF expression vs. uninduced), late (2 h MazF expression vs. uninduced), and temporal (2 h MazF expression vs. 0.5 h MazF expression). Cells were pulse labeled with Aha for 30 min with or without expression of MazF for the indicated duration. (B) Principal component analysis of samples from MazF-induced (circles) and uninduced (triangles) cells at early (red) and late (blue) timepoints. (C) Pearson correlation coefficients for normalized TMT abundances across all samples. (D) Fold change distributions (top panel) and regression against ribosome footprint ratios reported in Culviner and Laub[17] (bottom panel) for early (red) and late (blue) comparisons. Error bars represent 95% confidence intervals. (E) Pearson correlation coefficients between log fold change and nucleotide triplet count for early and late comparisons. (F) Fold change distributions for all proteins, those identified as MazF regulon members, or those reported as selectively upregulated by MazF expression.[14,23] Labels indicate the results of nonparametric testing between the protein subset and its complement with correction for multiple hypothesis testing (two-sided Mann-Whitney U test). (G) UpSet plot for all identified proteins or for those lacking statistically significant evidence of downregulation upon MazF activation across all three comparisons. The dot plot indicates the cardinality and identity of the intersections.

timepoint fold changes, the ACA triplet provided the strongest association and statistical significance (95% CI: -0.083 ± 0.017 ACA^{-1}; FDR = 3.1×10^{-18}) (Table 4.1). We did not see any consistent enrichment of the reported MazF regulon or of the limited pool of mutually identified proteins reported in the inaugural 2-DE analysis of MazF expression (Fig. 4.2F).[14,23] Instead, differential expression analysis across the three comparisons demonstrated that proteins either differentially upregulated (fold change > 0, FDR-adjusted p-value < 0.05) or lacking statistically significant downregulation (fold change < 0, FDR-adjusted p-value > 0.05) were associated with smaller CDS lengths bearing an expectedly smaller number of MazF recognition sites (Fig. 4.2G). We identified a group of 24 proteins meeting these criteria across all three comparisons (Table S4.2).

MazF disrupts metabolism and induces cellular stress responses

Competitive functional enrichment analysis of proteins ranked by their fold changes yielded numerous enriched annotations among upregulated and downregulated proteins in each comparison (Fig. 4.3). Notably, terms describing ribosome biogenesis and RNA binding were enriched in MazF-expressing cells. Additional functionalities enriched in MazF-induced cells include osmoregulation-associated symporter activity and cold shock, suggesting the translational inhibition invoked by MazF may activate specialized stress responses. Functionalities associated with downregulated proteins were associated with central metabolism, iron assimilation, and motility. Comparison of the abundance fold changes at both time points to footprint fold changes in ribosome profiling from Culviner and Laub yielded mutually upregulated cold shock response proteins across both studies (Fig. S4.4). The cold shock response system was also identified as upregulated in transcriptional profiling of cells expressing MazF.[19]

MazF activates expression of the major cold shock chaperone CspA

Among the most strongly and consistently upregulated proteins following MazF expression across all comparisons was the major cold shock RNA chaperone CspA, which enhances translation by eliminating obstructive, thermodynamically favorable RNA secondary structure at low temperatures (Fig. 4.4A).[24] Of the nine paralogous cold shock proteins (CSPs) encoded in the MG1655 genome, four (CspA, CspB, CspG, and CspE) are cold inducible, and three (CspA, CspB, and CspG) were consistently upregulated or unperturbed in synthesis (Fig. S4.5).[25] Additional members of the cold shock response prioritized in protein synthesis after MazF expression

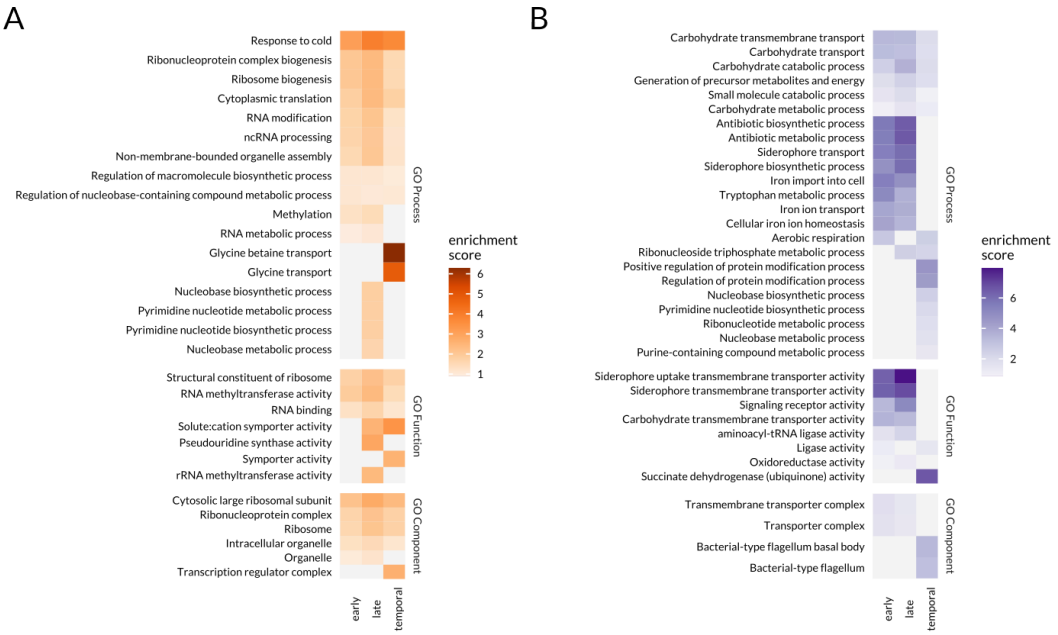

Figure 4.3: **Functional enrichment analysis of proteins across comparison groups.** (A) Enriched annotations (FDR < 0.05) across the three Gene Ontology domains among upregulated proteins across the three comparisons. (B) Enriched annotations (FDR < 0.05) across the three Gene Ontology domains among down-regulated proteins across the three comparisons.

included initiation factor IF-3 and ribosome binding and biogenesis factor RbfA. Other proteins exhibiting consistent synthesis included the tRNA chaperone TruB and the ribosome silencing factor RsfS, recently implicated in ribosome biogenesis.[26] Together, these proteins illustrate a reallocation of protein synthesis capacity towards stress response regulators and biogenesis factors dedicated to the restoration of translation upon MazF activation.

We next sought to confirm the upreguation of CspA by direct observation with immunoblotting. We inserted CspA under the control of its natural promoter and C-terminally fused to either an ACA-free FLAG or 3xFLAG epitope into a low copy vector, and we introduced these plasmids to both MO::*mazE-mazF* and MO::*mazE*. Cells bearing these plasmids were grown to exponential phase and treated with aTc. Installation of the 3xFLAG epitope provided adequate sensitivity to observe a gradual accumulation of CspA-3xFLAG in cells expressing MazF (Fig. 4.4B), whereas cells lacking MazF exhibited static and comparatively low expression. Next, to directly validate the upregulation of CspA, we inserted an ACA-free C-terminal 3xFLAG epitope at the *cspA* locus in MazF-inducible or MazF-deficient strains with a scarless Cas9 cloning system.[27] Strains were then monitored for expression

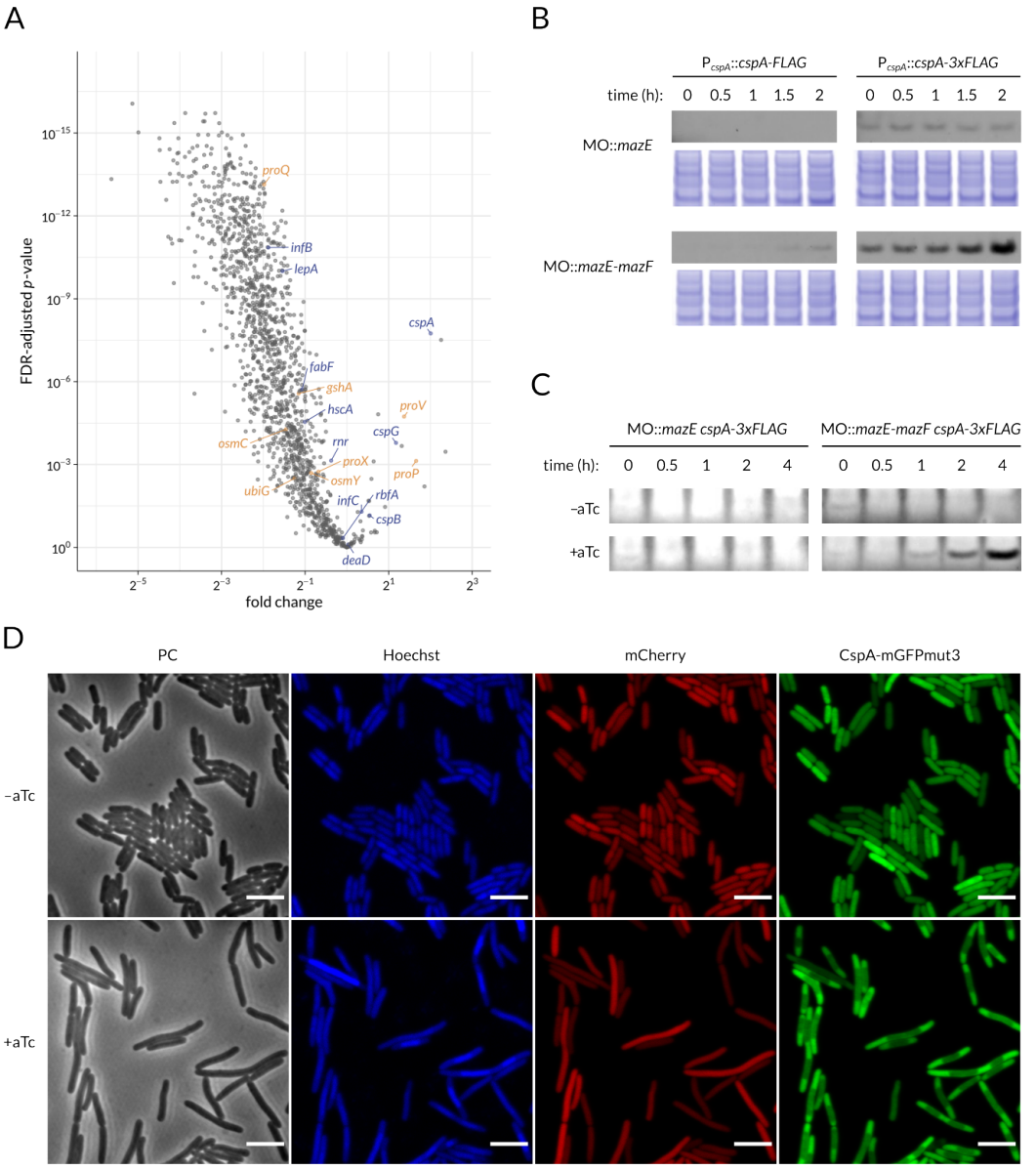

Figure 4.4: **MazF activation triggers the cold shock response.** (A) Fold changes for the temporal MazF expression comparison (2 h MazF expression vs. 0.5 h MazF expression) plotted against their FDR-adjusted *p*-values. Proteins functioning in cold shock response and hyperosmotic stress are indicated in blue and orange, respectively. (B) Immunoblotting validation of CspA-FLAG or CspA-3xFLAG expressed ectopically from a vector encoding the *cspA* transcriptional unit. (C) Immunoblotting validation of CspA-3xFLAG in strains bearing a C-terminal insertion of the 3xFLAG epitope at the *cspA* locus. (D) Confocal fluorescence microscopy of MO::*mazE-mazF cspA-mGFPmut3* cells with or without expression of MazF. Scale bar represents 5 µm.

of CspA-3xFLAG with or without induction of MazF expression. Expression of CspA increased gradually over the course of the time series in a manner exclusive to the MazF-bearing cells also induced for MazF expression (Fig. 4.4C). We also inserted an ACA-free mGFPmut3 at the C terminus of CspA and examined MazF-induced cells by confocal fluorescence microscopy (Fig. 4.4D). In this experiment, cells were additionally treated with IPTG to induce expression of mCherry for assessment of protein synthesis capacity. In agreement with electron microscopy imaging, MazF-expressing cells were elongated relative to untreated controls and presented compacted nucleoids.[28] While mCherry signal was uniform, CspA-mGFPmut3 signal was reduced in the nucleoid, suggesting CspA may primarily localize with mature RNAs in the cytoplasm under these conditions. Several nucleoid-associated proteins were among the top upregulated proteins in the immediate translational response to MazF activation (Fig. S4.6).

4.4 Discussion

To satisfy the requirement of MazF for ssRNA substrates, investigators have exploited the proficiency of recombinant CspA in unfolding structured RNA to establish recognition motifs of several MazF homologs in vitro.[29–31] However, to our knowledge, the role of endogenous CspA in mediating MazF activity has not been explored and merits further investigation. Transcriptome-wide structural profiling of RNAs under cold shock induction highlighted the concerted action of two critical acclimation actuators: Ribonuclease R, required for degradation of prohibitively structured mRNA upon cold shock, and CspA, which globally remodeled mRNA structure to enable translation.[32] In contrast with this role, CspA and other CSPs may aggravate translation upon MazF activation in light of their potentiation of endoribonuclease activity.

Subinhibitory doses of antibiotics that inhibit translation by directly targeting the ribosome elicit either the heat or cold shock responses, demonstrating a generalized role of these systems beyond their namesakes.[33] The widespread mRNA depletion and concomitant translational inhibition invoked by MazF may represent another stimulus of the cold shock response as suggested by our analysis and others.[17,19] However, while we do not observe any notable upregulation of proteins associated with the MazF-modified ribosome hypothesis, these results do not rule out the involvement of other modes of post-transcriptional regulation, directly or indirectly attributable to MazF due to its activity or the roles of the differentially expressed proteins observed here, in regulating translation. The cold shock response has an established translational bias elicited by differential abundance and recruitment of translation initiation factors, resulting in preferential translation of several cold shock proteins upregulated in our data.[34] Further, in meta-analysis of data from Culviner and Laub, we note that the CspA mRNA has a MazF recognition site coincident with a cleavage event in its 5' UTR, which destabilizes the transcript at 37 °C (Fig. S4.8).[35,36] Despite the three MazF recognition sites in the CspA coding sequence, this protein was the most consistently and dramatically upregulated across all comparisons. Whether post-transcriptional events or translational bias with commonalities to the cold shock response contribute to the differential expression seen here may be subject of further study.

While our data are not consistent with translational activation of the reported MazF regulon, several features comport with reports of a role of MazF in phage infection,

as we see upregulation of two phage shock response proteins (PspB and PspC) along with strong reduction in synthesis of outer membrane proteins such as those bearing the TonB domain that are exploited for phage attachment.[37,38] Functional enrichment in our meta-analysis of ribosome profiling data from Culviner and Laub also highlights the recruitment of ribosomes to phage shock– and cold shock–related coding sequences (Fig. S4.7).

Studies of the MazF toxin have placed strong emphasis on proteins upregulated in absolute abundance, and very few proteins meet this criterion in our analysis. However, we emphasize that reallocation of protein synthesis is physiologically meaningful in this and other stress response contexts, which frequently involve growth arrest and a concomitant global reduction of protein synthesis. In this light, our results indicate BONCAT provides sufficient sensitivity for profiling protein synthesis in such systems and may be useful in characterizing other toxins of the widespread toxin-antitoxin family.

4.5 Acknowledgements

E.J.M. was supported in part by a National Science Graduate Research Fellowship. This work was supported by the Caltech Center for Environmental Microbial Interactions.

4.6 Materials and Methods

Strains and plasmid construction

For constructing P_{cspA}-cspA-FLAG and P_{cspA}-cspA-3xFLAG, the *cspA* transcription unit was amplified from the MG1655 chromosome and inserted into a pBbS5k backbone with Gibson assembly. Side-directed mutagenesis with phosphorylated primers was then used to insert the FLAG or 3xFLAG epitope.

For chromosomal modification of the *cspA* locus, mutation and epitope tag insertion with the Cas9 system from Jiang et al. was performed as described.[27] The N20 sequence for the single guide RNA targeting the *cspA* locus was screened for specificity with the Cas-OFFinder tool was inserted into pTarget with site-directed mutagenesis to construct pTarget-cspA(N20).[39] The donor DNA for constructing CspA-3xFLAG and CspA-mGFPmut3 insertion mutants was amplified from P_{cspA}-cspA-3xFLAG and inserted into pTarget-cspA(N20). Insertion and sequence fidelity were verified by PCR and Sanger sequencing. A DNA fragment encoding an ACA-free mGFP-mut3 for Gibson assembly was directly synthesized (Integrated DNA Technologies) and substituted for the 3xFLAG sequence in pTarget cspA-3xFLAG using Gibson assembly to construct pTarget cspA-mGFPmut3.

Media and growth conditions

M9 glycerol medium consisted of 0.2% glycerol, 1 mM $MgCl_2$, 0.1 mM $CaCl_2$, 3 g/L KH_2PO_4, 0.5 g/L NaCl, 6.78 g/L Na_2HPO_4, and 1 g/L NH_4Cl supplemented with 5 mg/L L-methionine and 40 mg/L of the remaining canonical amino acids. Kanamycin and spectinomycin were used at 30 µg/mL and 50 µg/mL, respectively, for plasmid maintenance where appropriate. Growth conditions were 37 °C at 250 RPM in an orbital shaker unless otherwise noted.

Protein labeling

M9 glycerol medium was inoculated with an overnight culture of MO::*mazE-mazF* to an OD_{600} of 0.001. When OD_{600} reached 0.2, two 1.5 mL aliquots of cells were apportioned in microcentrifuge tubes, and each was either treated with 1 mM Aha or left untreated for 30 min shaking at 800 RPM and 37 °C on a thermoshaker. Cultures were then treated with 300 ng/mL anhydrotetracycline to induce MazF expression, and 1.5 mL aliquots of cells were removed at 30 min intervals, treated with 1 mM Aha or left untreated for 30 min at 37 °C and 800 RPM, and lysed. For lysis, cells were collected at 18000 RCF for 1 min, resuspended in 0.5% SDS in

PBS for lysis, and heat treated at 95°C for 15 min. Lysates were stored at -20°C for further analysis.

In-gel fluorescence detection

For click reaction conjugation, protein lysates were thawed and clarified for 20 min at 18,000 RCF. Protein concentrations were assessed with bicinchoninic assay (Thermo Scientific) and equalized with the addition of lysis buffer. Click reactions were performed as outlined in Hong et al. with 10 μM TAMRA-alkyne, 250 μM $CuSO_4$, 1.25 mM tris(3-hydroxypropyltriazolylmethyl)amine (THPTA), 5 mM (+)-sodium L-ascorbate, and 5 mM aminoguanidine hydrochloride.[40] Reactions proceeded for 1 h at 37 °C and were then mixed with 10X Reducing Agent and 4X LDS Fluorescence-Compatible Loading Dye (Thermo Fisher), heat treated at 70°C for 10 min, and resolved by SDS-PAGE on 4-12% Bis-Tris NuPAGE gels (Invitrogen). Gels were destained in 10% acetic acid, 40% methanol, and 50% deionized water overnight. The next day, gels were rehydrated in deionized water and imaged on a Typhoon Trio (GE). Gels were then incubated in InstantBlue Coomassie Protein Stain (Expedeon) for at least 2 h and imaged again on a Typhoon Trio (GE).

Proteomic analysis

Replicate cultures in M9 glycerol medium were inoculated with overnight cultures of MO::*mazE-mazF* to an OD_{600} of 0.001. When OD_{600} reached 0.2, cultures in the MazF-induced condition were treated with 300 ng/mL aTc to induce MazF expression. Then, cultures were treated with 1 mM Aha for 30 min for the early time point or incubated for an additional 1.5 h and treated with 1 mM Aha for 30 min for the late time point. After labeling, cultures were centrifuged at 5000 RCF for 10 min and then resuspended in 4% SDS in 100 mM Tris (pH 8.0) supplemented with cOmplete EDTA-free protease inhibitor for lysis. Lysates were heat treated at 95°C for 10 min and frozen at -80°C for further processing.

Thawed lysates were probe sonicated for 30 s at 20% amplitude (QSonica) with 2 s pulses and 1 s rests to shear chromosomal DNA. Lysates were then clarified by microcentrifugation at 18000 RCF for 20 min and transferred to new microcentrifuge tubes. Protein concentrations were assessed with bicinchoninic assay (Thermo Scientific) and equalized across samples with the addition of lysis buffer.

Chemoproteomic enrichment was performed as described in Glenn et al.[41] Lysates were brought to 500 μL by the addition of 1% SDS in PBS and then treated with 100

µL of 600 mM chloroacetamide in 0.8% SDS in PBS at 65°C for 30 min in the dark at 1200 RPM on a thermoshaker. For each replicate, 80 µL of dibenzocyclooctyne-agarose (DBCO-agarose) beads (Click Chemistry Tools) was washed thrice in 0.8% SDS in PBS and collected at 1500 RCF for 1 min. Lysates were then mixed with 600 µL of a solution of 8M urea and 0.85 M NaCl in PBS. Beads resuspended in 80 µL 0.8% SDS in PBS were then added to the lysates and incubated at room temperature on a rotary wheel in the dark for 24 h.

Beads were then collected at 1500 RCF for 1 min, washed with 1 mL deionized water, and collected again at 1500 RCF for 1 min. Supernatants were then discarded, and beads were resuspended in 500 µL of 5 mM dithiothreitol (DTT) (Millipore Sigma) and incubated for 15 min at 70°C shaking at 1200 RPM on a thermoshaker in the dark. Supernatants were then removed after collecting beads at 1500 RCF for 1 min, and beads were resuspended in 500 µL of 40 mM chloroacetamide (Millipore Sigma) and incubated for 30 min in the dark at room temperature on a rotary wheel.

For washing, beads were transferred to a PolyPrep gravity column (Bio-Rad) and washed ten times with 5 mL of 0.8% SDS in PBS (wash solution A), ten times with 8M urea in 100 mM Tris-Base (pH 8) (wash solution B), and ten times with 20% acetonitrile in deionized water (wash solution C). Columns were capped before addition of the second washes, and beads were incubated in the wash solution for 10 min, 30 min, and 10 min for solutions A, B, and C, respectively.

Beads were resuspended in 1.5 mL of 50 mM ammonium bicarbonate (Millipore Sigma) in a 10% acetonitrile and 90% deionized water solution, collected at 1.5k RCF for 1 min, and concentrated to 100 µL. Lys-C was added to 0.5 ng/µL for 4 h of digestion at 37 °C in the dark on a thermoshaker set to 1200 RPM. Trypsin was then added to 1 ng/µL for digestion overnight.

The next day, samples were mixed with 100 µL of 20% acetonitrile in deionized water, and beads were collected at 1500 RCF for 1 min. Supernatants were then transferred to centrifugation columns (Pierce). This process was repeated twice with an increased volume of 150 µL of 20% acetonitrile in deionized water. Peptide solutions were vacuum concentrated (SpeedVac) to dryness and resuspended in 20 µL of 0.2% formic acid in mass spectrometry-grade water. Samples were then desalted with C18 ZipTips (EMD Millipore) per manufacturers instructions and submitted for TMT labeling and LC-MS/MS analysis.

LC-MS/MS analysis

Enriched peptide samples were labelled with TMTpro (Thermo Fisher Scientific) according to the manufacturer's protocol, quenched with 5% hydroxylamine for 15min, pooled, and lyophilized to dryness. Samples were resuspended in 5% acetonitrile/0.5% TFA, desalted on Pierce C18 spin columns (Thermo Scientific #89870), and the eluted peptides dried by lyophilization. Peptides were then resuspended in 2% acetonitrile/0.2% formic acid for LC-MS analysis.

Liquid chromatography-mass spectrometry (LC-MS) analysis was carried out on an EASY-nLC 1200 (Thermo Fisher Scientific) coupled to an Orbitrap Eclipse mass spectrometer (Thermo Fisher Scientific) equipped with a Nanospray Flex ion source. 1 µg peptides were directly loaded onto an Aurora 25cm x 75µm ID, 1.6µm C18 column (Ion Opticks, Victoria, Australia) heated to 50°C. The peptides were separated with a 180 min gradient at a flow rate of 350 nL/min as follows: 2–6% Solvent B (11 min), 6-25% B (124 min), 25-40% B (45 min), 40-98% B (1 min), and held at 98% B (15 min). 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.

FAIMS was used alternating between 3 cycles at -35V, -50V, and -65V with cycle times of 1.2, 1.0, and 0.8 seconds respectively. MS1 spectra were acquired in the Orbitrap at 120K resolution with a scan range from 350-1600 m/z, an AGC target of 1e6, and a maximum injection time of 50 ms in Profile mode. Features were filtered for monoisotopic peaks with a charge state of 2-7 and minimum intensity of 2.5e4, with dynamic exclusion set to exclude features after 1 time for 45 seconds with a 5-ppm mass tolerance. HCD fragmentation was performed with fixed collision energy of 32% after quadrupole isolation of features using an isolation window of 0.5 m/z, an AGC target of 5e4, and maximum injection time of 86 ms. MS2 scans were then acquired in the Orbitrap at 50k resolution in Centroid mode with first mass fixed at 110.

Bioinformatic analysis

Analysis of LC-MS data was performed in Proteome Discoverer 2.4 (Thermo Scientific). Raw files were searched using the SequestHT search algorithm and the UniProt *E. coli* K-12 MG1655 reference proteome (UP000000625). Percolator FDR thresholds were 0.05 (relaxed) and 0.01 (strict). Trypsin was set as the cleavage enzyme, and the missed cleavage tolerance was set to 2. Dynamic modifications

were set as oxidation of methionine and N-terminal acetylation. Static modifications were set as N-terminal TMTpro modification, TMTpro modification of lysine, and carbamidomethylation of cysteine. The minimum peptide length was set to 6, and peptide FDRs were set to 0.01 (relaxed) and 0.001 (strict). The co-isolation threshold for reporter ion quantification was set to 50 with a minimum average signal to noise threshold of 10.

All normalization and differential expression analysis was conducted with the R package *limma*.[42] Protein abundances were log2-transformed and normalized with the cyclic Loess method within replicate groups. Normalized abundances were fit to a linear model for empirical Bayes–moderated *t*-testing of each contrast with Benjamini-Hochberg adjustment of *p*-values for FDR control.[43] Competitive functional enrichment analysis of protein lists ranked by fold change was performed on the STRING database server.[44] UpSet plots were rendered with the *ComplexUpSet* R package.[45] Enriched Gene Ontology terms (FDR < 0.05) within each domain were clustered and reduced with the binary cut method from the *simplifyEnrichment* R package.[46] The top 2 annotations within each enrichment score–ranked cluster were used for plotting. Differences in fold change distributions between protein subsets described in Amitai et al. and Sauert et al. and their respective complement were assessed by two-sided Mann-Whitney *U* test with *p*-values corrected for multiple hypothesis testing with the Benjamini-Hochberg procedure.[14,23,43]

For linear modeling of the fold changes, the late timepoint fold changes were used as a dependent variable, and the CDS length and nucleotide triplet count for the indicated triplet (Table S4.1) were used as the two predictor variables. Coefficients corresponding to the triplet counts were then ranked by *t*-statistic, and *p*-values were adjusted for multiple hypothesis testing with the Benjamini-Hochberg procedure.[43]

For meta analysis of paired-end RNA sequencing and ribosome profiling datasets from Culviner and Laub, FASTQ files acquired from the NCBI Short Read Archive were aligned to the MG1655 chromosome as originally reported using bowtie2. Indexing and transformation of count data were performed as described.[47]

Immunoblotting

For evaluating CspA expression and epitope sensitivity with ectopic expression from the *cspA* transcription unit, M9 glycerol was inoculated with a 1:200 dilution of an overnight culture of MO::*mazE* or MO::*mazE-mazF* bearing plasmid P_{cspA}-cspA-FLAG or P_{cspA}-cspA-3xFLAG. Cells were grown to an OD_{600} of 0.2, treated

with 300 ng/mL anhydrotetracycline, and collected at the indicated time points by centrifugation at 18000 RCF for 1 min. Cells were resuspended in 0.5% SDS in PBS for lysis and heat treated at 95°C for 15 min. Lysates were stored for further processing at -20°C.

For screening expression of CspA from the *cspA* locus, M9 glycerol was inoculated with a 1:200 dilution of an overnight culture of MO::*mazE cspA-3xFLAG* or MO::*mazE-mazF cspA-3xFLAG*. Cells were grown to an OD_{600} of 0.2, treated with 300 ng/mL anhydrotetracycline, and collected at the indicated time points by centrifugation at 18000 RCF for 1 min. Cells were resuspended in 0.5% SDS in PBS for lysis and heat treated at 95°C for 15 min. Lysates were stored for further processing at -20°C.

Protein lysates were thawed, clarified for 20 min at 18,000 RCF, and transferred to new microcentrifuge tubes. Protein concentrations were assessed with bicinchoninic assay (Thermo Scientific) and equalized with the addition of lysis buffer. Lysates were then mixed with 10X Reducing Agent and 4X LDS Fluorescence-Compatible Loading Dye (Thermo Fisher), heat treated at 70°C for 10 min, and resolved by SDS-PAGE on 4-12% Bis-Tris NuPAGE gels (Invitrogen). Proteins were transferred to nitrocellulose membranes (0.2-µm pore size, Thermo Fisher Scientific) with the iBlot 2 Dry Blotting System (Invitrogen). Protein loading was assessed by staining with SYPRO Ruby (Thermo Fisher) per manufacturer instructions and imaged on a Typhoon Trio (GE). Membranes were then blocked with 5% milk in PBST for 1 h and treated with anti-FLAG M2 antibody (Millipore Sigma) diluted 1:5000 overnight at 4°C. Membranes were washed in four consecutive rounds with PBST for 5 min each, treated with Goat anti-Mouse Alexa Fluor 647 antibody (Invitrogen #A32728) diluted 1:10000 for 1.5 h, washed again in four consecutive rounds with PBST for 5 min each, and then imaged on a Typhoon Trio (GE).

Confocal fluorescence microscopy

M9 glycerol was inoculated with 1:200 dilution from an overnight culture of MO::*mazE-mazF* cspA-mGFPmut3. Cells were grown to an OD_{600} of 0.2, treated with 300 ng/mL anhydrotetracycline for 4 h, stained with 1 µg/mL Hoechst 34580 for 20 min, and spotted on 1.5 wt% PBS agarose pads. Cells were imaged on an LSM 800 with Airyscan (ZEISS) at the Beckman Imaging Facility at the Beckman Institute at Caltech using a 100×/1.4 NA Ph3 objective.

4.7 Supplemental Information

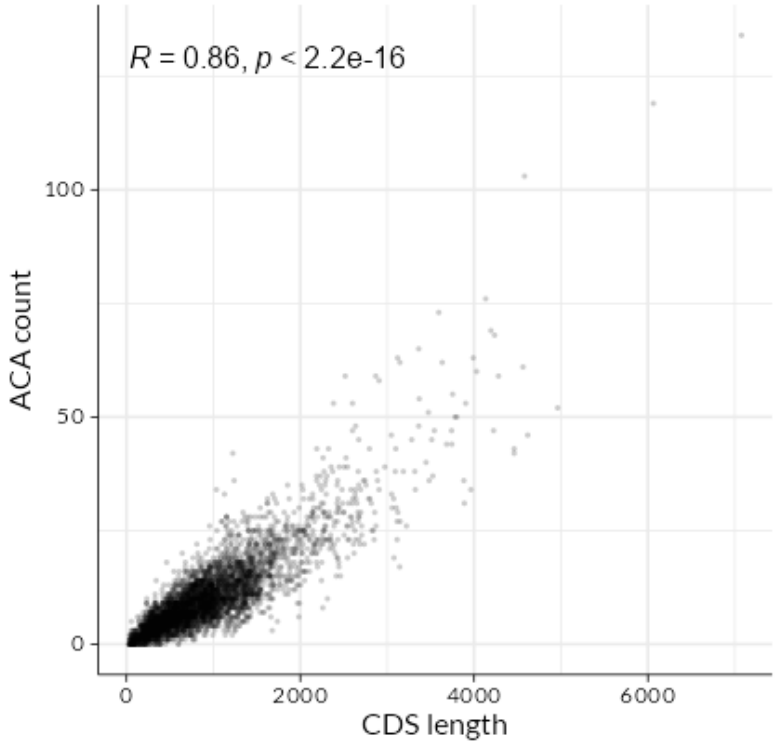

Figure S4.1: **ACA count and CDS length are strongly correlated.** Nucleotide triplets were counted for all protein-coding genes in the *E. coli* MG1655 proteome. The expected strong correlation between the total count of the ACA triplet in a protein's coding sequence and the length of the coding sequence is observed.

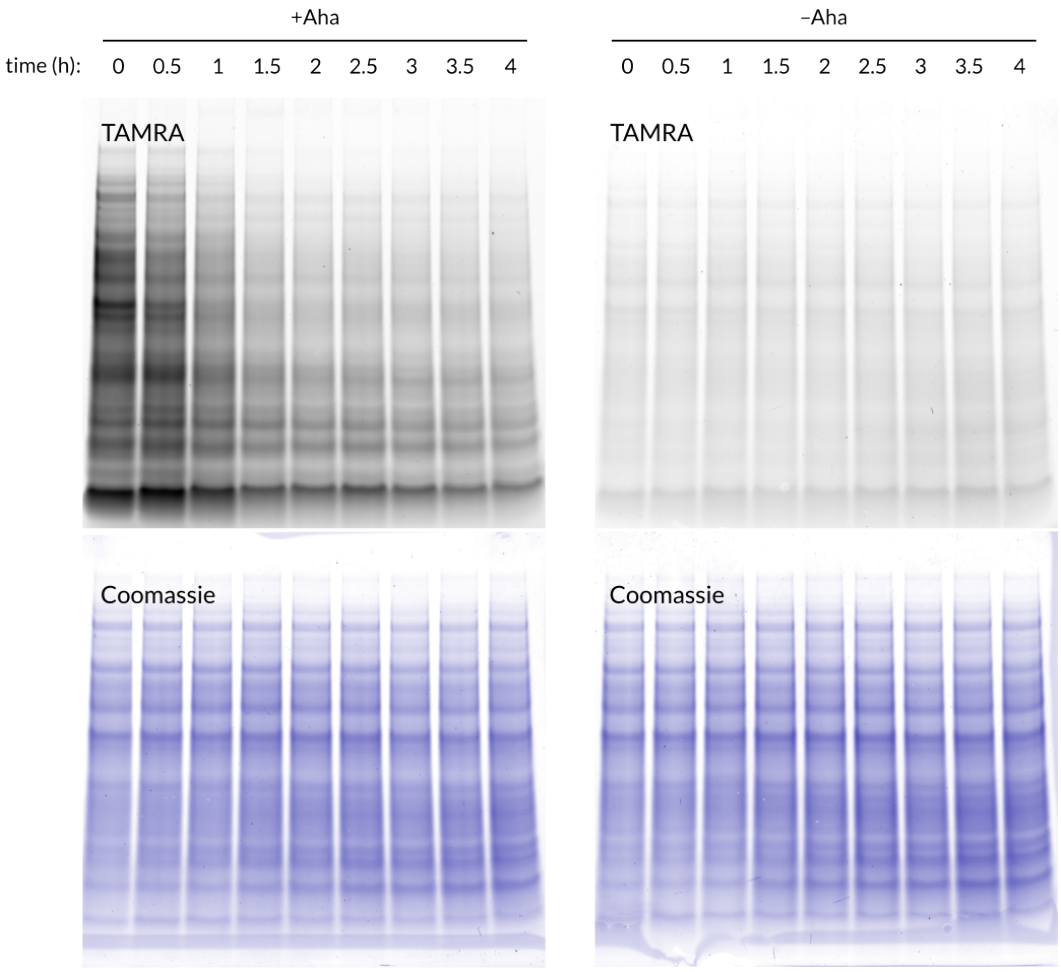

Figure S4.2: **In-gel fluorescence detection of MazF-induced cells treated or untreated with Aha.** In-gel fluorescence detection of Aha-labeled or unlabeled proteins following chromosomal expression of MazF.

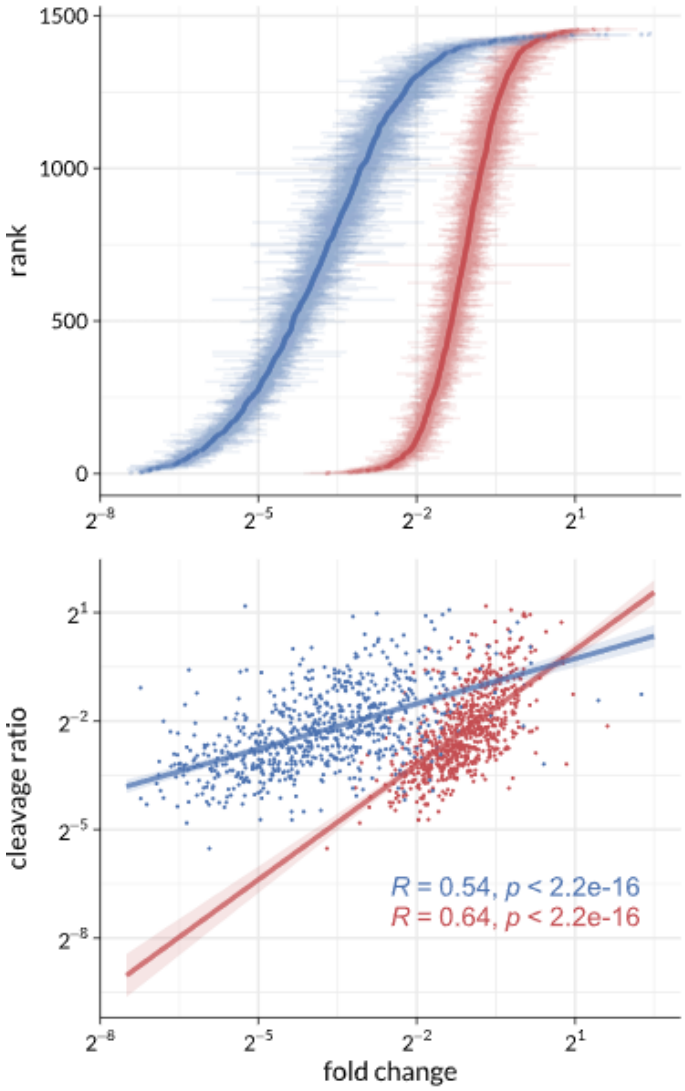

Figure S4.3: **Fold changes are correlated with RNA cleavage ratios from Culviner and Laub.** Fold changes across early (red) and late (blue) comparisons exhibit correlation with RNA cleavage ratios from paired-end RNA-seq analysis in Culviner and Laub.[17]

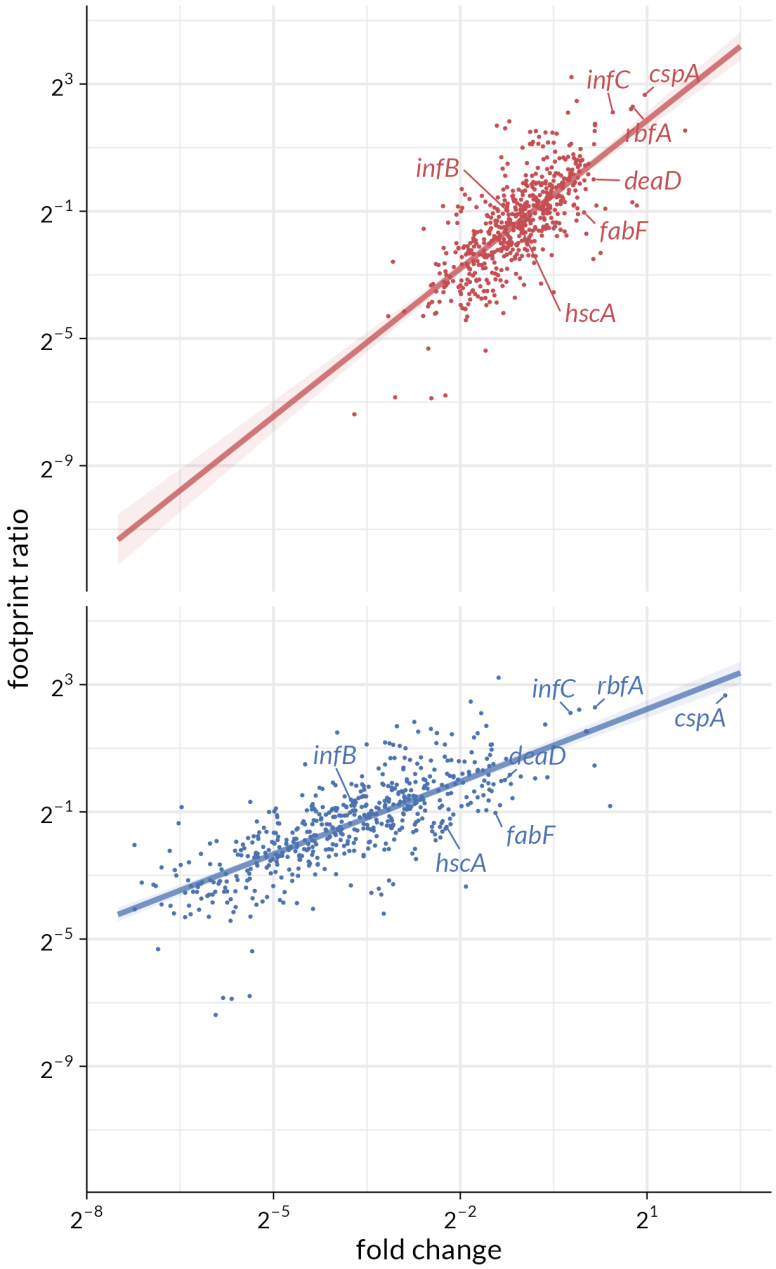

Figure S4.4: **Cold shock response proteins are upregulated across studies.** Linear regression between ribosome profiling footprint ratios from Culviner and Laub and abundance fold changes at early (red) and late (blue) time points from BONCAT-based proteomic profiling.[17] Mutually identified members of the cold shock response system are indicated.

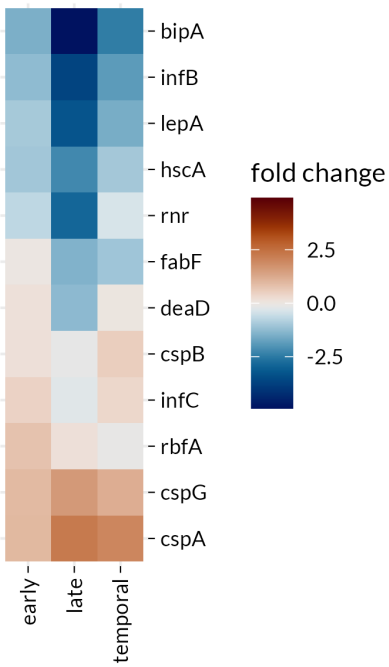

Figure S4.5: **Consistent upregulation or continuation of CSP synthesis.** Heatmap representation of the fold changes among observed cold shock regulon proteins across all three comparisons. CspA, CspB, and CspG are either upregulated or unperturbed in their synthesis.

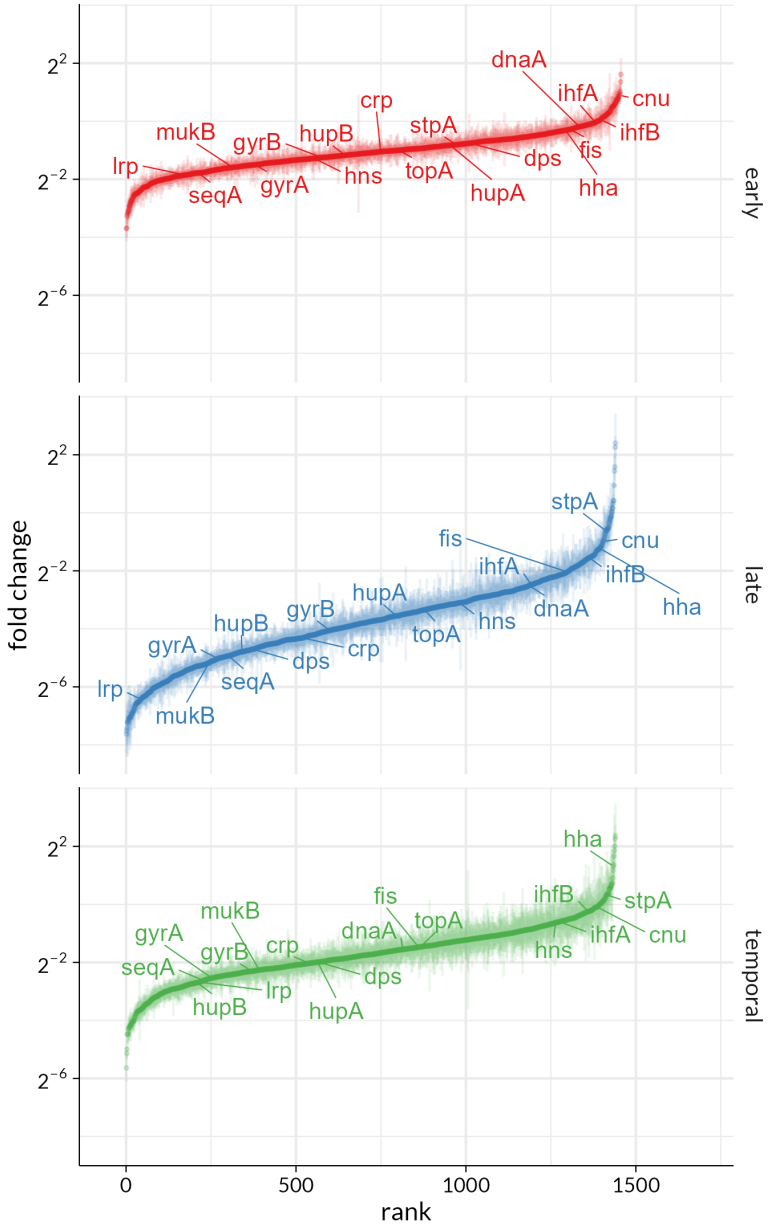

Figure S4.6: **Reallocation of protein synthesis towards nucleoid-associated proteins.** Several nucleoid-associated proteins are in the top percentiles of protein expression.

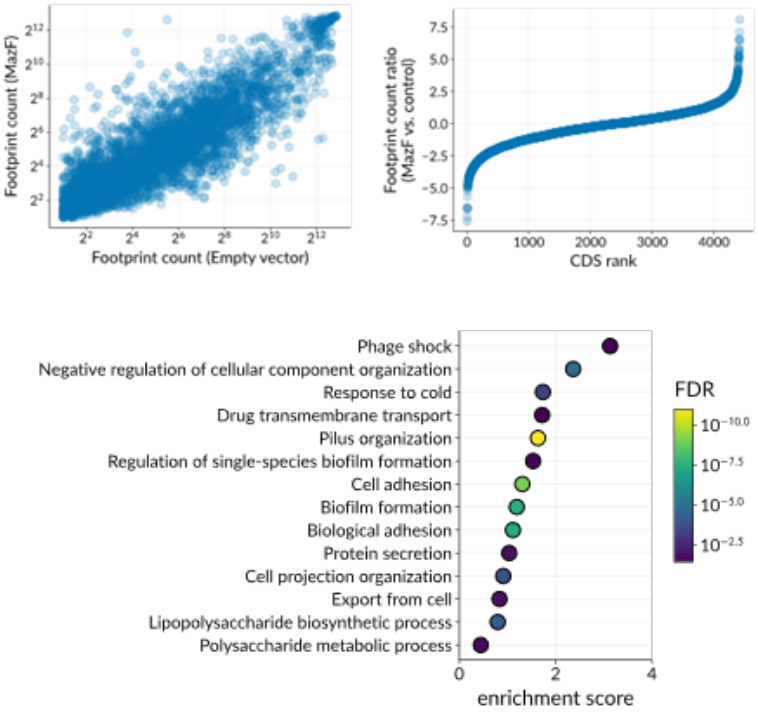

Figure S4.7: **Functional enrichment analysis of ribosome profiling footprint counts in MazF-induced cells from Culviner and Laub.** (A) Scatter plot of footprint counts between MazF-induced and empty vector controls. (B) Footprint count ratio distribution between MazF-induced and empty vector controls. (C) Competitive enrichment of proteins ranked by their footprint count ratios identifies functionalities consistent with our proteomic analysis.

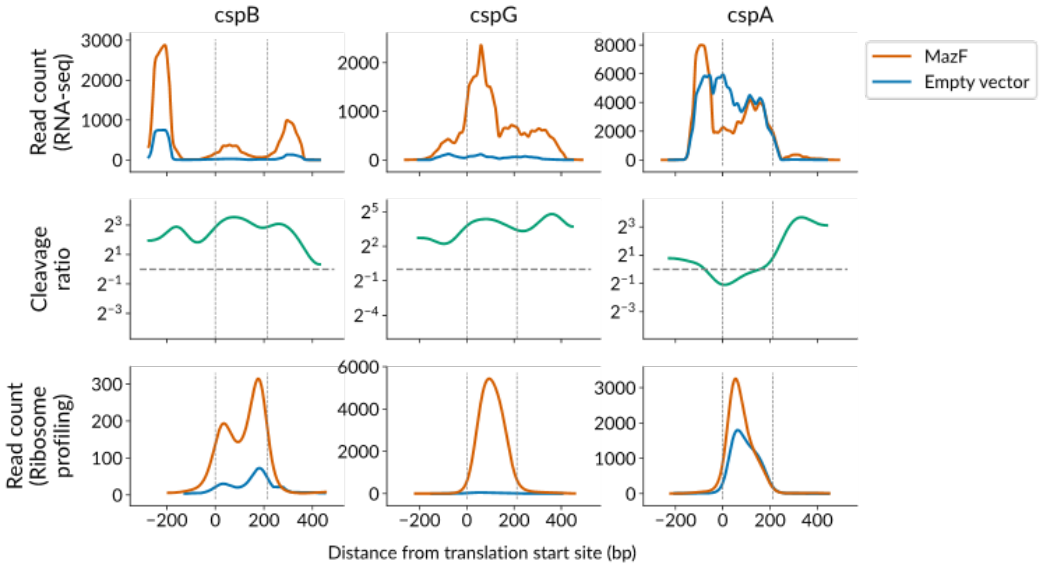

Figure S4.8: **Meta-analysis of cleavage ratios from Culviner and Laub for upregulated cold shock chaperones CspA, CspB, and CspG.** The top row indicates the total reads covering the indicated nucleotide position for CspB, CspG, and CspA. The middle row indicates the cleavage ratio, which reports the ratio of total read counts between MazF-induced cells and empty vector controls, as a function of the indicated nucleotide positions in the transcript. The bottom row indicates the read counts at the indicated nucleotide position in the transcript derived from ribosome profiling. All three proteins bear evidence of upregulation in our data and in Culviner and Laub despite a putative cleavage in the 5' UTR of the CspA transcript.

Table 4.1: Coefficients for nucleotide triplet count from regressions of log fold change on CDS length and the count of the indicated nucleotide triplet. Adjusted p-values for FDR control were computed using the Benjamini-Hochberg procedure.[43]

rank	triplet	estimate	std.error	CIL	CIR	statistic	p.value	p.adjust
1	ACA	-0.0825	0.00886	-0.0999	-0.0651	-9.31	4.86e-20	3.11e-18
2	CAA	-0.0476	0.0071	-0.0615	-0.0336	-6.7	3.1e-11	1.95e-09
3	AAA	-0.034	0.00581	-0.0454	-0.0226	-5.84	6.32e-09	3.92e-07
4	AAC	-0.0385	0.00664	-0.0515	-0.0255	-5.8	8.23e-09	5.02e-07
5	ACT	-0.0479	0.00894	-0.0654	-0.0303	-5.35	1.01e-07	6.06e-06
6	CAC	-0.045	0.0088	-0.0623	-0.0277	-5.11	3.61e-07	2.13e-05
7	TTG	0.0298	0.00694	0.0162	0.0434	4.3	1.83e-05	0.00106
8	GCT	0.0272	0.00646	0.0145	0.0399	4.21	2.74e-05	0.00156
9	CGT	0.0248	0.00598	0.0131	0.0366	4.16	3.41e-05	0.00191
10	TCG	0.0297	0.00779	0.0144	0.045	3.81	0.000146	0.00804
11	CTA	-0.0426	0.0112	-0.0646	-0.0205	-3.79	0.000158	0.00851
12	GAC	-0.0297	0.00799	-0.0454	-0.014	-3.72	0.000208	0.011
13	TGC	0.0237	0.00663	0.0107	0.0367	3.58	0.000357	0.0186
14	ACC	-0.0243	0.0069	-0.0378	-0.0107	-3.52	0.000447	0.0228
15	GCG	0.0165	0.00475	0.0072	0.0258	3.48	0.000523	0.0262
16	AAT	-0.0241	0.00706	-0.0379	-0.0102	-3.41	0.000675	0.0331
17	CGC	0.0171	0.00508	0.00712	0.0271	3.36	0.00079	0.0379
18	GAG	0.0301	0.00941	0.0116	0.0485	3.2	0.00142	0.0668
19	GTG	0.0215	0.00688	0.00797	0.035	3.12	0.00185	0.0851
20	GTC	0.0266	0.00884	0.0093	0.044	3.01	0.00263	0.118
21	TTT	0.0193	0.00695	0.00566	0.0329	2.78	0.00558	0.245
22	GTT	0.0202	0.00745	0.00555	0.0348	2.71	0.00689	0.296
23	GAA	-0.0122	0.00477	-0.0215	-0.00279	-2.55	0.011	0.46
24	TGT	0.0222	0.00887	0.00477	0.0396	2.5	0.0126	0.515
25	TGG	0.014	0.00563	0.00294	0.025	2.48	0.0131	0.524
26	TAC	-0.0203	0.00903	-0.038	-0.0026	-2.25	0.0246	0.961
27	CCA	-0.0191	0.00858	-0.0359	-0.00226	-2.23	0.0262	0.996
28	GCC	0.0126	0.00576	0.00133	0.0239	2.19	0.0286	1
29	CTT	0.0185	0.00935	0.000157	0.0368	1.98	0.0481	1
30	ATA	-0.017	0.00897	-0.0346	0.000577	-1.9	0.058	1
31	AGA	-0.0138	0.00762	-0.0287	0.00118	-1.81	0.0709	1
32	TCA	-0.0167	0.00932	-0.035	0.00157	-1.79	0.0732	1
33	CTC	0.0202	0.0114	-0.00228	0.0426	1.76	0.0782	1
34	ATG	-0.0128	0.00734	-0.0272	0.00159	-1.75	0.0811	1
35	GAT	0.0122	0.00705	-0.00168	0.026	1.72	0.0849	1
36	AGC	0.0103	0.00611	-0.00168	0.0223	1.69	0.092	1
37	GGT	0.01	0.00602	-0.00179	0.0218	1.66	0.0961	1
38	GCA	-0.0107	0.00647	-0.0234	0.00199	-1.65	0.0983	1

Table 4.1 continued from previous page

rank	triplet	estimate	std.error	CIL	CIR	statistic	p.value	p.adjust
39	TTA	0.00995	0.00676	-0.00331	0.0232	1.47	0.141	1
40	TGA	-0.01	0.00696	-0.0237	0.00366	-1.44	0.151	1
41	ATT	0.00726	0.00571	-0.00394	0.0185	1.27	0.204	1
42	AAG	-0.00617	0.00519	-0.0163	0.00402	-1.19	0.235	1
43	TCT	0.00966	0.00842	-0.00686	0.0262	1.15	0.251	1
44	CTG	0.00504	0.00488	-0.00454	0.0146	1.03	0.303	1
45	TAA	-0.00864	0.00873	-0.0258	0.00848	-0.99	0.323	1
46	GGG	0.00923	0.00947	-0.00934	0.0278	0.975	0.33	1
47	CAG	-0.00516	0.00578	-0.0165	0.00618	-0.893	0.372	1
48	CCC	0.0106	0.012	-0.013	0.0341	0.879	0.38	1
49	CAT	-0.00538	0.00865	-0.0223	0.0116	-0.622	0.534	1
50	CCG	0.00367	0.00598	-0.00806	0.0154	0.614	0.54	1
51	CCT	0.00448	0.00978	-0.0147	0.0237	0.457	0.647	1
52	TAT	0.00342	0.00768	-0.0117	0.0185	0.445	0.656	1
53	AGG	-0.0043	0.0101	-0.0241	0.0155	-0.426	0.67	1
54	GTA	0.00311	0.00777	-0.0121	0.0184	0.4	0.689	1
55	AGT	-0.00395	0.0106	-0.0248	0.0169	-0.372	0.71	1
56	CGG	-0.00204	0.00613	-0.0141	0.00999	-0.332	0.74	1
57	GGA	0.00252	0.00783	-0.0128	0.0179	0.322	0.748	1
58	ACG	-0.00217	0.00781	-0.0175	0.0132	-0.278	0.781	1
59	TTC	-0.00153	0.00748	-0.0162	0.0131	-0.204	0.838	1
60	TCC	0.0014	0.00896	-0.0162	0.019	0.156	0.876	1
61	TAG	0.0023	0.0164	-0.0298	0.0344	0.141	0.888	1
62	ATC	-0.00097	0.00745	-0.0156	0.0136	-0.131	0.896	1
63	GGC	0.000652	0.00505	-0.00925	0.0106	0.129	0.897	1
64	CGA	-0.00058	0.00667	-0.0137	0.0125	-0.0876	0.93	1

Table 4.2: The 24 proteins bearing no statistically significant evidence of downregulation across all three differential expression comparisons

Protein	Description
CspA	Cold shock protein CspA
CspB	Cold shock-like protein CspB
CspG	Cold shock-like protein CspG
DsbB	Disulfide bond formation protein B
GltP	Proton/glutamate-aspartate symporter
InfC	Translation initiation factor IF-3
PanZ	PanD regulatory factor
PdeL	Cyclic di-GMP phosphodiesterase PdeL
ProP	Proline/betaine transporter
PspC	Phage shock protein C
RbfA	Ribosome-binding factor A
RnpA	Ribonuclease P protein component
RplQ	50S ribosomal protein L17
RpmE	50S ribosomal protein L31
RpmJ	50S ribosomal protein L36 1
RsfS	Ribosomal silencing factor RsfS
SixA	Phosphohistidine phosphatase SixA
TruB	tRNA pseudouridine synthase B
Usg	USG-1 protein
YccF	Inner membrane protein YccF
YciA	Acyl-CoA thioester hydrolase YciA
YecF	Uncharacterized protein YecF
YefM	Antitoxin YefM
YggN	Uncharacterized protein YggN

Strains

Table 4.3: Strains used in this study

Strain	Description	Reference
MO::*mazE*	MG1655 derivative with arabinose-inducible chromosomal MazE expression cassette	[22]
MO::*mazE-mazF*	MG1655 derivative with arabinose-inducible chromosomal MazE expression cassette and anhydrotetracycline-inducible chromosomal MazF expression cassette	[22]
MO::*mazE cspA-3xFLAG*	C-terminal 3xFLAG insertion at *cspA* locus	This work
MO::*mazE-mazF cspA-3xFLAG*	C-terminal 3xFLAG insertion at *cspA* locus	This work
MO::*mazE cspA-mGFPmut3*	C-terminal mGFPmut3 insertion at *cspA* locus	This work
MO::*mazE-mazF cspA-mGFPmut3*	C-terminal mGFPmut3 insertion at *cspA* locus	This work

Plasmids

Table 4.4: Plasmids used in this study

Plasmid	Description	Reference
pTarget	sgRNA expression vector	[27]
pCas	Cas9 expression vector	[27]
pBbS5k	Low-copy vector with SC101 origin of replication	[48]
P_{cspA}-cspA	CspA driven by the CspA promoter replacing the expression cassette in pBbS5k	This work
P_{cspA}-cspA-FLAG	CspA-FLAG driven by the CspA promoter replacing the expression cassette in pBbS5k	This work
P_{cspA}-cspA-3xFLAG	CspA-3xFLAG driven by the CspA promoter replacing the expression cassette in pBbS5k	This work
pTarget-N20(cspA)	Cloning intermediate with sgRNA expression targeting Cas9 to *cspA* locus	This work
pTarget cspA-3xFLAG	sgRNA expression targeting Cas9 to *cspA* locus with cspA-3xFLAG donor DNA	This work
pTarget cspA-mGFPmut3	sgRNA expression targeting Cas9 to *cspA* locus with cspA-mGFPmut3 donor DNA	This work

Primers and Sequences

Table 4.5: Primers used in this study

Primer	Sequence (5' to 3')	Purpose
B-cspA-SC101-fwd	AGAATGCGCAACGCCGTAAG tgttctggtgttgctagttt gttatc	Insertion of P_{cspA}::*cspA* transcription unit into pBbS5k backbone
B-cspA-SC101-rev	CAATTTCGGTACAGGACGTT gcgcaacgcaattaatgtaa gttag	Insertion of P_{cspA}::*cspA* transcription unit into pBbS5k backbone
cspA-fwd	AACGTCCTGTACCGAAATTG AGC	Insertion of P_{cspA}::*cspA* transcription unit into pBbS5k backbone
cspA-rev	CTTACGGCGTTGCGCATTC	Insertion of P_{cspA}::*cspA* transcription unit into pBbS5k backbone
cspA-FLAG-fwd	/5Phos/GACTATAAGGACG ACGACGATAAGtaatctctg cttaaaagcacagaatctaa g	Insertion of FLAG into P_{cspA}-cspA
cspA-FLAG-rev	/5Phos/caggctggttacg ttaccag	Insertion of FLAG into P_{cspA}-cspA
cspA-3xFLAG-fwd	/5Phos/GATATCGACTATA AGGACGACGACGATAAGtaa tctctgcttaaaagcacaga atctaag	Insertion of 3xFLAG into P_{cspA}-cspA
cspA-3xFLAG-rev	/5Phos/GTGGTCCTTATAG TCACCGTCGTGGTCCTTATA GTCcaggctggttacgttac cag	Insertion of 3xFLAG into P_{cspA}-cspA
pTarget-N20-ins-rev	/5Phos/actagtattatac ctaggactgagctag	Insertion of N20 sequence for targeting *cspA* locus into pTarget

Table 4.5 continued from previous page

Primer	Sequence (5' to 3')	Purpose
cspA-N20-fwd	/5Phos/TTTAAGCAGAGAT TACAGGCgttttagagctag aaatagcaagttaaaataag g	Insertion of N20 sequence for targeting *cspA* locus into pTarget
cspA-donor-fwd	AACGTCCTGTACCGAAATTG AG	Amplification of CspA-3xFLAG donor DNA sequence from P_{cspA}-cspA-3xFLAG for targeting *cspA* locus into pTarget
cspA-donor-rev	GAGCTGGCGGCGTTTTTAGC CTGCAAATTGAAAGAGTAAG AGTCTTCGGCGGGAAATTAT TCCCGCCTTActtacggcgt tgcgcattc	Amplification of CspA-3xFLAG donor DNA sequence from P_{cspA}-cspA-3xFLAG for targeting *cspA* locus into pTarget
cspA-donor-ins-rev	CAATTTCGGTACAGGACGTT gtagggataacagggtaata gatctaagc	Insertion of CspA-3xFLAG donor DNA into pTarget
cspA-donor-2-ins-fwd	GCTAAAAACGCCGCCAGCTC tcgagttcatgtgcagctcc	Insertion of CspA-3xFLAG donor DNA into pTarget
cspA-gfp-fwd	TAATCTCTGCTTAAAAGCAC AGAATCTAAGATC	Insertion of CspA-mGFPmut3 donor DNA sequence for targeting *cspA* locus into pTarget
cspA-gfp-rev	CAGGCTGGTTACGTTACCAG C	Insertion of CspA-mGFPmut3 donor DNA sequence for targeting *cspA* locus into pTarget

Table 4.6: Sequences in this study

Name	Sequence (5’ to 3’)
mGFPmut3-ACAfree	CTGGTAACGTAACCAGCCTGAGCGGTGGCGGCGGTATGTC AAAGGGGGAAGAACTGTTTACGGGTGTTGTTCCGATTCTG GTCGAGTTGGATGGGGACGTGAACGGGCATAAGTTCAGTG TATCAGGCGAAGGTGAGGGTGATGCCACTTATGGTAAATT AACGCTTAAGTTCATCTGTACTACTGGTAAACTGCCTGTC CCTTGGCCGACTTTAGTTACGACGTTTGGTTACGGGGTTC AGTGTTTTGCTCGCTACCCAGATCATATGAAGCAGCATGA TTTTTTTAAGTCTGCCATGCCTGAGGGTTATGTGCAAGAA CGTACCATTTTCTTCAAGGATGATGGGAATTATAAAACCC GTGCCGAGGTAAAGTTCGAAGGTGATACGCTGGTTAATCG TATCGAATTAAAGGGGATTGACTTCAAAGAAGACGGTAAT ATCCTTGGCCATAAGTTGGAATATAACTATAACTCGCATA ATGTCTATATCATGGCGGATAAGCAGAAAAATGGAATCAA GGTGAATTTCAAAATTCGTCATAATATCGAAGACGGATCT GTTCAACTTGCTGACCATTATCAGCAGAATACTCCTATTG GCGATGGTCCAGTCCTTCTGCCGGATAATCATTATTTATC CACCCAGTCGAAGCTGAGTAAAGACCCGAACGAAAAGCGC GATCATATGGTGTTATTGGAGTTTGTTACCGCTGCAGGAA TCACTCATGGCATGGACGAGTTATATAAGTAATCTCTGCT TAAAAGCAC

4.8 Supplementary Datasets

Dataset 4.1. Proteomic results for MazF expression profiling.

This dataset describes all proteins identified through LC-MS/MS analysis of BONCAT-enriched proteins from the MazF expression screening. Column 1 provides the NCBI Gene Symbol for the protein. Columns 2-17 provide the TMT reporter ion abundances for each sample. Columns 18-26 provide the estimated fold change for each of the three comparisons (early, late, temporal) along with their *p*-values and FDR-adjusted *p*-values.

References

(1) Pandey, D. P.; Gerdes, K. Toxin-antitoxin loci are highly abundant in free-living but lost from host-associated prokaryotes. *Nucleic Acids Res* **2005**, *33*, 966–976.

(2) Gerdes, K.; Rasmussen, P. B.; Molin, S. Unique type of plasmid maintenance function: postsegregational killing of plasmid-free cells. *Proc Natl Acad Sci U S A* **1986**, *83*, 3116–3120.

(3) Harms, A.; Brodersen, D. E.; Mitarai, N.; Gerdes, K. Toxins, Targets, and Triggers: An Overview of Toxin-Antitoxin Biology. *Mol Cell* **2018**, *70*, 768–784.

(4) Pedersen, K.; Zavialov, A. V.; Pavlov, M. Y.; Elf, J.; Gerdes, K.; Ehrenberg, M. The bacterial toxin RelE displays codon-specific cleavage of mRNAs in the ribosomal A site. *Cell* **2003**, *112*, 131–140.

(5) Tan, Q.; Awano, N.; Inouye, M. YeeV is an Escherichia coli toxin that inhibits cell division by targeting the cytoskeleton proteins, FtsZ and MreB. *Mol Microbiol* **2011**, *79*, 109–118.

(6) Germain, E.; Castro-Roa, D.; Zenkin, N.; Gerdes, K. Molecular mechanism of bacterial persistence by HipA. *Mol Cell* **2013**, *52*, 248–254.

(7) Bernard, P.; Couturier, M. Cell killing by the F plasmid CcdB protein involves poisoning of DNA-topoisomerase II complexes. *J Mol Biol* **1992**, *226*, 735–745.

(8) Jiang, Y.; Pogliano, J.; Helinski, D. R.; Konieczny, I. ParE toxin encoded by the broad-host-range plasmid RK2 is an inhibitor of Escherichia coli gyrase. *Mol Microbiol* **2002**, *44*, 971–979.

(9) LeRoux, M.; Culviner, P. H.; Liu, Y. J.; Littlehale, M. L.; Laub, M. T. Stress Can Induce Transcription of Toxin-Antitoxin Systems without Activating Toxin. *Mol Cell* **2020**, *79*, 280–292.

(10) LeRoux, M.; Laub, M. T. Toxin-Antitoxin Systems as Phage Defense Elements. *Annu Rev Microbiol* **2022**, *76*, 21–43.

(11) nas, D.; Fraikin, N.; Goormaghtigh, F.; Van Melderen, L. Biology and evolution of bacterial toxin-antitoxin systems. *Nat Rev Microbiol* **2022**, *20*, 335–350.

(12) Gerdes, K.; Christensen, S. K.; bner-Olesen, A. Prokaryotic toxin-antitoxin stress response loci. *Nat Rev Microbiol* **2005**, *3*, 371–382.

(13) Zhang, Y.; Zhang, J.; Hoeflich, K. P.; Ikura, M.; Qing, G.; Inouye, M. MazF cleaves cellular mRNAs specifically at ACA to block protein synthesis in Escherichia coli. *Mol Cell* **2003**, *12*, 913–923.

(14) Amitai, S.; Kolodkin-Gal, I.; Hananya-Meltabashi, M.; Sacher, A.; Engelberg-Kulka, H. Escherichia coli MazF leads to the simultaneous selective synthesis of both 'death proteins' and 'survival proteins'. *PLoS Genet* **2009**, *5*, e1000390.

(15) Kolodkin-Gal, I.; Hazan, R.; Gaathon, A.; Carmeli, S.; Engelberg-Kulka, H. A linear pentapeptide is a quorum-sensing factor required for mazEF-mediated cell death in Escherichia coli. *Science* **2007**, *318*, 652–655.

(16) Vesper, O.; Amitai, S.; Belitsky, M.; Byrgazov, K.; Kaberdina, A. C.; Engelberg-Kulka, H.; Moll, I. Selective translation of leaderless mRNAs by specialized ribosomes generated by MazF in Escherichia coli. *Cell* **2011**, *147*, 147–157.

(17) Culviner, P. H.; Laub, M. T. Global Analysis of the E. coli Toxin MazF Reveals Widespread Cleavage of mRNA and the Inhibition of rRNA Maturation and Ribosome Biogenesis. *Mol Cell* **2018**, *70*, 868–880.

(18) Mets, T.; Kasvandik, S.; Saarma, M.; li, Ü.; Tenson, T.; Kaldalu, N. Fragmentation of Escherichia coli mRNA by MazF and MqsR. *Biochimie* **2019**, *156*, 79–91.

(19) Venturelli, O. S.; Tei, M.; Bauer, S.; Chan, L. J. G.; Petzold, C. J.; Arkin, A. P. Programming mRNA decay to modulate synthetic circuit resource allocation. *Nat Commun* **2017**, *8*, 15128.

(20) Babin, B. M.; Bergkessel, M.; Sweredoski, M. J.; Moradian, A.; Hess, S.; Newman, D. K.; Tirrell, D. A. SutA is a bacterial transcription factor expressed during slow growth in Pseudomonas aeruginosa. *Proc Natl Acad Sci U S A* **2016**, *113*, 597–605.

(21) Babin, B. M.; Atangcho, L.; van Eldijk, M. B.; Sweredoski, M. J.; Moradian, A.; Hess, S.; Tolker-Nielsen, T.; Newman, D. K.; Tirrell, D. A. Selective Proteomic Analysis of Antibiotic-Tolerant Cellular Subpopulations in Pseudomonas aeruginosa Biofilms. *mBio* **2017**, *8*, 10.1128/mbio.01593–17.

(22) Mok, W. W.; Park, J. O.; Rabinowitz, J. D.; Brynildsen, M. P. RNA Futile Cycling in Model Persisters Derived from MazF Accumulation. *mBio* **2015**, *6*, e01588–01515.

(23) Sauert, M.; Wolfinger, M. T.; Vesper, O.; ller, C.; Byrgazov, K.; Moll, I. The MazF-regulon: a toolbox for the post-transcriptional stress response in Escherichia coli. *Nucleic Acids Res* **2016**, *44*, 6660–6675.

(24) Jiang, W.; Hou, Y.; Inouye, M. CspA, the major cold-shock protein of Escherichia coli, is an RNA chaperone. *J Biol Chem* **1997**, *272*, 196–202.

(25) Yamanaka, K.; Fang, L.; Inouye, M. The CspA family in Escherichia coli: multiple gene duplication for stress adaptation. *Mol Microbiol* **1998**, *27*, 247–255.

(26) Nikolay, R.; Hilal, T.; Schmidt, S.; Qin, B.; Schwefel, D.; Vieira-Vieira, C. H.; Mielke, T.; rger, J.; Loerke, J.; Amikura, K.; gel, T.; Ueda, T.; Selbach, M.; Deuerling, E.; Spahn, C. M. T. Snapshots of native pre-50S ribosomes reveal a biogenesis factor network and evolutionary specialization. *Mol Cell* **2021**, *81*, 1200–1215.

(27) Jiang, Y.; Chen, B.; Duan, C.; Sun, B.; Yang, J.; Yang, S. Multigene editing in the Escherichia coli genome via the CRISPR-Cas9 system. *Appl Environ Microbiol* **2015**, *81*, 2506–2514.

(28) Cho, J.; Carr, A. N.; Whitworth, L.; Johnson, B.; Wilson, K. S. MazEF toxin-antitoxin proteins alter Escherichia coli cell morphology and infrastructure during persister formation and regrowth. *Microbiology (Reading)* **2017**, *163*, 308–321.

(29) Zhu, L.; Phadtare, S.; Nariya, H.; Ouyang, M.; Husson, R. N.; Inouye, M. The mRNA interferases, MazF-mt3 and MazF-mt7 from Mycobacterium tuberculosis target unique pentad sequences in single-stranded RNA. *Mol Microbiol* **2008**, *69*, 559–569.

(30) Zhu, L.; Inoue, K.; Yoshizumi, S.; Kobayashi, H.; Zhang, Y.; Ouyang, M.; Kato, F.; Sugai, M.; Inouye, M. Staphylococcus aureus MazF specifically cleaves a pentad sequence, UACAU, which is unusually abundant in the mRNA for pathogenic adhesive factor SraP. *J Bacteriol* **2009**, *191*, 3248–3255.

(31) Yamaguchi, Y.; Nariya, H.; Park, J. H.; Inouye, M. Inhibition of specific gene expressions by protein-mediated mRNA interference. *Nat Commun* **2012**, *3*, 607.

(32) Zhang, Y.; Burkhardt, D. H.; Rouskin, S.; Li, G. W.; Weissman, J. S.; Gross, C. A. A Stress Response that Monitors and Regulates mRNA Structure Is Central to Cold Shock Adaptation. *Mol Cell* **2018**, *70*, 274–286.

(33) VanBogelen, R. A.; Neidhardt, F. C. Ribosomes as sensors of heat and cold shock in Escherichia coli. *Proc Natl Acad Sci U S A* **1990**, *87*, 5589–5593.

(34) Giuliodori, A. M.; Brandi, A.; Gualerzi, C. O.; Pon, C. L. Preferential translation of cold-shock mRNAs during cold adaptation. *RNA* **2004**, *10*, 265–276.

(35) Fang, L.; Jiang, W.; Bae, W.; Inouye, M. Promoter-independent cold-shock induction of cspA and its derepression at 37 degrees C by mRNA stabilization. *Mol Microbiol* **1997**, *23*, 355–364.

(36) Goldenberg, D.; Azar, I.; Oppenheim, A. B. Differential mRNA stability of the cspA gene in the cold-shock response of Escherichia coli. *Mol Microbiol* **1996**, *19*, 241–248.

(37) Cui, Y.; Su, X.; Wang, C.; Xu, H.; Hu, D.; Wang, J.; Pei, K.; Sun, M.; Zou, T. Bacterial MazF/MazE toxin-antitoxin suppresses lytic propagation of arbitrium-containing phages. *Cell Rep* **2022**, *41*, 111752.

(38) Alawneh, A. M.; Qi, D.; Yonesaki, T.; Otsuka, Y. An ADP-ribosyltransferase Alt of bacteriophage T4 negatively regulates the Escherichia coliMazF toxin of a toxin–antitoxin module. *Molecular Microbiology* **2016**, *99*, 188–198.

(39) Bae, S.; Park, J.; Kim, J. S. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* **2014**, *30*, 1473–1475.

(40) Hong, V.; Presolski, S. I.; Ma, C.; Finn, M. G. Analysis and optimization of copper-catalyzed azide-alkyne cycloaddition for bioconjugation. *Angew Chem Int Ed Engl* **2009**, *48*, 9879–9883.

(41) Glenn, W. S.; Stone, S. E.; Ho, S. H.; Sweredoski, M. J.; Moradian, A.; Hess, S.; Bailey-Serres, J.; Tirrell, D. A. Bioorthogonal Noncanonical Amino Acid Tagging (BONCAT) Enables Time-Resolved Analysis of Protein Synthesis in Native Plant Tissue. *Plant Physiol* **2017**, *173*, 1543–1553.

(42) Ritchie, M. E.; Phipson, B.; Wu, D.; Hu, Y.; Law, C. W.; Shi, W.; Smyth, G. K. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **2015**, *43*, e47.

(43) Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **1995**, *57*, 289–300.

(44) Szklarczyk, D.; Kirsch, R.; Koutrouli, M.; Nastou, K.; Mehryary, F.; Hachilif, R.; Gable, A. L.; Fang, T.; Doncheva, N. T.; Pyysalo, S.; Bork, P.; Jensen, L. J.; von Mering, C. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res* **2023**, *51*, D638–D646.

(45) Krassowski, M.; Arts, M.; Lagger, C.; Max krassowski/complex-upset: v1.3.5, 2022.

(46) Gu, Z.; Hübschmann, D. simplifyEnrichment: A Bioconductor Package for Clustering and Visualizing Functional Enrichment Results. *Genomics Proteomics Bioinformatics* **2023**, *21*, 190–202.

(47) Langmead, B.; Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **2012**, *9*, 357–359.

(48) Lee, T. S.; Krupa, R. A.; Zhang, F.; Hajimorad, M.; Holtz, W. J.; Prasad, N.; Lee, S. K.; Keasling, J. D. BglBrick vectors and datasheets: A synthetic biology platform for gene expression. *J Biol Eng* **2011**, *5*, 12.

Appendix A

OTHER CONTRIBUTIONS

Contributions to Jones, J. et al.
Published as:

1. Jones, J.; MacKrell, E. J.; Wang, T. Y.; Lomenick, B.; Roukes, M. L.; Chou, T. F. Tidyproteomics: an open-source R package and data object for quantitative proteomics post analysis and visualization. *BMC Bioinformatics* **2023**, *24*, 239, DOI: 10.1186/s12859-023-05360-7,

Abstract

Background The analysis of mass spectrometry-based quantitative proteomics data can be challenging given the variety of established analysis platforms, the differences in reporting formats, and a general lack of approachable standardized post-processing analyses such as sample group statistics, quantitative variation and even data filtering. We developed *tidyproteomics* to facilitate basic analysis, improve data interoperability and potentially ease the integration of new processing algorithms, mainly through the use of a simplified data-object.

Results The R package *tidyproteomics* was developed as both a framework for standardizing quantitative proteomics data and a platform for analysis workflows, containing discrete functions that can be connected end-to-end, thus making it easier to define complex analyses by breaking them into small stepwise units. Additionally, as with any analysis workflow, choices made during analysis can have large impacts on the results and as such, *tidyproteomics* allows researchers to string each function together in any order, select from a variety of options and in some cases develop and incorporate custom algorithms.

Conclusions *Tidyproteomics* aims to simplify data exploration from multiple platforms, provide control over individual functions and analysis order, and serve as a tool to assemble complex repeatable processing workflows in a logical flow. Datasets in *tidyproteomics* are easy to work with, have a structure that allows for biological annotations to be added, and come with a framework for developing additional analysis tools. The consistent data structure and accessible analysis and plotting tools also offers a way for researchers to save time on mundane data manipulation tasks.

Contributions

I developed an interactive application for graphical user interface–based quality control, filtering, normalization, differential expression analysis, and ontology enrichment of LC-MS/MS–based proteomics data for the *tidyproteomics* R package. I published and maintained the GitHub repository for the application, a separate documentation website for the application, and the publicly accessible deployed application on the Caltech Proteome Exploration Laboratory server.

Contributions to Altshuller, M. et al.

Published as:

1. Altshuller, M.; He, X.; MacKrell, E. J.; Wernke, K. M.; Wong, J. W. H.; Sellés-Baiget, S.; Wang, T. Y.; Chou, T. F.; Duxin, J. P.; Balskus, E. P.; Herzon, S. B.; Semlow, D. R. The Fanconi anemia pathway repairs colibactin-induced DNA interstrand cross-links. *bioRxiv* **2024**, DOI: 10.1101/2024.01.30.576698,

Abstract

Colibactin is a secondary metabolite produced by bacteria present in the human gut and is implicated in the progression of colorectal cancer and inflammatory bowel disease. This genotoxin alkylates deoxyadenosines on opposite strands of host cell DNA to produce DNA interstrand cross-links (ICLs) that block DNA replication. While cells have evolved multiple mechanisms to resolve ("unhook") ICLs encountered by the replication machinery, little is known about which of these pathways promote resistance to colibactin-induced ICLs. Here, we use *Xenopus* egg extracts to investigate replication-coupled repair of plasmids engineered to contain site-specific colibactin-ICLs. We show that replication fork stalling at a colibactin-ICL leads to replisome disassembly and activation of the Fanconi anemia ICL repair pathway, which unhooks the colibactin-ICL through nucleolytic incisions. These incisions generate a DNA double-strand break intermediate in one sister chromatid, which can be repaired by homologous recombination, and a monoadduct ("ICL remnant") in the other. Our data indicate that translesion synthesis past the colibactin-ICL remnant depends on Polη and a Polκ-REV1-Polζ polymerase complex. Although translesion synthesis past colibactin-induced DNA damage is frequently error-free, it can introduce T>N point mutations that partially recapitulate the mutation signature associated with colibactin exposure *in vivo*. Taken together, our work provides a biochemical framework for understanding how cells tolerate a naturally-occurring and clinically-relevant ICL.

Contributions

I conducted quality control, normalization, and differential expression analysis of LC-MS/MS–based proteomics data. I developed a specialized interactive application for visualizing protein abundances, differential expression, gene ontology enrichment, time-series clustering, and time-series correlational analysis.