QUANTITATIVE, TIME-RESOLVED PROTEOMIC ANALYSIS USING BIO-ORTHOGONAL NON-CANONICAL AMINO ACID TAGGING

Thesis by

John David Bagert

In Partial Fulfillment of the Requirements for the degree

of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2015

(Defended May 20, 2015)

To the memory of Katie Fisher

ACKNOWLEDGEMENTS

Throughout my time at Caltech, I've been fortunate to interact with many creative, intelligent scientists from diverse academic backgrounds. First I thank my advisor, Dave Tirrell, for his guidance and support throughout the years. Dave exemplifies the highest standards of scientific rigor; he has been and will continue to be a model to which I aspire throughout my life. I also thank my committee members, Frances Arnold, Rob Phillips, and Jared Leadbetter for their advice and support.

I thank past and present members of the Tirrell lab for providing a rich, healthy, and fun learning and work environment. During my early years, Frank Truong, John Ngo, Alborz Mahdavi, Beverly Lu, and Jim Van Deventer helped me to get comfortable in lab and learn molecular biology techniques. I had many fruitful discussions with Brett Babin about BONCAT, covering everything from experimental techniques to statistical analysis. I also thank my collaborators, Julia van Kessel and Bonnie Bassler, for their hard work and for sharing their expertise.

Present and former members of the Proteome Exploration Laboratory operated and maintained LC-MS/MS instruments and were fundamental to my research. In particular, I thank Sonja Hess for discussions about mass spectrometry-based proteomics and Mike Sweredoski for talks about and help with bioinformatic analysis.

I thank Matt Smith and Ethan Ogilby for their wonderful friendship. I thank my family for their love and ceaseless encouragement and support. Last, I thank my partner, June Wicks, who is a continual source of happiness and inspiration in my life.

ABSTRACT

Bio-orthogonal non-canonical amino acid tagging (BONCAT) is an analytical method that allows the selective analysis of the subset of newly synthesized cellular proteins produced in response to a biological stimulus. In BONCAT, cells are treated with the non-canonical amino acid L-azidohomoalanine (Aha), which is utilized in protein synthesis in place of methionine by wild-type translational machinery. Nascent, Aha-labeled proteins are selectively ligated to affinity tags for enrichment and subsequently identified via mass spectrometry. The work presented in this thesis exhibits advancements in and applications of the BONCAT technology that establishes it as an effective tool for analyzing proteome dynamics with time-resolved precision.

Chapter 1 introduces the BONCAT method and serves as an outline for the thesis as a whole. I discuss motivations behind the methodological advancements in Chapter 2 and the biological applications in Chapters 2 and 3.

Chapter 2 presents methodological developments that make BONCAT a proteomic tool capable of, in addition to identifying newly synthesized proteins, accurately quantifying rates of protein synthesis. I demonstrate that this quantitative BONCAT approach can measure proteome-wide patterns of protein synthesis at time scales inaccessible to alternative techniques.

In Chapter 3, I use BONCAT to study the biological function of the small RNA regulator CyaR in *Escherichia coli*. I correctly identify previously known CyaR targets, and validate several new CyaR targets, expanding the functional roles of the sRNA regulator.

In Chapter 4, I use BONCAT to measure the proteomic profile of the quorum sensing bacterium *Vibrio harveyi* during the time-dependent transition from individual- to group-behaviors. My analysis reveals new quorum-sensing-regulated proteins with diverse functions,

including transcription factors, chemotaxis proteins, transport proteins, and proteins involved in iron homeostasis.

Overall, this work describes how to use BONCAT to perform quantitative, time-resolved proteomic analysis and demonstrates that these measurements can be used to study a broad range of biological processes.

Table of Contents

LIST OF FIGURES

Figure *Page*

LIST OF TABLES

Table *Page*

CHAPTER 1

THE DEVELOPMENT AND APPLICATION OF QUANTITATIVE BIO-ORTHOGONAL NON-CANONICAL AMINO ACID TAGGING FOR TIME-RESOLVED PROTEOMICS

Introduction and summary

In 2006, the Tirrell and Schuman groups introduced a method for the selective analysis of newly synthesized cellular proteins called bio-orthogonal non-canonical amino acid tagging (BONCAT).[1] In BONCAT, cells in culture are treated with the non-canonical amino acid L-azidohomoalanine (Aha), which is used in place of methionine by wild-type translational machinery.[2] Once incorporated into nascent proteins, Aha serves as a chemical handle for the selective ligation of fluorescent dyes for visualization or affinity tags for enrichment and subsequent identification via mass spectrometry.[1,3]

After the introduction of BONCAT, several methodological hurdles remained that were essential for the advancement and expanded application of the technique. First, BONCAT had been successfully used as a qualitative tool for identifying newly synthesized proteins produced in response to a biological stimulus, but the method could not yet establish a quantitative relationship between identified proteins and the stimulus. Such information was necessary for accurately determining whether a protein's production was influenced by the stimulus of interest and, if so, how much. Second, the extent to which Aha incorporation itself perturbed endogenous protein synthesis was not well understood. Measuring these effects was necessary for identifying Aha-labeling strategies that were sufficiently mild to avoid unintentionally perturbing protein synthesis. Last, BONCAT had not been benchmarked against alternative methods for analyzing newly synthesized proteins. Pulsed metabolic labeling with amino acid isotopologs, for example, can distinguish nascent proteins by mass.[4] Because the BONCAT protocol, in particular the tagging and enrichment of Aha-labeled proteins, is more experimentally laborious than that of stable isotope labeling, we had to determine whether BONCAT offered significant advantages for probing protein synthesis. Additionally, the conditions under which to expect such improvements should also be understood.

Chapter 2 of this thesis presents methodological improvements that address these challenges and questions. We demonstrate that a combined BONCAT and stable isotope labeling with amino acids in cell culture (SILAC) approach is capable of both identifying and quantifying newly synthesized proteins with high fidelity. We use quantitative proteomics to measure how Aha labeling perturbs protein synthesis and identify an Aha labeling strategy that abolishes these effects, even for long labeling times. Last, we show that the combined BONCAT-SILAC approach can measure proteome-wide patterns of protein synthesis at time scales inaccessible to stable isotope labeling techniques alone.

There are two general cases in which a proteomics experiment would strongly benefit from enhanced time resolution: when the experiment aims to measure the proteins that change first in response to a stimulus, and when the experiment aims to define the temporal order of changes in protein synthesis over a prolonged period of time. Chapters 3 and 4, respectively, present biological applications of BONCAT that exemplify these general cases.

Chapter 3 describes a proteomic approach for determining the biological function of small RNA (sRNA) regulators in bacteria. The most prevalent class of sRNAs regulate protein production post-transcriptionally by base-pairing with target mRNA molecules and modulating their abundance or translation. Base-pairing regions are often short and non-contiguous, making target prediction difficult. We use BONCAT to measure the proteome-wide changes due to the expression of the sRNA CyaR on the time scale of minutes. We identify a small group of rapidly-regulated proteins, which includes previously known direct CyaR targets. We go on to validate several new direct CyaR targets, expanding the functional roles of the sRNA regulator in carbon metabolism, osmoregulation, and transcriptional regulation.

In Chapter 4 we use BONCAT to study quorum sensing in the model bacterium *Vibrio harveyi*. Quorum sensing is a method of chemical communication that bacteria use to assess

their population density and to change their behavior in response to fluctuations in the cell number and species composition of the community. In this work we use BONCAT to characterize the time-dependent transition of the quorum-sensing bacterium *Vibrio harveyi* from individual- to group-behaviors. We identify quorum-sensing-regulated proteins at early, intermediate, and late stages of the transition, and show how temporal changes in quorum-sensing proteins can be programmed by both transcriptional and post-transcriptional mechanisms. Our analysis revealed new quorum-sensing-regulated proteins with diverse functions, including transcription factors, chemotaxis proteins, transport proteins, and proteins involved in iron homeostasis.

Appendices A–D present brief summaries of work resulting from collaborations during my PhD that are not directly related to this thesis. These results have either been published or are currently being prepared for submission for publication.

The work presented in this thesis demonstrates BONCAT as an effective tool for analyzing protein synthesis with time-resolved precision. I expect the experimental and analytical techniques developed here will be useful and broadly applicable for studying proteome dynamics in a variety of biological processes.

References

(1) Dieterich, D. C., Link, A. J., Graumann, J., Tirrell, D. A., and Schuman, E. M. (2006) Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT). *Proc. Natl. Acad. Sci. U.S.A. 103*, 9482–7.

(2) Kiick, K. L., Saxon, E., Tirrell, D. A., and Bertozzi, C. R. (2002) Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation. *Proc. Natl. Acad. Sci. U.S.A. 99*, 19–24.

(3) Beatty, K. E., and Tirrell, D. A. (2008) Two-color labeling of temporally defined protein populations in mammalian cells. *Bioorg. Med. Chem. Lett. 18*, 5995–9.

(4) Schwanhäusser, B., Gossen, M., Dittmar, G., and Selbach, M. (2009) Global analysis of cellular protein translation by pulsed SILAC. *Proteomics 9*, 205–9.

CHAPTER 2

QUANTITATIVE, TIME-RESOLVED PROTEOMIC ANALYSIS BY COMBINING BIO-ORTHOGONAL NON-CANONICAL AMINO ACID TAGGING AND PULSED STABLE ISOTOPE LABELING BY AMINO ACIDS IN CELL CULTURE

This chapter has been published as:

Bagert, J. D., Xie, Y. J., Sweredoski, M. J., Qi, Y., Hess, S., Schuman, E. M., and Tirrell, D. A. (2014) 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* 13, 1352–8.

ABBREVIATIONS

BONCAT: Bioorthogonal noncanonical amino acid tagging

pSILAC: Pulsed stable isotope labeling with amino acids in cell culture

Aha: L-azidohomoalanine

$Aha^{30:1}$: A mixture of Aha (1 mM) and Met (33 µM)

ncAA: Noncanonical amino acid

H/M: Heavy to medium ratio

GO: Gene ontology

SUMMARY

An approach to proteomic analysis that combines bioorthogonal noncanonical amino acid tagging (BONCAT) and pulsed stable isotope labeling with amino acids in cell culture (pSILAC) provides accurate quantitative information about rates of cellular protein synthesis on time scales of minutes. The method is capable of quantifying 1400 proteins produced by HeLa cells during a 30-min interval, a time scale that is inaccessible to isotope labeling techniques alone. Potential artifacts in protein quantification can be reduced to insignificant levels by limiting the extent of noncanonical amino acid tagging. We find no evidence for artifacts in protein identification in experiments that combine the BONCAT and pSILAC methods.

INTRODUCTION

Methods for the analysis of cellular protein synthesis should be quantitative and fast. In 2006, Dieterich and coworkers introduced a proteomics discovery tool called bioorthogonal noncanonical amino acid tagging (BONCAT), in which noncanonical amino acids (ncAAs) with bioorthogonal functional groups (e.g. azides or alkynes) are used as metabolic labels to distinguish new proteins from old (1, 2). Labeled proteins can be conjugated to fluorescent reporters for visualization or affinity tags for purification and subsequent identification by mass spectrometry (3). Because the ncAA probe can be introduced to cells in a well-defined "pulse," affinity purification removes pre-existing proteins and provides both reduced sample complexity and excellent time resolution.

The methionine (Met) surrogate L-azidohomoalanine (Aha) has become standard in the application of BONCAT methodologies. Using Aha and fluorescent tagging, Tcherkezian et al. observed co-localization of the DCC receptor with sites of protein synthesis, providing support for the role of netrin as a stimulant of extranuclear protein production in neurons (4). Combining Aha labeling and 2D gel electrophoresis, Yoon et al. discovered that the protein lamin B2 is synthesized in axons and crucial to mitochondrial function and axon maintenance in Xenopus retinal glial cells (5). Aha has also been used to study histone turnover (6), protein palmitoylation (7), pathogen amino acid uptake (8), inflammatory response (9), and local translation in neuronal dendrites and axons (10). These labeling techniques have been expanded to tissue and animal culture, where Aha has been used to profile protein synthesis in rat hippocampal brain slices (11, 12) and zebrafish embryos (13).

The development of fast, reliable, quantitative BONCAT methods will enable new insights into proteome dynamics in response to biological stimuli. Recent work by Eichelbaum et al. combined Aha labeling with stable isotope labeling to measure lipopolysaccharide-stimulated

protein secretion by macrophages (14). Using similar approaches, Somasekharan et al. identified a set of proteins that are translationally regulated by the Y-box binding protein-1 (YB-1) in TC-32 Ewing sarcoma cells (15), and Howden et al. monitored changes in protein expression following stimulation of primary T cells with phorbol 12-myristate 13-acetate and ionomycin (16).

A concern that arises in the use of Aha (as it does for all chemical probes of biological processes) is that the protocols used for Aha labeling might perturb cellular protein synthesis. The development of ncAAs as reliable analytic tools hinges on our ability to understand and minimize such unintended effects. For Aha, previous work has shown that protein labeling does not visibly alter cellular morphology in dissociated hippocampal neurons or HEK293 cells, and 1D gels reveal no discrepancies between the proteomes of Aha- and Met-treated cells (1). These experiments, however, offer only coarse measures of effects on protein synthesis, and as Aha labeling is frequently coupled to mass spectrometry-based proteomic analysis, the biological effects of Aha treatment must be investigated with equivalent sensitivity and resolution.

Here we report sound methods for fast, reliable measurement of proteome dynamics via noncanonical amino acid tagging. First, we use the quantitative proteomics technique pulsed stable isotope labeling with amino acids in cell culture (pSILAC) to investigate potential unintended effects of Aha labeling on protein abundance in HeLa cell cultures, and we develop a strategy for minimizing these effects. Second, we show that a combined BONCAT-pSILAC approach, capable of both enriching and quantifying newly synthesized proteins, yields detailed proteomic information on time scales that are inaccessible to isotope labeling techniques alone.

EXPERIMENTAL PROCEDURES

pSILAC in HeLa cell culture – HeLa cells were maintained in DMEM (Gibco) supplemented with 10% FBS (Gibco) and 1% penicillin/streptomycin (Gibco) in a humidified incubator at 37 °C and 5% CO_2. For each pSILAC experiment, 2.1 million cells were seeded in 2 T-75 flasks and grown for 24 h. Cultures were washed with warm PBS twice and resuspended in custom lysine-free and Met-free DMEM (Gibco) supplemented with either "medium" lysine (D4 L-lysine, Cambridge Isotope Laboratories) or "heavy" lysine (U-$^{13}C_6$ U-$^{15}N_2$ L-lysine, Cambridge Isotope Laboratories) at 1 mM. Cultures were also supplemented with either Met (1 mM), Aha (1 mM), or Aha$^{30:1}$ (1 mM Aha, 33 µM Met) as indicated for each experiment. Aha was synthesized as previously described (17). pSILAC experiments measuring changes in protein abundance upon treatment with Aha or Aha$^{30:1}$ and Met were conducted with four biological replicates, two of which were arranged as label swap experiments. pSILAC experiments with pulse durations of 4 h and 30 min were performed with three biological replicates. After the desired labeling time, cells were removed from the flask by trypsinization and pelleted at 4 °C. Cells were lysed in 2% SDS in PBS by heating to 90 °C for 10 min. DNA was digested with Benzonase (Sigma) and lysates were cleared by centrifugation. Protein concentrations were measured with the BCA protein quantitation kit (Thermo Scientific).

BONCAT – BONCAT experiments were carried out as described in the HeLa cell pSILAC protocol, with a few modifications. T-150 flasks were seeded with 4 million cells prior to each experiment. The larger culture size compensates for the relatively small amounts of protein that are produced during short pulses. During the pulse, both medium and heavy cultures were supplemented with either Aha or Aha$^{30:1}$. Each BONCAT experiment was conducted with three biological replicates. Protein synthesis was halted prior to cell lysis by addition of cycloheximide

(Sigma) to 100 µg/ml. Cells were lysed in freshly prepared 2% SDS in PBS with 100 mM chloroacetamide (Sigma) to alkylate free cysteines in proteins. Cysteine alkylation reduces thiol addition of cyclooctyne reagents, and increases the specificity of tagging of Aha-labeled proteins (18). After mixing heavy and medium lysates, Aha-labeled proteins were conjugated to a biotin tag by strain-promoted azide-alkyne click chemistry (19). DBCO-sulfo-biotin tag (Click Chemistry Tools) was added to 1 mg of mixed lysates to a final concentration of 12 µM and allowed to react for 15 min, after which the reaction was quenched with excess Aha. Tagged proteins were captured with Streptavidin Plus UltraLink Resin (Thermo Scientific), washed with 64 column volumes of 1% SDS in PBS, and eluted by boiling the resin in 1 mM biotin in 1% SDS in PBS for 15 min. Eluted proteins were concentrated on a 3 kDa molecular weight cut-off centrifugation filter (Amicon) prior to SDS-PAGE. Separation of excess or unreacted biotin tag prior to streptavidin capture was unnecessary due to the small quantity of tag used in the click reaction.

GeLC-MS – Proteins were separated on precast 4-12% polyacrylamide gels (Invitrogen) and visualized with the colloidal blue staining kit (Invitrogen). Lanes were cut into 8 gel pieces and destained by iterative washing with 50 mM ammonium bicarbonate (Baker) and 1:1 (v/v) 50 mM ammonium bicarbonate and acetonitrile (LC-MS grade, Fluka). Proteins were reduced in 7 mM DTT (Research Products International) in 50 mM ammonium bicarbonate at 50 °C for 30 min. After removing the DTT solution, proteins were alkylated with freshly prepared 40 mM chloroacetamide (Sigma) in 50 mM ammonium bicarbonate for 20 min in the dark. Gel pieces were washed with 50 mM ammonium bicarbonate and acetonitrile for 5 min each. Proteins were digested with endoproteinase LysC (Mako) at 3 ng/µl in 50 mM Tris (Sigma), pH 8.5, overnight at 37 °C. Digested peptides were extracted from gel pieces by washing in 1% formic acid/2% acetonitrile for 5 min, 1:1 acetonitrile/water for 5 min, and 1% formic acid in

acetonitrile for 5 min. Extracted peptides were desalted with custom packed C_{18} columns as described in Rappsilber et al. (20), concentrated by lyophilization, and resuspended in 0.1% formic acid (Sigma) prior to LC-MS/MS.

NanoLC-Mass Spectrometry Analysis – All liquid chromatography-mass spectrometry experiments were performed on an EASY-nLC (Proxeon Biosystems, now Thermo Scientific, Waltham, MA) connected to a hybrid LTQ-Orbitrap or LTQ-FT (Thermo Scientific) equipped with a nano-electrospray ion source (Proxeon Biosystems, now Thermo Scientific). pSILAC experiments with 24-h pulse times were analyzed on the LTQ-FT, and experiments comparing pSILAC and BONCAT with 4-h and 30-min pulse times were analyzed on the LTQ-Orbitrap. Peptides on the Orbitrap were separated on a 15-cm reversed phase analytical column (75 µm ID) packed in-house with 3 µm C_{18AQ} beads (ReproSil-Pur C_{18AQ}) using a 60 min elution from 0% to 30% solvent B at a flow rate of 350 nL/min. Solvent A was 0.2% formic acid, 2% acetonitrile, and 97.8% water. Solvent B was 0.2% formic acid, 19.8% water, and 80% acetonitrile. For the LTQ-FT, a 90 min gradient from 0% to 40% solvent B was used. The mass spectrometers were operated in data-dependent mode to automatically switch between MS and MS/MS scans, essentially as described (21). Survey full scan mass spectra were acquired in the Orbitrap (300 – 1700 *m/z*), following accumulation of 500,000 ions, with a resolution of 60,000 at 400 *m/z*. The top ten most intense ions from the survey scan were isolated and, after the accumulation of 5000 ions, fragmented in the linear ion trap by collisionally induced dissociation (collisional energy 35% and isolation width 2 Da). Precursor ion charge state screening was enabled and all singly charged and unassigned charge states were rejected. A reject mass list with the major streptavidin contaminants was used. The dynamic exclusion list was set with a maximum retention time of 90 s and a relative mass window of 10 ppm.

Survey full scan mass spectra were acquired in the LTQ-FT with an *m/z* of 400 – 1600, following accumulation of 1,000,000 ions, with a resolution of 50,000 at 400 *m/z*. The top seven most intense ions from the survey scan were isolated and, after the accumulation of 5000 ions, fragmented in the linear ion trap by collisionally induced dissociation (collisional energy 35% and isolation width 3 Da). Precursor ion charge state screening was enabled and all singly charged and unassigned charge states were rejected. The dynamic exclusion list was set with a maximum retention time of 60 s and a relative mass window of 10 ppm.

Protein Identification and Quantification – MaxQuant (v. 1.3.0.5) was used to process the Thermo RAW files. All default parameters were used, except LysC was specified as the enzyme and requantify was disabled. Up to 2 missed cleavages were allowed. Met oxidation (+15.9949) and N-terminal acetylation (+42.0106) were specified as variable modifications, and carbamidomethyl cysteine (+57.0125) was specified as a fixed modification. In all Aha labeling experiments, Aha (-4.9863) and L-2,4-diaminobutanoate (-30.9768), a product of reduction of Aha, were specified as variable modifications for Met. In BONCAT experiments, DBCO-sulfo-biotin (+648.2115) was specified as an additional variable modification for Met. Medium (+4.0251) and heavy (+8.0142) lysine labels were specified in all experiments. Mass tolerance for precursor ions and fragment ions were 7 ppm and 0.5 Da, respectively. In pSILAC experiments, multiplicity was set to 3 and light, medium, and heavy peptides were specified. In BONCAT experiments multiplicity was set to 2, specifying only medium and heavy lysine labels. The human database searched (IPI v 3.54) consisted of 75,710 sequences, 262 of which were common contaminants. The database was appended to a decoy database of equal size in MaxQuant. Protein and peptide false discovery rates were fixed at 1% using a target decoy approach. All members of a protein family with shared identifications were reported in one

protein group. Individual evidence ratios were calculated by MaxQuant as the ratio of peak areas. Overall protein ratios were reported as the mean of the individual experiment ratios. Only ratios associated with unique or razor peptides were used for quantification. Annotated spectra for proteins identified by single peptides and detailed information on peptide and protein identifications are available in the supplementary information. We report only proteins quantified by at least two evidences in each set of experiments. In pSILAC experiments comparing Met *vs*. Met, Aha *vs*. Met, and $Aha^{30:1}$ *vs*. Met-treated cultures, we report only proteins quantified in both arrangements of the label swap experiments, and peptides that contained Met were discarded when calculating protein ratios to avoid unreliable quantifications due to unmodified counterpart peptides.

Protein Ratio Statistics – Overall protein ratios and their standard errors were calculated using a hierarchical model combined with bootstrap estimates and pooled variance estimates at the peptide level. Briefly, a global estimate of measurement error is calculated using pooled variance from the protein ratio in each replicate. Next, a hierarchical model of the overall protein ratio is calculated by first calculating the protein replicate ratio as the median of the peptide ratios in each replicate and then calculating the overall protein ratio as the mean of the protein replicate ratios. Finally, the standard error of the overall protein ratio is calculated using a bootstrap procedure where resampling with replacement occurs within the hierarchical model at both the replicate and peptide level and each peptide ratio in the bootstrap procedure is augmented by adding a random “noise” effect drawn from a normal distribution with mean zero and standard deviation equal to the previously calculated global estimate of measurement error. In total, 1,000 bootstrap iterations are performed. The standard error of the overall protein ratio is then calculated as the standard deviation of the bootstrapped overall protein

ratios. Z-tests can then be used to calculate p-values of overall protein ratios with respect to a 1-to-1 ratio.

RESULTS

In pSILAC, metabolic labeling with amino acid isotopologs is used to determine relative rates of protein translation in different cell populations (22). To measure the effect of Aha treatment on protein abundance, cells grown in standard "light" media were split and shifted into isotopically "medium" and "heavy" culture media containing either Met, Aha, or a mixture of Aha and Met (Fig. 2.1A). After labeling times of 24 h in HeLa cell culture, paired medium and heavy cultures were lysed and mixed in a 1:1 ratio. Separate experiments in which both medium and heavy populations were treated with Met served as controls. Proteins were prepared for LC-MS/MS by standard gel separation and in-gel digestion protocols. Raw MS data were processed with the MaxQuant quantitative proteomics software (23).

The extent of replacement of Met by Aha can be controlled by adjusting the relative concentrations of the two amino acids in the culture medium (24). A set of scouting experiments in *E. coli* suggested that treating cells with a 30:1 ratio of Aha to Met might limit the extent of replacement to levels that would allow enrichment of newly synthesized proteins with minimal effects on protein abundance (supplemental Fig. 2.S1 and 2.S2). We then investigated similar conditions, hereafter referred to as $Aha^{30:1}$, for quantitative, time-resolved proteomic analysis of HeLa cells in culture.

We performed pSILAC experiments with 24-h labeling times to determine whether $Aha^{30:1}$ labeling causes differences in protein abundance relative to Met-labeled controls. Of the 1257 quantified proteins produced during the 24-h pulse, only three exhibited H/M ratios significantly different from 1 (Fig. 2.1B). The 78-kDa glucose-regulated protein (Grp78), an ER-

localized chaperone, was upregulated 1.2-fold, while the mannose-P-dolichol utilization defect 1 protein (MPDU1) and S-adenosylmethionine (SAM) synthase isoform type 2 (MAT2A) were upregulated 2.4- and 5.1-fold, respectively. MPDU1 is an ER membrane protein required for use of mannose-P-dolichol in the synthesis of GPI anchors and lipid-linked oligosaccharides (25). SAM synthase catalyzes the conversion of methionine to SAM, an important methyl donor in the cell (26). In these experiments the extent of replacement of Met by Aha was 6% (supplemental Table 2.S1). The ratio of k_{cat}/K_m values for activation of Aha and Met by the *E. coli* methionyl-tRNA synthetase is 1:390, which for the Aha$^{30:1}$ labeling condition predicts a replacement rate of 7.7%, in close agreement with the measured rate (27).

In contrast to the results just described, treatment of HeLa cells with Aha alone (i.e., without added Met) for a 24-h period caused substantial changes in the abundance of many proteins. Statistically significant differences were noted for 362 of the 1001 quantified proteins, 101 of which showed differences greater than 2-fold (Fig. 2.1B, supplemental Table 2.S2).

Gene ontology (GO) analysis suggests that protein folding and translation are among the cellular processes affected most significantly by Aha labeling in the absence of added Met (Fig. 2.2A). The heat shock 70 kDa protein 1A/1B (HSPA1A), a member of the protein folding group, showed the largest increase (9.6-fold) in the Aha-treated samples. Other upregulated heat shock proteins and chaperones included the heat shock 105 kDa protein (HSPH1), endoplasmin (HSP90B1), calnexin (CANX), and several DnaJ homologs (DNAJB11, DNAJA1, and DNAJB1), which were upregulated 2 − 3.5-fold. In the translation group, most of the strongly downregulated proteins were ribosomal proteins, which also constituted 17 of the 19 most downregulated proteins overall.

Among proteins in other GO groups, fatty acid synthase (FASN) showed the largest decrease (−5.6-fold) in Aha-treated cells. FASN produces long-chain fatty acids and is sensitive to

amino acid levels in the cell; deprivation of single essential amino acids, including Met, diminishes FASN abundance (28). Regulation of SAM synthase is also linked to Met abundance, and the type 2 isoform identified here is upregulated in the absence of Met (29). These examples suggest that Met starvation, independent of Aha incorporation, probably contributes to the changes in protein abundance observed in Met-depleted cultures. The extent to which the observed changes are due to Aha incorporation, Met depletion, or a combination of the two, however, is unknown.

A similar representation of the results of labeling of HeLa cells with a 30:1 mixture of Aha and Met (Fig. 2.2B) confirms that light labeling yields reliable measurements of relative protein abundances. The full lists of proteins from $Aha^{30:1}$ and Aha-treated cultures and their quantitative analyses are available in supplemental Table 2.S2.

Because BONCAT provides a convenient means of identifying newly synthesized cellular proteins, we examined the pSILAC experiments for evidence that Aha labeling might give rise to artifacts in protein identification. Peptides containing medium and heavy lysine labels allowed for direct comparison of proteins identified in either Aha- or Met-pulsed cultures from individual pSILAC experiments. For each experiment we determined the total number of proteins identified, the number shared between the sample treated with either Aha or $Aha^{30:1}$ and the Met control, and the number of proteins found only in the sample culture (supplemental Table 2.S1). In each case the number of proteins found only in the sample culture constituted between 0.5% and 6.1% of the total, and no increase was observed in Aha-treated cultures compared to those treated with Met (Fig. 2.3). Thus we find no evidence that Aha labeling, even under conditions of Met depletion, gives rise to artifacts in protein identification.

Both BONCAT and pSILAC can be used to measure changes in the cellular proteome that occur during time windows defined by amino acid pulse-labeling protocols. The key difference

between the two methods is that BONCAT physically separates newly synthesized proteins prior to MS analysis, while pSILAC distinguishes and quantifies new proteins informatically. By combining the BONCAT and pSILAC approaches, newly synthesized proteins produced during a pulse can be both enriched and quantified (Fig. 2.4A). Here we show that quantitative BONCAT can be used as a general approach for monitoring rapid proteomic changes, and we compare the performance of BONCAT and pSILAC for short labeling times.

For short pulse-labeling times, newly synthesized proteins make up a small fraction of the total proteome. We expected that enriching newly synthesized proteins by pulse-labeling with Aha would aid in their detection and quantification by mass spectrometry for short pulse times. We compared the performance of the combined BONCAT-pSILAC approach to that of pSILAC alone at pulse times of 4 h and 30 min in HeLa cell culture. In BONCAT experiments, Aha-labeled proteins were enriched by conjugation to a DBCO-biotin tag via strain-promoted azide-alkyne click chemistry and purification on a streptavidin resin (30). Enriched proteins were prepared for LC-MS/MS as described in the experimental methods.

Using the combined BONCAT-pSILAC method in triplicate experiments, we quantified 1931 and 1529 newly synthesized proteins in HeLa cell cultures within a 4-h window with the Aha and Aha$^{30:1}$ labeling strategies, respectively. Similar pSILAC experiments quantified 589 proteins within the 4-h pulse (Fig. 2.4B). For a pulse time of 30 min, pSILAC quantified only 9 newly synthesized proteins, while BONCAT quantified 1484 and 416 proteins with the Aha and Aha$^{30:1}$ labeling strategies, respectively. Comparisons of protein identifications produced similar results (supplemental Fig. 2.S3). BONCAT-pSILAC experiments yielded H/M protein ratios that were accurate and consistent, showing narrow distributions centered about a value of 1 (Fig. 2.4C). BONCAT replicates were reproducible, sharing an average of 89% of quantified proteins between replicates (supplemental Table 2.S3). Individual BONCAT experiments provided 5 – 7

times more peptide evidences than parallel pSILAC experiments in the 4-h pulse, and 35 – 175 times more evidences in the 30-m pulse experiments (Fig. 2.4D, supplemental Fig. 2.S4). Prior to enrichment, newly synthesized proteins in the 4-h and 30-min experiments made up approximately 6% and 3% of the total proteome, respectively. After BONCAT enrichment, newly synthesized proteins constituted an average of 80 – 83% and 52 – 65% of the total protein in the 4-h and 30-min experiments, respectively (Fig. 2.4E). The list of quantified proteins from the 4-h and 30-min pSILAC and BONCAT experiments is available in supplemental Table 2.S4.

DISCUSSION

Pulse-labeling with Aha allows fast, accurate and sensitive detection of changes in the cellular proteome. Complete replacement of Met by Aha, even for pulses of 24 h in HeLa cell culture, does not significantly change the identities of proteins detected by mass spectrometry, although differences in protein abundance are observed under such conditions. To mitigate the latter effect, we describe an Aha$^{30:1}$ labeling strategy that minimizes perturbations in protein abundance while maintaining a level of labeling that is sufficient for modification by click chemistry and affinity enrichment. The Aha$^{30:1}$ labeling approach is recommended for use in studies in which preservation of natural protein abundances is of the utmost importance. Alternatively, full Aha labeling increases the yield of affinity enrichment and can be used to identify larger numbers of newly synthesized proteins, especially under conditions in which enrichment is more challenging, as in experiments that use short pulse times.

A comparison of the BONCAT and pSILAC methods shows that affinity enrichment of newly synthesized proteins enhances the time resolution of proteomic analysis. A combined BONCAT-pSILAC approach enables acquisition of extensive, quantitative proteomic information within 30 min in HeLa cell culture, a time scale that is inaccessible to isotope labeling techniques

alone. We expect quantitative BONCAT-pSILAC experiments to be especially useful for monitoring proteome dynamics, i.e., for identifying early, middle, and late changes in protein production in response to biological cues.

ACKNOWLEDGMENTS

This work was supported by National Institutes of Health grant NIH RO1 GM062523, the Institute for Collaborative Biotechnologies through grant W911NF-09-0001 from U.S. Army Research Office, the Joseph J. Jacobs Institute for Molecular Engineering for Medicine, the Betty and Gordon Moore Foundation through Grant GBMF775, and the Beckman Institute. Y.J.X. acknowledges funding from the Caltech Summer Undergraduate Research Fellowships (SURF) program. We thank Kai Yuet for providing the *E. coli* KY2 strain and the PEL staff for technical support.

FIGURES

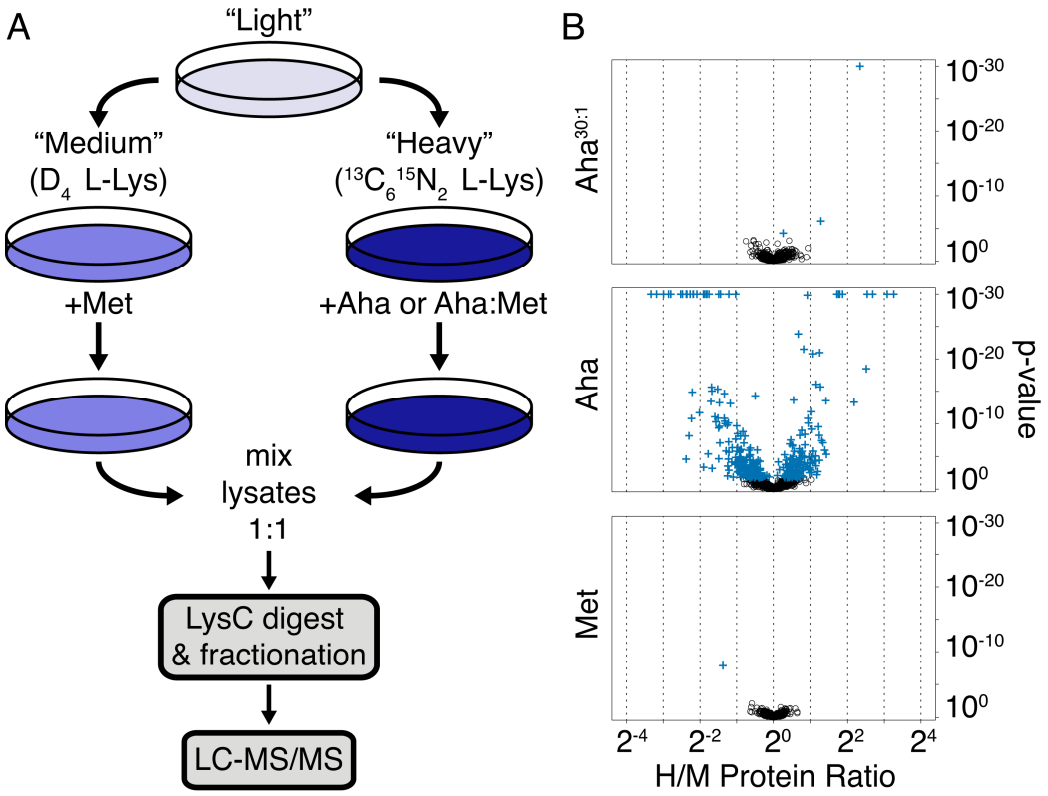


Figure 2.1. Quantifying the effects of Aha labeling on protein abundance in HeLa cell cultures. (A) We used pSILAC to determine the effects of Aha labeling on protein abundance. Incorporation of "medium" and "heavy" isotope-labeled lysine allowed quantification of proteins from cultures treated with Aha or with mixtures of Aha and Met, relative to cultures treated with Met only. (B) Global proteomic effects of Aha labeling in HeLa cell cultures. Quantified proteins are compared for cultures treated with $Aha^{30:1}$ *vs*. Met (top), Aha *vs*. Met (middle), and Met *vs*. Met (bottom). Proteins that show statistically significant differences from an H/M ratio of 1 (Benjamini-Hochberg FDR < 0.05) are marked by the blue cross symbol. H/M ratios greater than 1 indicate increased expression of proteins in the presence of Aha. Proteins with p-values less than 10^{-30} were plotted at 10^{-30}.

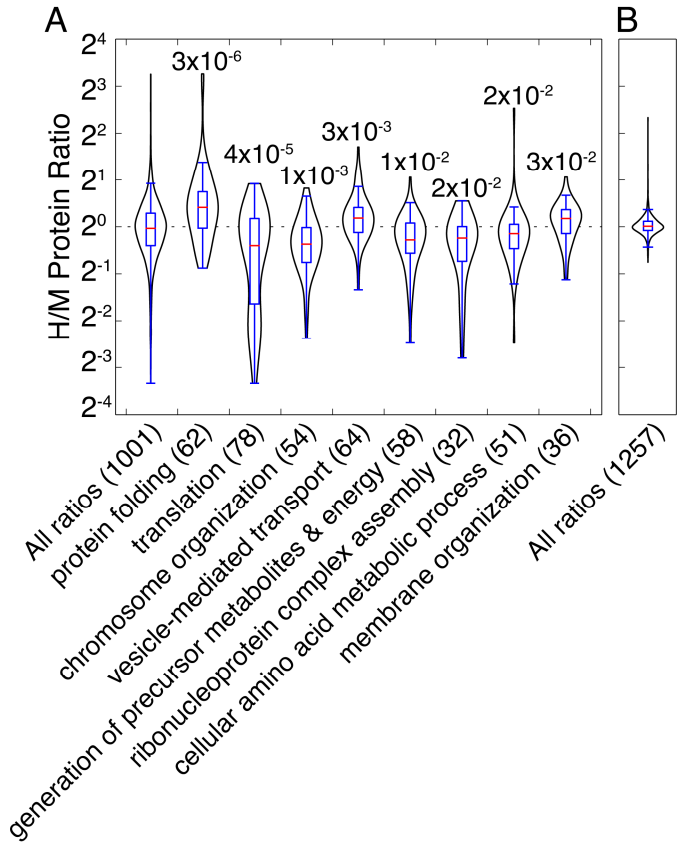


Figure 2.2. Gene ontology analysis of proteins affected by Aha treatment. (A) Quantified proteins from Aha *vs*. Met experiments were mapped to gene ontology groups, and the ratio distributions of the groups were compared to the total protein distribution by a two-sample K–S test ($p < 0.05$). Violin plots of the distributions are shown with p-values and the number of proteins in each group. (B) A similar violin plot for total quantified proteins from Aha$^{30:1}$ experiments confirms the reliability of relative protein abundances measured in such experiments.

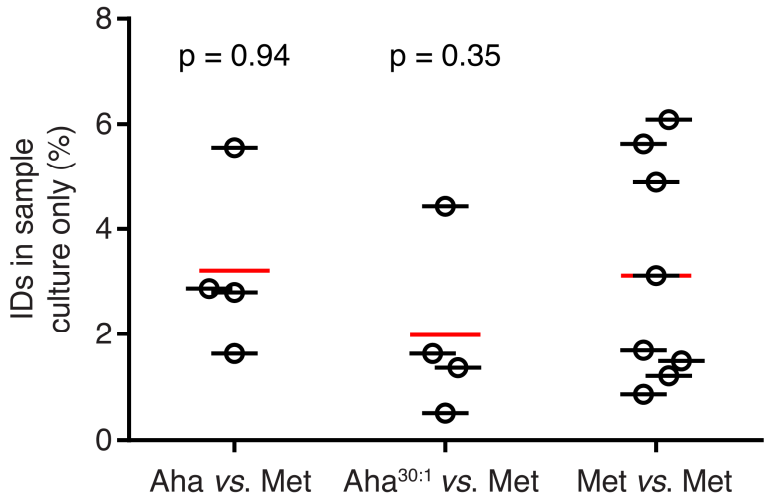


Figure 2.3. Aha labeling does not cause artifacts in protein identification. Proteins identified in sample cultures treated with Aha, Aha$^{30:1}$, or Met are compared to proteins identified in their respective control cultures treated with Met only. Comparisons of protein identifications are made between medium and heavy-labeled proteins from individual pSILAC experiments. The percentage of identifications in the sample culture only are plotted for each of the four replicate experiments, and the average of these values is marked by a red line. Unpaired, two-sample *t*-tests comparing Aha *vs*. Met and Aha$^{30:1}$ *vs*. Met to the Met *vs*. Met control failed to reject the null hypothesis that average values are identical (p-values are shown). Due to the symmetry of Met *vs*. Met control experiments, the number of identification comparisons is effectively doubled.

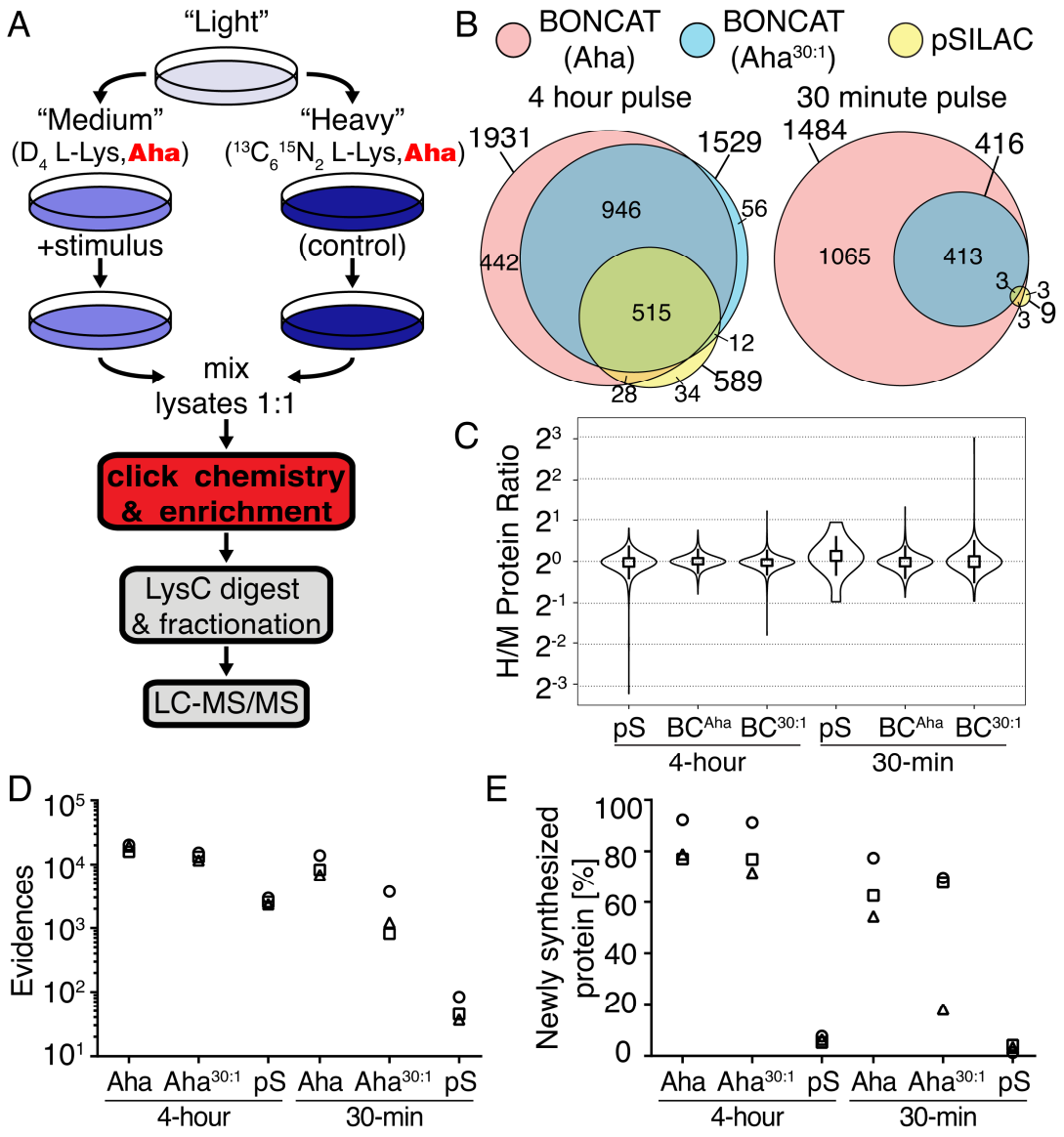


Figure 2.4. Comparison of the BONCAT and pSILAC approaches in HeLa cell culture. (A) Combining BONCAT with pulsed isotope labeling allows enrichment, identification and quantification of newly synthesized proteins. Deviations from a standard pSILAC protocol are shown in red. In a quantitative BONCAT experiment, cultures are treated during the pulse with both lysine isotopologs and Aha. After mixing lysates, Aha-labeled proteins are conjugated to affinity tags, enriched by affinity purification, and analyzed by standard MS protocols. (B)

Quantified proteins from the BONCAT and pSILAC methods are compared for experiments performed during 4-h and 30-min pulses. Venn diagrams show total proteins quantified in each experiment and proteins shared between experiments. (C) BONCAT experiments provide accurate protein quantification. Each violin plot shows a density distribution of H/M protein ratios, a box plot that indicates the span from the 25th to the 75th percentile, and whiskers that extend to 1.5-fold times the inner quartile range from the box edges. (D) The total number of peptide evidences and (E) the percent of newly synthesized protein are plotted for individual BONCAT and pSILAC experiments. BONCAT (BC), pSILAC (pS).

SUPPLEMENTARY FIGURES AND TABLES

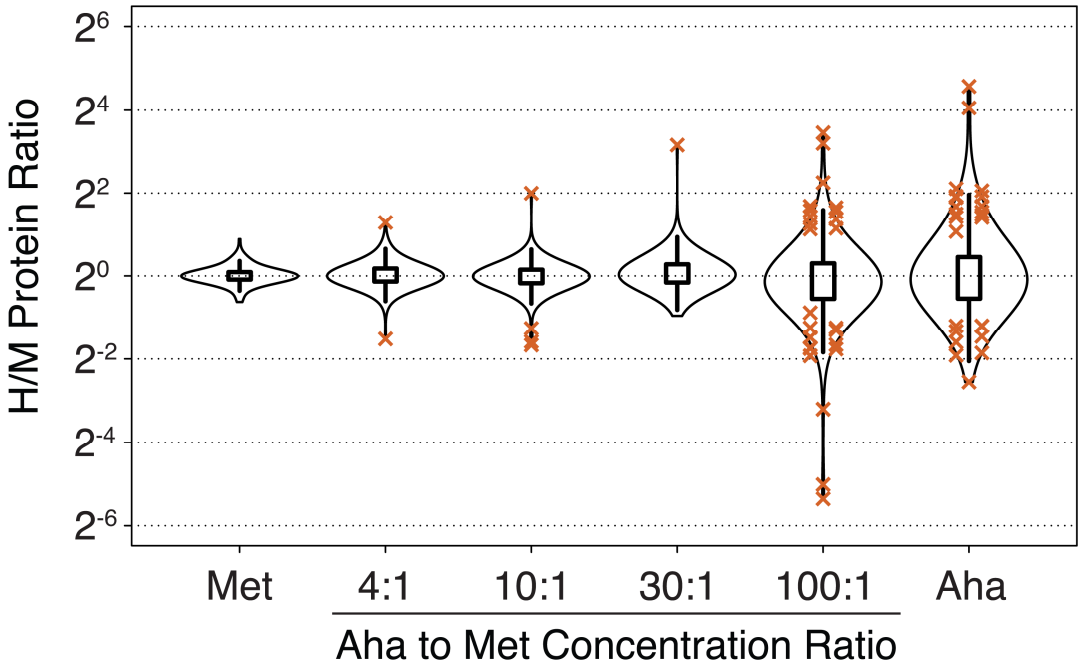


Figure 2.S1. Small-scale pSILAC experiments in *E. coli* showed how labeling conditions affect protein abundances. Experiments followed the scheme outlined in Fig. 2.1A of the manuscript. The total concentration of Aha in each experiment was 1 mM, and the labeling time was 1 hour. Each violin plot shows a density distribution of protein ratios, a box plot that indicates the span from the 25th to the 75th percentile, and whiskers that extend to 1.5-fold times the inner quartile range from the box edges. Proteins that show H/M ratios with statistically significant differences greater than 1.5-fold (Benjamini-Hochberg FDR < 0.05) are represented by the orange symbol x. The 100:1 and Aha-only cultures show increases in the breadth of the distribution and the number of outlier proteins, suggesting that Aha incorporation or Met starvation under these conditions alters protein abundances. Cultures treated with lower ratios of Aha to Met yield H/M distributions that resemble more closely that of the Met control.

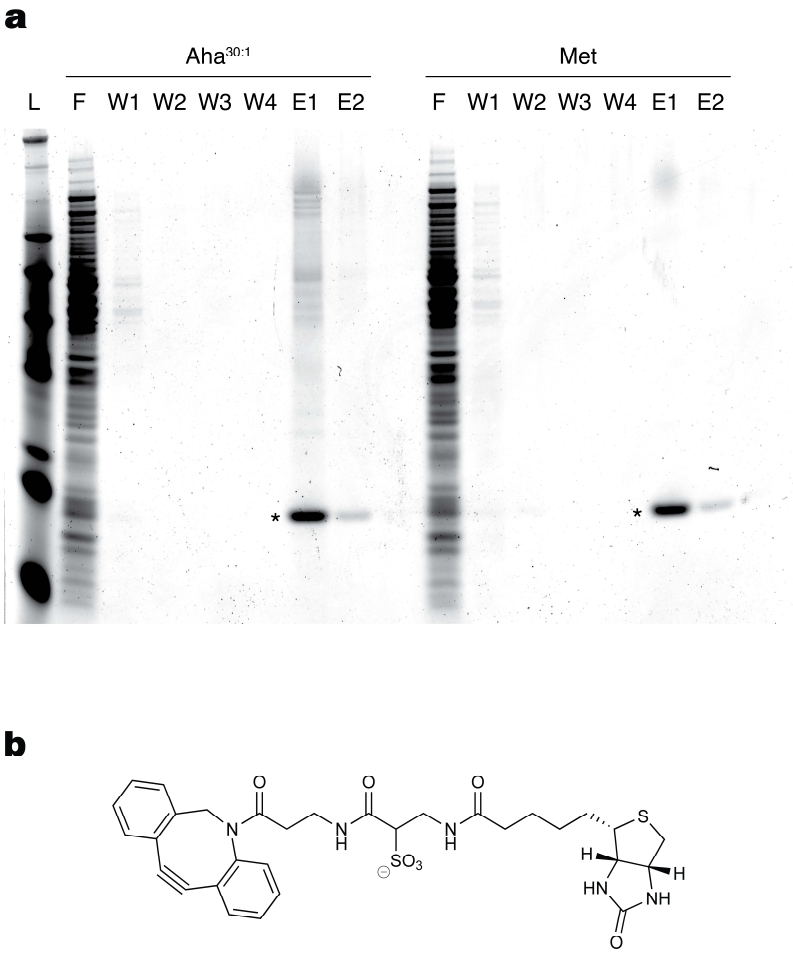


Figure 2.S2. Aha$^{30:1}$-labeled proteins contain sufficient Aha for selective tagging and enrichment by click chemistry. (a) Aha-labeled proteins were conjugated to a DBCO-biotin tag by strain-promoted click chemistry and purified on streptavidin resin. Met-labeled samples show no proteins in the elution fractions. Lane abbreviations are defined as follows: L – SeeBlue® ladder, F – flowthrough, W – washes, and E – elutions. The symbol * marks streptavidin protein. (b) Structure of the DBCO-sulfo-biotin tag used for affinity enrichment.

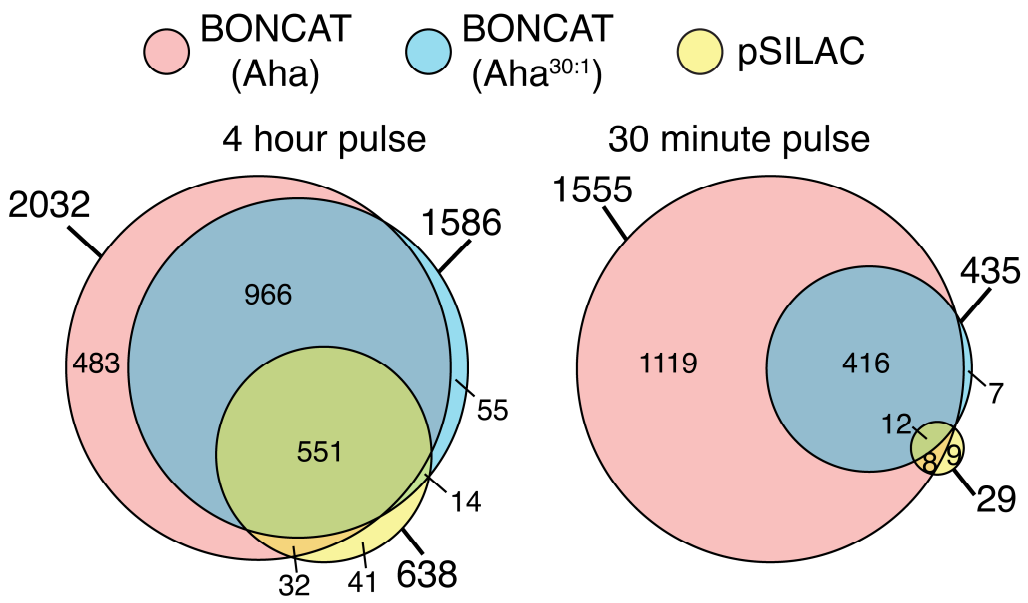


Figure 2.S3. Comparison of newly synthesized proteins identified in BONCAT-pSILAC and pSILAC experiments in HeLa cell culture for 4-h and 30-min pulse times. We required newly synthesized proteins to be identified by at least two evidences of production during the amino acid pulse, i.e. a medium or heavy lysine or an identified SILAC pair or triplet. This figure is similar to that shown in Figure 2.4 of the main text, but includes all identified proteins, even if they could not be quantified.

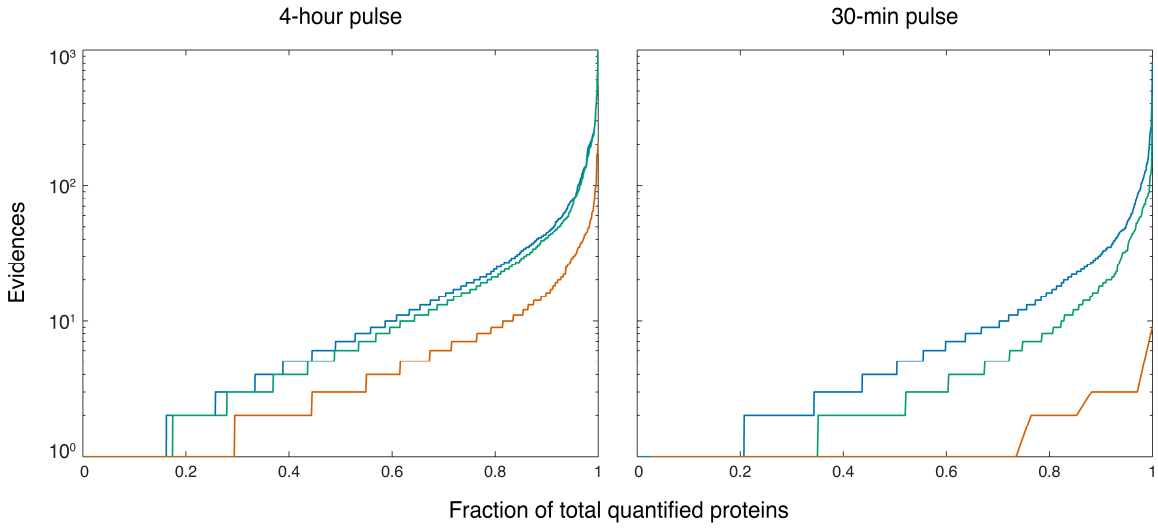


Figure 2.S4. For short pulse-labeling times, BONCAT provides more quantitative information per protein than pSILAC. For both 4 hour and 30 min pulses, BONCAT experiments using Aha (blue line) or Aha$^{30:1}$ (green line) result in substantially more quantified evidences per protein than pSILAC alone (orange line).

Table 2.S1. Protein identifications and Aha incorporation in HeLa cell 24-hour pSILAC experiments. Proteins identified in sample cultures treated with Aha, Aha$^{30:1}$, or Met are compared to proteins in their respective control cultures treated with Met only. Comparisons of protein identifications are made between medium and heavy-labeled proteins from individual pSILAC experiments. We calculated Aha incorporation rates in the pSILAC experiments by comparing MS intensities of Met-containing peptides between paired medium and heavy cultures. As Aha incorporation in a peptide increases, the corresponding Met-containing peptide signal intensity decreases. Therefore, changes in the H/M ratio of Met-containing peptides can be used to infer the extent of Aha replacement. The median H/M ratio of peptide evidences in each experiment provides a robust estimate of the extent of Aha incorporation, and 95th percentile intervals are shown. Because this calculation relies on accurate H/M ratios of Met-containing peptides, it is not reliable for full Aha labeling experiments, in which Aha incorporation is very high.

	Protein IDs				Protein IDs not shared				Aha Incorporation		
Experiment	Medium (K+4)	Heavy (K+8)	Total	Shared	Medium		Heavy		Median	2.5th PCTL	97.5th PCTL
30:1(K+4) vs. Met #1	982	1025	1041	966	16	1.63%	59	5.76%	5.36%	4.02%	6.43%
30:1(K+4) vs. Met #2	1382	1450	1457	1375	7	0.51%	75	5.17%	6.89%	5.45%	7.97%
30:1(K+8) vs. Met #1	981	1016	1026	971	10	1.02%	45	4.43%	5.25%	4.04%	6.50%
30:1(K+8) vs. Met #2	1410	1395	1429	1376	34	2.41%	19	1.36%	5.10%	4.14%	6.07%
Aha(K+4) vs. Met #1	714	891	911	694	20	2.80%	197	22.11%			
Aha(K+4) vs. Met #2	1045	1238	1255	1028	17	1.63%	210	16.96%			
Aha(K+8) vs. Met #1	821	809	866	764	57	6.94%	45	5.56%			
Aha(K+8) vs. Met #2	1229	1221	1264	1186	43	3.50%	35	2.87%			
Met vs. Met #1	1008	1044	1059	993	15	1.49%	51	4.89%			
Met vs. Met #2	1042	1100	1109	1033	9	0.86%	67	6.09%			
Met vs. Met #3	1423	1444	1468	1399	24	1.69%	45	3.12%			
Met vs. Met #4	1154	1208	1222	1140	14	1.21%	68	5.63%			

Table 2.S2. Proteins and peptides data for 24-h pSILAC experiments in HeLa cell culture comparing cultures pulsed with Aha, Aha$^{30:1}$, and Met. This table is available as a downloadable excel file in the online version of this thesis.

Table 2.S3. Percent of shared protein quantifications from experimental replicates in the 4-hour and 30-min BONCAT and pSILAC experiments.

Pulse length	Replicates compared	pSILAC	BONCAT (Aha)	BONCAT ($Aha^{30:1}$)
4 h	Exp. 1 & 2	70.8%	92.2%	86.4%
	Exp. 1 & 3	76.6%	90.3%	88.4%
	Exp. 2 & 3	80.8%	90.2%	91.0%
30 min	Exp. 1 & 2	25.7%	93.6%	92.5%
	Exp. 1 & 3	25.0%	93.5%	89.3%
	Exp. 2 & 3	42.9%	89.2%	74.0%

Table 2.S4. Proteins and peptides data for BONCAT and pSILAC experiments in HeLa cell culture with 4-hour and 30-min pulses. This table is available as a downloadable excel file in the online version of this thesis.

REFERENCES

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

2. Dieterich, D. C., Lee, J. J., Link, A. J., Graumann, J., Tirrell, D. A., and Schuman, E. M. (2007) Labeling, detection and identification of newly synthesized proteomes with bioorthogonal non-canonical amino-acid tagging. *Nat. Protoc.* **2**, 532–540

3. Szychowski, J., Mahdavi, A., Hodas, J. J. L., Bagert, J. D., Ngo, J. T., Landgraf, P., Dieterich, D. C., Schuman, E. M., and Tirrell, D. A. (2010) Cleavable Biotin Probes for Labeling of Biomolecules via Azide-Alkyne Cycloaddition. *J. Am. Chem. Soc.* **132**, 18351–18360

4. Tcherkezian, J., Brittis, P. A., Thomas, F., Roux, P. P., and Flanagan, J. G. (2010) Transmembrane Receptor DCC Associates with Protein Synthesis Machinery and Regulates Translation. *Cell* **141**, 632–644

5. Yoon, B. C., Jung, H., Dwivedy, A., Hare, C. M. O., Zivraj, K. H., and Holt, C. E. (2011) Local Translation of Extranuclear Lamin B Promotes Axon Maintenance. *Cell* **148**, 752–764

6. Deal, R. B., Henikoff, J. G., and Henikoff, S. (2010) Genome-wide kinetics of nucleosome turnover determined by metabolic labeling of histones. *Science (80-.).* **328**, 1161–1164

7. Zhang, M. M., Tsou, L. K., Charron, G., Raghavan, A. S., and Hang, H. C. (2010) Tandem fluorescence imaging of dynamic S-acylation and protein turnover. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 8627–8632

8. Ouellette, S. P., Dorsey, F. C., Moshiach, S., Cleveland, J. L., and Carabeo, R. A. (2011) Chlamydia species-dependent differences in the growth requirement for lysosomes. *PLoS One* **6**, e16783

9. Choi, K.-Y. G., Lippert, D. N. D., Ezzatti, P., and Mookherjee, N. (2012) Defining TNF-α and IL-1β induced nascent proteins: Combining bio-orthogonal non-canonical amino acid tagging and proteomics. *J. Immunol. Methods* **382**, 189–195

10. Melemedjian, O. K., Asiedu, M. N., Tillu, D. V, Peebles, K. A., Yan, J., Ertz, N., Dussor, G. O., and Price, T. J. (2010) IL-6- and NGF-induced rapid control of protein synthesis and nociceptive plasticity via convergent signaling to the eIF4F complex. *J. Neurosci.* **30**, 15113–15123

11. Dieterich, D. C., Hodas, J. J. L., Gouzer, G., Shadrin, I. Y., Ngo, J. T., Triller, A., Tirrell, D. A., and Schuman, E. M. (2010) In situ visualization and dynamics of newly synthesized proteins in rat hippocampal neurons. *Nat. Neurosci.* **13**, 897–905

12. Hodas, J. J. L., Nehring, A., Höche, N., Sweredoski, M. J., Pielot, R., Hess, S., Tirrell, D. A., Dieterich, D. C., and Schuman, E. M. (2012) Dopaminergic modulation of the hippocampal neuropil proteome identified by bio-orthogonal non-canonical amino-acid tagging (BONCAT). *Proteomics* **12**, 2464–2476

13. Hinz, F. I., Dieterich, D. C., Tirrell, D. A., and Schuman, E. M. (2012) Noncanonical amino acid labeling in vivo to visualize and affinity purify newly synthesized proteins in larval zebrafish. *ACS Chem. Neurosci.* **3**, 40–49

14. Eichelbaum, K., Winter, M., Diaz, M. B., Herzig, S., and Krijgsveld, J. (2012) Selective enrichment of newly synthesized proteins for quantitative secretome analysis. *Nat. Biotechnol.* **30**, 984–990

15. Somasekharan, S. P., Stoynov, N., Rotblat, B., Leprivier, G., Galpin, J. D., Ahern, C. a, Foster, L. J., and Sorensen, P. H. B. (2012) Identification and quantification of newly synthesized proteins translationally regulated by YB-1 using a novel Click-SILAC approach. *J. Proteomics* **77**, e1–e10

16. Howden, A. J. M., Geoghegan, V., Katsch, K., Efstathiou, G., Bhushan, B., Boutureira, O., Thomas, B., Trudgian, D. C., Kessler, B. M., Dieterich, D. C., Davis, B. G., and Acuto, O. (2013) QuaNCAT: quantitating proteome dynamics in primary cells. *Nat. Methods* **10**, 343–346

17. Link, A. J., Vink, M. K. S., and Tirrell, D. A. (2007) Synthesis of the functionalizable methionine surrogate azidohomoalanine using Boc-homoserine as precursor. *Nat. Protoc.* **2**, 1884–1887

18. Van Geel, R., Pruijn, G. J. M., van Delft, F. L., and Boelens, W. C. (2012) Preventing thiol-yne addition improves the specificity of strain-promoted azide-alkyne cycloaddition. *Bioconjugate Chem.* **23**, 392–398

19. Debets, M. F., van Berkel, S. S., Schoffelen, S., Rutjes, F. P. J. T., van Hest, J. C. M., and van Delft, F. L. (2010) Aza-dibenzocyclooctynes for fast and efficient enzyme PEGylation via copper-free (3+2) cycloaddition. *Chem. Commun.* **46**, 97–99

20. Rappsilber, J., Mann, M., and Ishihama, Y. (2007) Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* **2**, 1896–1906

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

22. Schwanhäusser, B., Gossen, M., Dittmar, G., and Selbach, M. (2009) Global analysis of cellular protein translation by pulsed SILAC. *Proteomics* **9**, 205–209

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

24. Nowatzki, P. J., Franck, C., Maskarinec, S. A., Ravichandran, G., and Tirrell, D. A. (2008) Mechanically Tunable Thin Films of Photosensitive Artificial Proteins: Preparation and Characterization by Nanoindentation. *Macromolecules* **41**, 1839–1845

25. Hirata, T., Fujita, M., Kanzawa, N., Murakami, Y., Maeda, Y., and Kinoshita, T. (2013) Glycosylphosphatidylinositol mannosyltransferase II is the rate-limiting enzyme in glycosylphosphatidylinositol biosynthesis under limited dolichol-phosphate mannose availability. *J. Biochem.* **154**, 257–264

26. Markham, G. D., Hafner, E. W., Tabor, C. W., and Tabor, H. (1980) S-Adenosylmethionine synthetase from Escherichia coli. *J. Biol. Chem.* **255**, 9082–9092

27. Kiick, K. L., Saxon, E., Tirrell, D. A., and Bertozzi, C. R. (2002) Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 19–24

28. Dudek, S. M., and Semenkovich, C. F. (1995) Essential Amino Acids Regulate Fatty Acid Synthase Expression through an Uncharged Transfer RNA-dependent Mechanism. *J. Biol. Chem.* **270**, 29323–29329

29. Martínez-Chantar, M. L., Latasa, M. U., Varela-Rey, M., Lu, S. C., García-Trevijano, E. R., Mato, J. M., and Avila, M. a (2003) L-methionine availability regulates expression of the methionine adenosyltransferase 2A gene in human hepatocarcinoma cells. *J. Biol. Chem.* **278**, 19885–90

30. Sletten, E. M., and Bertozzi, C. R. (2011) Bioorthogonal Reactions. *Acc. Chem. Res.* **44**, 666–676

CHAPTER 3

A QUANTITATIVE PROTEOMICS APPROACH FOR DETERMINING THE FUNCTION OF BACTERIAL SMALL RNA REGULATORS

Abstract

Bacterial small RNA regulators have been identified in many species and play critical regulatory roles in the cell. Currently, the majority of identified sRNAs remain uncharacterized and the standard approaches for identifying sRNA targets focus on measuring transcript, not protein, abundances. Here, we introduce a proteomics method for determining bacterial sRNA function using bio-orthogonal non-canonical amino acid tagging (BONCAT). Using this approach, we measured the proteome-wide effects of the expression of the sRNA regulator CyaR on the time scale of minutes. We identify and confirm three new CyaR targets, expanding the functional role of CyaR to include new roles in carbon metabolism, osmoregulation and transcriptional regulation.

Introduction

Small RNA (sRNA) regulators in bacteria play critical regulatory roles in a variety of cellular processes, including membrane homeostasis, metabolism, quorum sensing, and virulence.[1] sRNAs have been found in a wide range of bacteria and are prevalent; in E. coli, for example, experimentally validated sRNAs comprise about 2% of all coding. Whole-genome expression profiling via microarray or deep sequencing have proven effective in the identification of sRNA regulators, however, determining sRNA function has proven more difficult. While some sRNAs bind and act upon target proteins directly, most modulate gene expression via base-pairing with target mRNA, and many sRNAs have been identified that regulate multiple targets.[2] sRNA target hybridization is often short (≤25 bp) and imperfect, making computational prediction challenging. As a result, experimental methods have proven necessary for comprehensively characterizing sRNA function.

The most common approach for identifying sRNA targets involves a brief expression of the sRNA followed by transcriptome analysis. The short expression time reduces off-target effects, focusing analysis toward genes that are more likely to be primary sRNA targets. This approach, however, is limited in that it can only identify targets that exhibit changes in transcript abundance, although sRNA regulatory mechanisms have been reported that modify translation rates independent of transcript abundance.[3,4] A proteomic approach for studying sRNA regulator function would be ideal; however, measuring proteome-wide changes in translation over very short time scales is difficult with standard quantitative proteomics approaches.

Bio-orthogonal non-canonical amino acid tagging (BONCAT) is a chemical biology tool that allows the enrichment and identification of newly synthesized cellular proteins that are produced in response to a biological stimulus of interest.[5,6] In the standard BONCAT protocol, a

cell culture is treated with the methionine (Met) surrogate L-azidohomoalanine (Aha), which replaces Met in protein synthesis (Figure 3.1A).[7] Aha-labeled proteins are then selectively conjugated to affinity tags via copper-catalyzed or strain-promoted azide-alkyne click chemistry, enriched by affinity chromatography, and identified by LC-MS/MS (Figure 3.1B). Because Aha can be introduced to cells in a well-defined pulse, BONCAT provides excellent temporal resolution for measurements of protein synthesis. We showed previously that combining BONCAT with stable isotope labeling with amino acids in cell culture (SILAC) can quantitatively measure the nascent proteome with a temporal resolution inaccessible to stable isotope labeling alone.[8]

In this work we use BONCAT to quantify protein translation of the *E. coli* proteome in response to expression of the sRNA regulator CyaR (cyclic AMP-activated RNA) on the time scale of minutes (Figure 3.1B). CyaR is a Crp-activated, Hfq-dependent sRNA regulator that is expressed during late-log and stationary phase growth.[9,10] CyaR is highly conserved in enterobacterial species and regulates proteins involved in membrane homeostasis, nitrogen assimilation, carbon metabolism, and quorum sensing.[9,11–13] These functions were determined via computational target prediction and transcriptomics experiments; no proteomics approaches, however, have been used to study CyaR regulation.[9,11–13]

Results and Discussion

In order to determine the time scales at which we could perform BONCAT experiments, we assessed levels of CyaR expression and Aha incorporation at several time points. Expression of CyaR under plac control (pNRD405) in a Δ*cyaR* strain (NRD359) was detectable within 2 min and reached peak expression in 10 min after IPTG induction (Figure 3.1C). To determine whether Aha could probe nascent protein synthesis at these times, we treated *E. coli* cultures with Aha

and assessed metabolic incorporation by conjugating Aha-labeled proteins to a fluorescent dye. Visualization of tagged proteins by in-gel fluorescence showed that Aha-labeled proteins could be readily detected 5–15 min after Aha treatment (Figure 3.1D).

Based on these results, we designed a combined BONCAT-SILAC experiment to measure the effects of CyaR expression on protein translation over a period of 15 min (Figure 3.1B, 3.S1A). *E. coli* strain JDB3 (Δ*cyaR* Δ*lysA*), with a vector either empty (pBRplac) or carrying *cyaR* (pNRD405) was grown in minimal media with either "light" or "heavy" SILAC labels. During mid-log growth both cultures were treated with IPTG and Aha to simultaneously induce CyaR expression and initiate labeling of protein synthesis. After 15 min, protein synthesis was halted by addition of chloramphenicol. Paired "light" and "heavy" cultures were lysed and mixed at equal protein concentration, and Aha-labeled proteins were conjugated to an alkyne-biotin tag with an acid-cleavable dialkoxydiphenylsilane linker via copper-catalyzed azide-alkyne click chemistry (Figure 3.1A).[14] Tagged proteins were enriched with a streptavidin resin and subsequently analyzed by LC-MS/MS (Figure 3.S1B).

In four replicate experiments, including two SILAC label swap experiments, we collected a total of 147,328 MS/MS spectra from 6271 peptides, identifying 1176 proteins produced during the 15-min pulse (Table 3.S1). Based on spectral counts, we estimated that Met replacement with Aha was low (1.4 Aha residues per enriched protein), which minimizes potential perturbation due to amino acid substitution (Figure 3.S2).[8] We measured significant difference in 26 proteins in response to CyaR expression (Figure 3.2A, Table 3.1). Regulated proteins were well-correlated between label swap experiments (Figure 3.S3A). Outlier protein quantifications from Aha-containing peptides, which provide direct evidence of protein synthesis during the pulse, correlated with quantifications from non-Aha-containing peptides, providing evidence that enriched proteins are indeed nascent (Figure 3.S3B).

Outlier proteins were associated with functional groups involving protein folding, in which all proteins were down-regulated, molecular transport, in which proteins were both up- and down-regulated, and metabolism, in which most proteins were up-regulated (Figure 3.3B). These proteins were enriched in functional associations (40 observed, 6.4 expected), largely due to the dense interaction network of proteins involved in protein folding. Interestingly, several of the outlier genes (*raiA*, *groL*, *groS*, *clpB*, *htpG*, *yacL*, *malE*, *mgtA*, *dnaK*, and *gcd*) are also regulated by cAMP-Crp, but in a direction opposite to that CyaR regulation.[15–18] CyaR may oppose Crp regulation in these co-regulated genes as a way to finely tune the cellular response to catabolite repression.

Included in the list of outlier proteins were three known directly regulated CyaR targets: OmpX, NadE, and LuxS.[9] Two more recently identified CyaR targets, PtsI and SdhA, were identified but were not affected by CyaR expression (Table 3.S1).[13] The remaining known target proteins, YqaE and YobF, were not identified in BONCAT experiments, possibly due to their small size (< 55 amino acids).[9,13] Additionally, YqaE contains a single, N-terminal, Met residue which is predicted to be cleaved by methionine aminopeptidase (MetAP).[19] Because MetAP also cleaves N-terminal Aha,[20] we would not expect to detect YqaE after BONCAT enrichment; about 1% of the E. coli proteome would be difficult to detect in Aha-based enrichments for similar reasons. The RNA polymerase sigma factor RpoS, a master regulator of the general stress response, was down-regulated in response to CyaR; this has been observed previously and explained by indirect regulatory mechanisms.[13,21]

We next determined whether outlier proteins are under direct regulatory control by CyaR. We created translational Gfp fusions of each putative target (5' UTR up to 10th codon) and measured protein expression in the presence of CyaR or a control RNA.[22,23] Importantly, the target-Gfp fusions are expressed by an orthogonal promoter, which requires regulation to be

post-transcriptional. Using this approach, we confirmed three additional targets as directly regulated by CyaR: Gcd, MscL, and YacL (Figure 3.3A, 3.S4). Surprisingly, we measured up-regulation of the MalE-Gfp fusion, which was inconsistent with our proteomics results. This suggests that CyaR interacts with the *malE* 5' UTR, but that the truncation used in this assay does not accurately represent this interaction. We next measured transcript levels of these genes after CyaR expression by quantitative PCR. Regulation of transcript levels of *gcd*, *mscL*, and *yacL* changed in ways that were consistent with BONCAT and Gfp-fusion measurements (Figure 3.3B).

The *gcd* mRNA encodes the membrane-bound quinoprotein glucose dehydrogenase (mGDH), which oxidizes D-glucose in the periplasm and feeds electrons into the respiratory chain.[24] For activity, mGDH requires the cofactor pyrroloquinoline quinone, which *E. coli* do not produce but do chemotax toward, suggesting that mGDH functions under more complex and heterogenous environmental conditions.[25–27] Transcription of *gcd* is directly regulated by Crp in a direction opposite to that of CyaR, which itself is activated by Crp. This regulatory structure, when comprised of a transcription factor and a sRNA, is a rare form called a mixed incoherent feedforward loop, which can achieve distinct regulatory effects that may be advantageous for precisely controlling cell behavior.[28,29] YacL is a 15 kDa, non-essential, acidic protein of unknown function.[30] Several studies have shown that YacL interacts closely with the RNA polymerase core and accessory factors, suggesting a role in transcriptional regulation.[31–33] The last identified target is the mechanosensitive channel of large conductance (MscL), the largest of seven mechanosensitive channels in *E. coli*.[34] MscL senses tension in the lipid membrane caused by osmotic down-shock and, at forces close to that which would cause lysis, opens a pore in the membrane to dissipate the osmotic gradient.[35] Overall, these results broaden the functional role

of CyaR to include new roles in carbon metabolism, osmoregulation and transcriptional regulation.

In summary, we introduced BONCAT as a new approach toward measuring sRNA function and detecting sRNA targets in bacteria. Because BONCAT offers enhanced temporal resolution compared to alternative proteomics techniques, we were able to measure alterations in protein synthesis in response to CyaR expression at the short time scales necessary for probing direct sRNA target regulation. Previous work using a transcriptomics approach identified four directly-regulated CyaR targets (*ompX*, *nadE*, *luxS*, and *yqaE*).[9] Improved methodologies in sRNA target prediction then led to the discovery of three additional CyaR targets (*ptsI*, *sdhA*, and *yobF*).[13] We identified three new CyaR targets (*gcd*, *mscL*, and *yacL*), including the first up-regulated CyaR target (*gcd*). Taken together, this and previous studies teach us that, in order to comprehensively characterize sRNA function, multiple methodological approaches should be used, including computational prediction, transcriptomics, and proteomics.

Materials and Methods

Cell culture. *E. coli* cultures were grown in a shaking incubator at 37 °C and 250 rpm in either LB medium or M9 minimal medium supplemented with 0.4 % (w/v) glucose and 0.08 % (w/v) 19 amino acids (−Met) with the appropriate antibiotic selection markers. For proteomics experiments, overnight cultures of *E. coli* strain JDB3 (Δ*cyaR* Δ*lysA*) with either pBRplac (control RNA) or pNRD405 (CyaR) were used to inoculate 20 ml cultures in minimal media at an OD_{600} of 0.008. Minimal media was Met-free and included either “light” L-Lysine or “heavy” L-Lysine (U-$^{13}C_6$ U-$^{15}N_2$ L-Lys, Cambridge Isotope Laboratories) (Figure 3.S1A). BONCAT experiments were performed with four replicates, two of which were label swap experiments. At an OD_{600} of 0.5, gene expression was induced and metabolic labeling was initiated by simultaneously adding

IPTG (100 μM) and Aha (1 mM). Aha was synthesized as described previously.[36] After a 15-min Aha pulse, protein synthesis was quenched by adding chloramphenicol (Sigma) to a final concentration of 100 μg/ml. Cells were collected by centrifugation and stored at -80 °C. The *E. coli* strains used in this study are listed in Table 3.S2.

DNA manipulations and mutant construction. *E. coli* DH10B was used for all cloning procedures. Restriction enzymes, shrimp alkaline phosphatase, T4 DNA ligase, and DNA polymerase were purchased from New England Biolabs. Plasmids were produced as described in Table 3.S3 using the primers listed in Table 3.S4, which were purchased from Integrated DNA technologies. *E. coli* JDB3 (Δ*cyaR* Δ*lysA*), a lysine auxotroph produced for efficient incorporation of SILAC labels, was created by λ Red recombineering using the plasmids and primers in Tables 3.S3 and 3.S4, respectively.[37]

BONCAT enrichment. Pelleted cells were resuspended in 1% SDS in PBS and lysed by heating at 90 °C for 10 min. Lysates were cleared by centrifugation, and the protein concentrations of lysate supernatants were determined with the BCA Protein Assay Kit (Life Technologies). Paired "light" and "heavy" cultures were mixed at equal concentrations. Alkyne-DADPS-biotin tag (0.1 mM) was conjugated to Aha-labeled proteins via azide-alkyne click chemistry as described in Hong et al.[38] Reactions were allowed to progress for 4 hr at room temperature. Alkyne-DADPS-biotin tag was synthesized as described previously.[14] Proteins were concentrated and purified by acetone precipitation and solubilized in 2% SDS in PBS. Solutions were diluted to 0.15% SDS in PBS, and proteins were captured by incubating with Streptavidin UltraLink Resin (Pierce) for 30 min at room temperature. Bound proteins were washed with 35 column volumes of 1% SDS in PBS and 10 column volumes of 0.1% SDS in ddH_2O. Alkyne-DADPS-biotin tag was cleaved by

incubating the resin in 5% formic acid in 0.1% SDS in ddH_2O for 1 hr at room temperature. Proteins were then washed with five column volumes of 0.1% SDS in ddH_2O, during which proteins remained bound, and subsequently eluted into fractions with 1% SDS in PBS. Protein enrichment was confirmed by SDS-PAGE, and fractions containing eluted protein were concentrated using three kDa MWCO spin filters (Amicon).

In-gel digestion. Concentrated proteins were separated on precast 4–12% polyacrylamide gels (Life Technologies), briefly stained with Colloidal Blue (Life Technologies), and washed with ddH_2O. Gel lanes were cut into 8 pieces and proteins were destained, reduced, alkylated, digested with LysC (Mako), and extracted as described in Bagert et al.[8] Extracted peptides were desalted using custom-packed C_{18} stage tip columns, lyophilized, and resuspended in 0.1% formic acid (Sigma).

Liquid Chromatography-Mass Spectrometry Analysis. Liquid chromatography-mass spectrometry experiments were performed on an EASY-nLC-Orbitrap mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) as described previously with the following specifications.[39] Peptides were loaded over 20 min at a flow rate of 500 nL/min onto a 16 cm analytical HPLC column (75 µm ID) packed in-house with 3 µm ReproSil-Pur C_{18}AQ resin (120 Å pore size, Dr. Maisch, Ammerbuch, Germany), which was enclosed in a column heater operating at 30 °C. Solvent A was 97.8% H_2O, 2% ACN, and 0.2% formic acid; solvent B was 19.8% H_2O, 80% ACN, and 0.2% formic acid. Peptides were separated with gradients of 0–30% solvent B (50 min), 30–100% B (1 min), and 100% B (8 min) at a flow rate of 350 nL/min. The Orbitrap was operated in data-dependent mode to switch automatically between survey full scans (m/z=300–1700) in

the Orbitrap and 10 CID MS/MS scans in the linear ion trap. CID was performed with helium at a normalized collision energy and activation time of 35% and 30 msec, respectively.

Protein and Quantification and Ratio Statistics. Thermo RAW files were processed by MaxQuant (v. 1.3.0.5) with the reference proteome for *E. coli* MG1655 (proteome ID UP000000625) and using default parameters with a few modifications. LysC was set as the enzyme and both re-quantify and match between runs were enabled. Multiplicity was set to two and the light and heavy (+8.0142) Lys labels specified. Variable modifications for Met were Aha (-4.9863), L-2,4-diaminobutanoate (-30.9768), a product of Aha reduction, alkyne-DADPS (+835.4300), and 5-hexyn-1-ol (+93.0868), a product of alkyne-DADPS cleavage, and were included in protein quantification. In post-processing analysis, only proteins with at least one evidence in both types of label swap experiments were considered. Individual protein ratio standard errors were calculated using both pooled variances and bootstrap statistical methods as described previously.[8] *p*-values were calculated using a Z-test for assessing whether ratios were significantly different from a value of 1; *p*-values were adjusted for multiple hypothesis testing using the Benjamini and Hochberg method.[40]

Northern Blot. Total RNA was extracted from cell cultures using the mirVana miRNA Isolation Kit with phenol (Ambion). RNA was separated using precast TBE-Urea Gels (Life Technologies), transferred to a nylon membrane (Life Technologies), and fixed with ultraviolet light (254 nm, 2 min). RNA was hybridized with fluorescent DNA probes (Integrated DNA Technologies) (Table 3.S4) at a concentration of 2.5 nM in ULTRAhyb Hybridization Buffer (Ambion), following the manufacturer's protocols. Blots were visualized on a Typhoon TRIO variable mode imager (GE Healthcare).

Fluorescence assays. Direct sRNA target regulation assays were performed in triplicate as described previously using the plasmids listed in Table 3.S3.[22,23] Overnight cultures of *E. coli* DH10B with one plasmid expressing either CyaR or a control RNA and a second plasmid expressing a putative target fused to the superfolder variant of *gfp* were used to inoculate cultures in minimal media at an OD_{600} of 0.03. Cultures were grown to an OD_{600} of 0.5, pelleted, and resuspended in 1X TBS. Each culture was diluted 100-fold into 1X TBS with 0.5 μM SYTO 62 red fluorescent nucleic acid stain (Life Technologies) and allowed to incubate for 30 min. Red fluorescence from SYTO 62-stained cells (640 nm ex, 675/25 nm em) was thresholded to aid in cell detection and Gfp fluorescence (488 nm ex, 533/30 nm em) of 20,000 events was measured for each sample on a BD Accuri C6 flow cytometer. Triplicate samples were averaged and normalized by their non-fluorescent negative controls to obtain relative fluorescent values.

Quantitative real-time PCR. qRT-PCR was used to measure target mRNA abundances in *E. coli* JDB3 (Δ*cyaR* Δ*lysA*) cultures 15 min after IPTG-induction of either CyaR (pNRD405) or a control RNA (pBRplac). Isolated RNA from triplicate cultures was used to produce cDNA using the SuperScript III RT kit (Life Technologies). qRT-PCR was performed using the QuantiFast SYBR Green PCR kit (Qiagen) following the manufacturer's instructions on an Mx3000P QPCR Systems instrument (Agilent). Relative gene expression levels were calculated using the $\Delta\Delta C_T$ method using *rrsH* as an internal standard.[41]

Acknowledgements

Michael J. Sweredoski and Sonja Hess assisted with proteomic and bioinformatic analyses. Susan Gottesman provided the pBRplac and pNRD405 plasmids and NRD359 *E. coli* strain. Jörg Vogel provided plasmids for sRNA fluorescence assays. We thank the Caltech PEL staff for support.

Figures and Tables

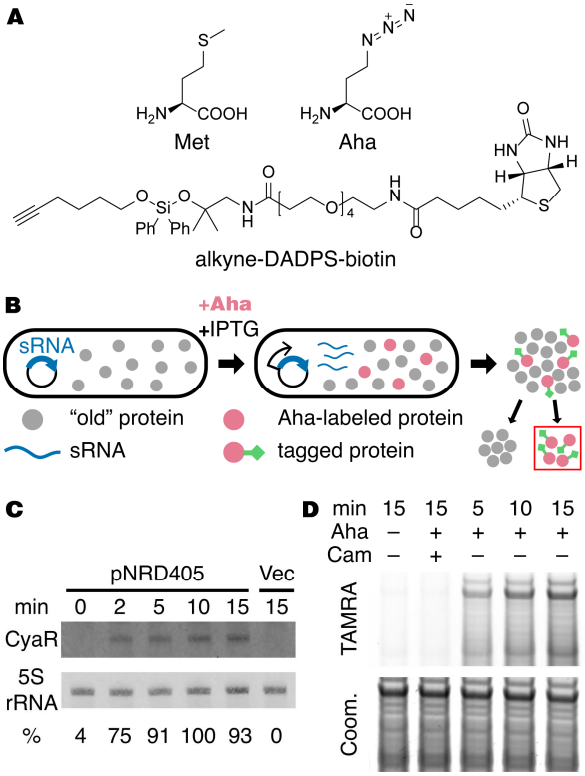


Figure 3.1. Using BONCAT for measuring the proteomic response to sRNA expression. (A) The chemical structures of L-Methionine, L-Azidohomoalanine, and the alkyne-DADPS-biotin probe. (B) Experimental outline for measuring the regulatory effects of sRNAs using BONCAT. Cells harboring a plasmid for sRNA expression are simultaneously treated with an inducer (e.g. IPTG) and Aha. Proteins produced during the pulse are Aha-labeled and can be conjugated to tags for affinity purification and subsequent proteomics analysis. (C) Northern blot showing expression of CyaR after induction with IPTG. Ribosomal RNA is shown as a loading control. pNRD405 and Vec (pBRplac) labels indicate strains with plasmids harboring *cyaR* and an empty vector, respectively. (D) Metabolic labeling of newly synthesized proteins with Aha. Cultures were treated with Aha (1 mM) for the indicated times. Aha-labeled proteins were tagged with an

alkyne-TAMRA probe via click chemistry and visualized by in-gel fluorescence. Cultures not treated with Aha and cultures treated with Aha and the protein synthesis inhibitor chloramphenicol showed no labeling. Cam, chloramphenicol. TAMRA, tetramethylrhodamine. Coom, coomassie.

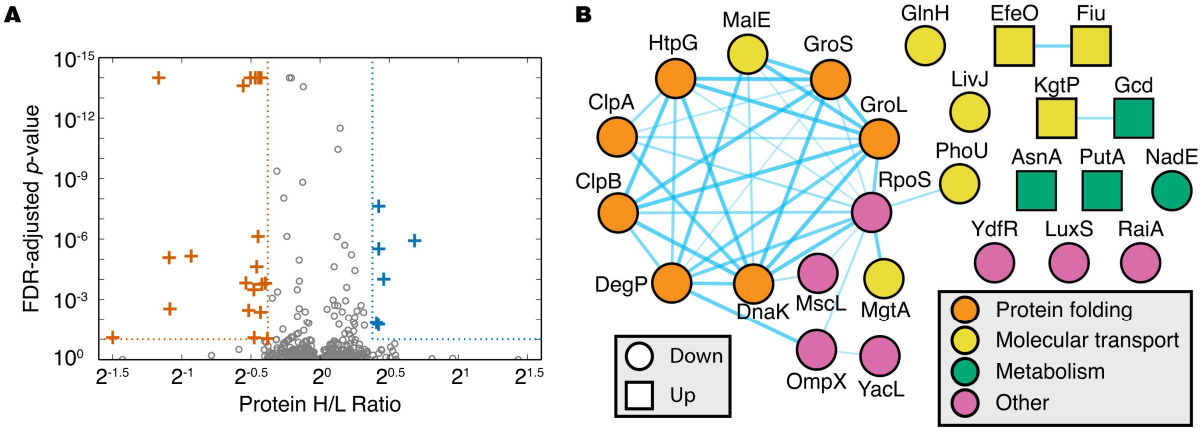


Figure 3.2. Determination of proteins significantly regulated by expression of CyaR. (A) Volcano plot showing the regulation and statistical significance of quantified proteins. Outlier proteins (crosses) were determined by three criteria: (1) proteins were quantified with differences greater than 1.3-fold, (2) protein ratios were quantified in both types of label swap experiments in the same direction, and (3) proteins had a false discovery rate-adjusted *p*-values less than 0.10 (Benjamini-Hochberg FDR).[40] (B) Functional protein association network of outlier proteins were enriched in interactions from the STRING database. Node color and shape represent functional associations and direction of regulation in response to CyaR, respectively.

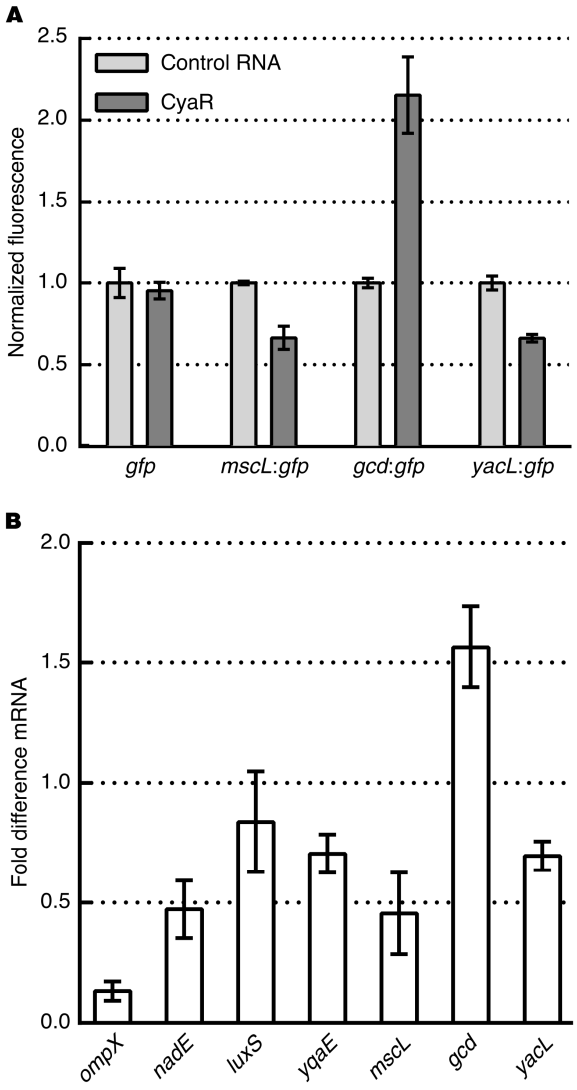


Figure 3.3. Direct regulation of target proteins by CyaR. (A) Target proteins fused to Gfp were assessed for CyaR regulation using flow cytometry. Fluorescence of Gfp fusions were compared in cells expressing either CyaR or a control RNA. (B) Target transcript regulation after a 15-min CyaR expression. mRNA values were measured using qRT-PCR and normalized to control cultures with an empty vector. Error bars designate standard error of the mean.

Table 3.1. List of outlier proteins significantly regulated by CyaR.

Uniprot ID	Gene name	Protein description[a]	Peptides	Log_2 H/L Ratio
P76160	*ydfR*	Uncharacterized protein	2	-1.50
P0A917	*ompX*	Outer membrane protein	3	-1.17
P0AD49	*raiA*	Ribosome-associated inhibitor	7	-1.09
P13445	*rpoS*	RNA polymerase sigma factor	1	-1.09
P0AEQ3	*glnH*	Glutamine-binding periplasmic protein	13	-0.93
P0A6F9	*groS*	Chaperone	4	-0.56
P0A742	*mscL*	Large-conductance mechanosensitive channel	3	-0.52
P63284	*clpB*	Chaperone	27	-0.50
P0A6Z3	*htpG*	Chaperone	23	-0.50
P0A9K7	*phoU*	Phosphate-transport system accessory protein	5	-0.48
P0A8E5	*yacL*	Uncharacterized protein	1	-0.48
P0A6F5	*groL*	Chaperone	31	-0.47
P0AEX9	*malE*	Maltose-binding periplasmic protein	14	-0.46
P0ABB8	*mgtA*	Magnesium-transporting ATPase, P-type 1	10	-0.45
P0C0V0	*degP*	Periplasmic serine endoprotease	13	-0.44
P0A6Y8	*dnaK*	Chaperone	26	-0.43
P0AD96	*livJ*	Leu/Ile/Val-binding protein	8	-0.43
P18843	*nadE*	NH_3-dependent NAD^+ synthetase	9	-0.42
P0ABH9	*clpA*	ATP-dependent Clp protease, ATP-binding subunit	17	-0.40
P45578	*luxS*	S-ribosylhomocysteine lyase	4	-0.38
P00963	*asnA*	Aspartate-ammonia ligase	3	0.41
P75780	*fiu*	Catecholate siderophore receptor	11	0.42
P0AEX3	*kgtP*	Alpha-ketoglutarate permease	1	0.42

P0AB24	*efeO*	Iron uptake system component	8	0.42
P15877	*gcd*	Quinoprotein glucose dehydrogenase	12	0.46
P09546	*putA*	Proline dehydrogenase	15	0.68

[a] Protein descriptions from Uniprot

Supplementary Figures and Tables

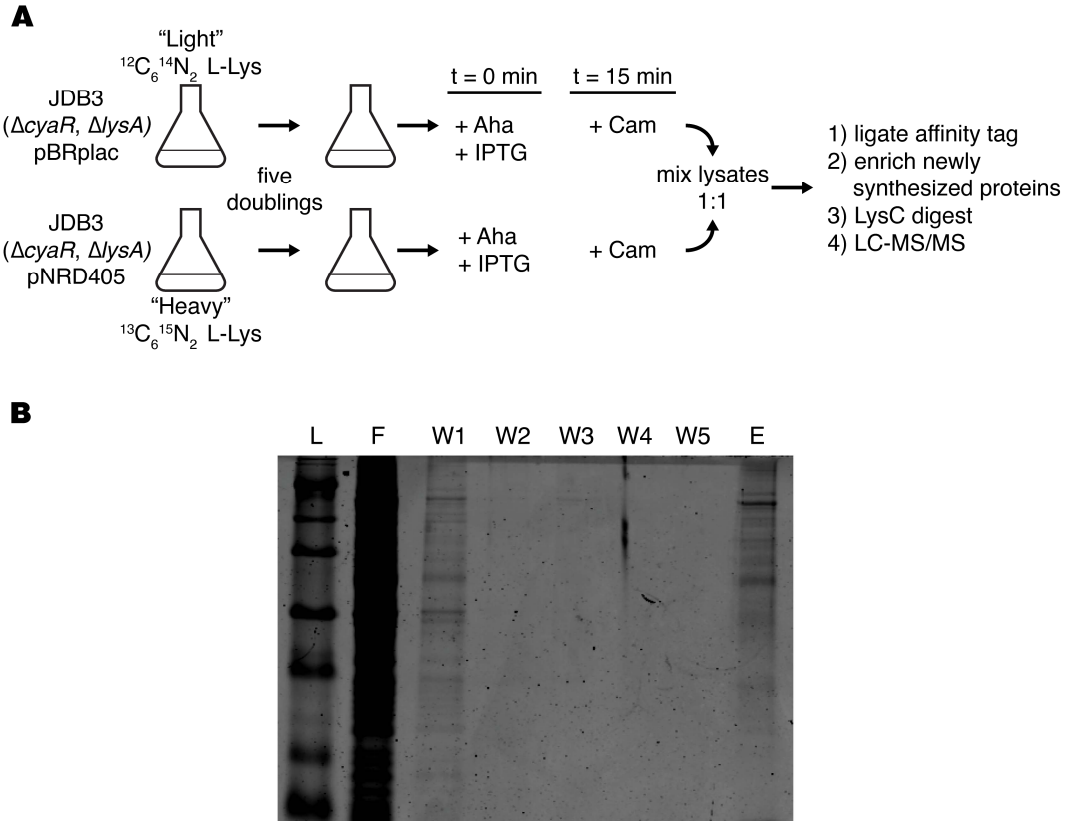


Figure 3.S1. (A) Detailed experimental outline of the combined BONCAT-SILAC approach for measuring protein regulation by CyaR. (B) Protein gel showing enrichment of Aha-labeled proteins. L — ladder, F — flowthrough, W1–5 — washes, E — elution.

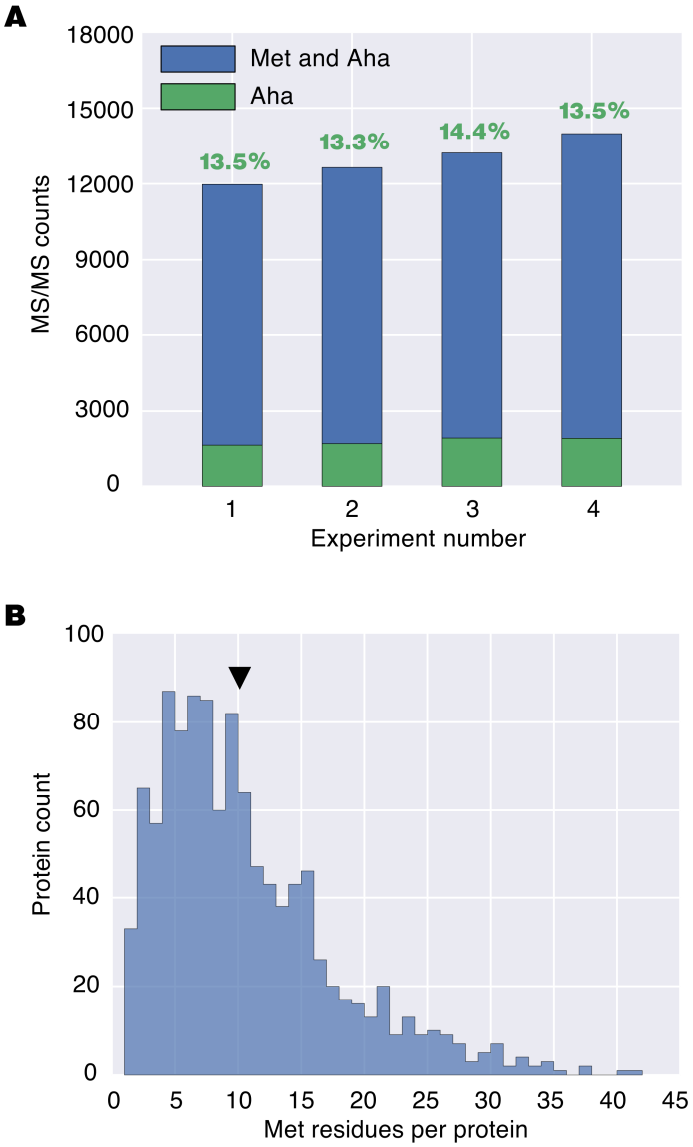


Figure 3.S2. Estimates of Met replacement with Aha for enriched proteins. (A) Total MS/MS spectra were counted for peptides containing Aha and Met and used to estimate the extent of Aha incorporation. All Aha-based modifications were included (e.g. Aha clicked with alkyne-DADPS-biotin tag successfully cleaved). Values above bar plots show the percent of Aha replacement in each experimental replicate. (B) Histogram showing the number of Met residues for all identified proteins. The arrow marker shows the mean value of the distribution, which,

when considering the incorporation rate calculated in part A, estimates that each enriched protein has, on average, about 1.4 Aha residues.

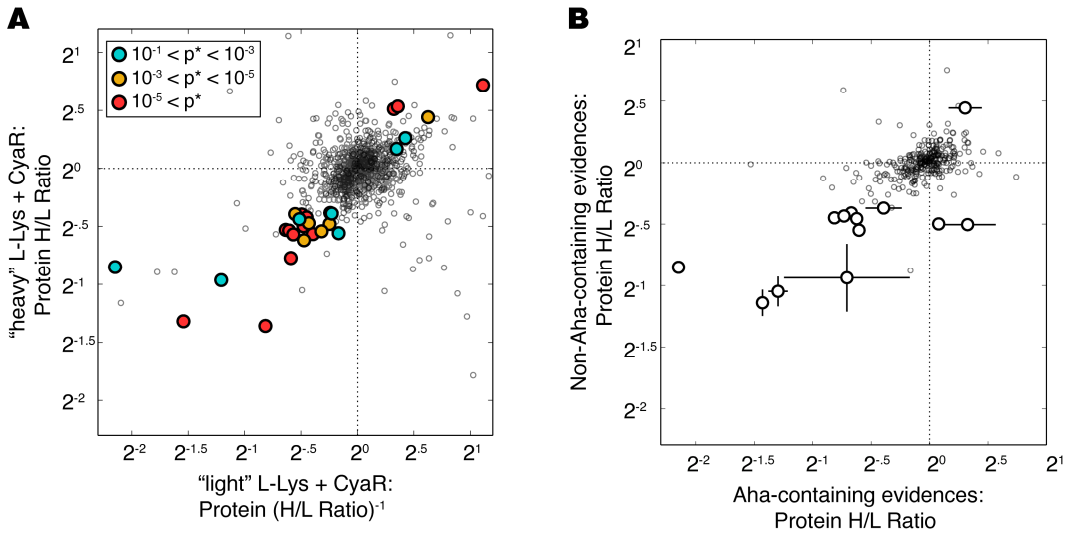


Figure 3.S3. Reproducibility of selected outlier proteins. (A) Consistency of outlier protein regulation in label swap experiments. In total, four proteomic experiments were performed: two in which CyaR was induced in the cultures labeled with the "heavy" L-Lys isotopolog (as depicted in Figure 3.S1) and two in which CyaR was induced in cultures with the "light" L-Lys isotopolog. Outlier proteins (colored markers) quantified in replicate label swap experiments were well correlated (R = 0.87). (B) Aha-containing evidences correlate with non-Aha-containing evidences for outlier proteins (R = .69). All Aha-containing evidences were included (similar to Figure 3.S2A). Error bars show standard error of the mean of quantified evidences.

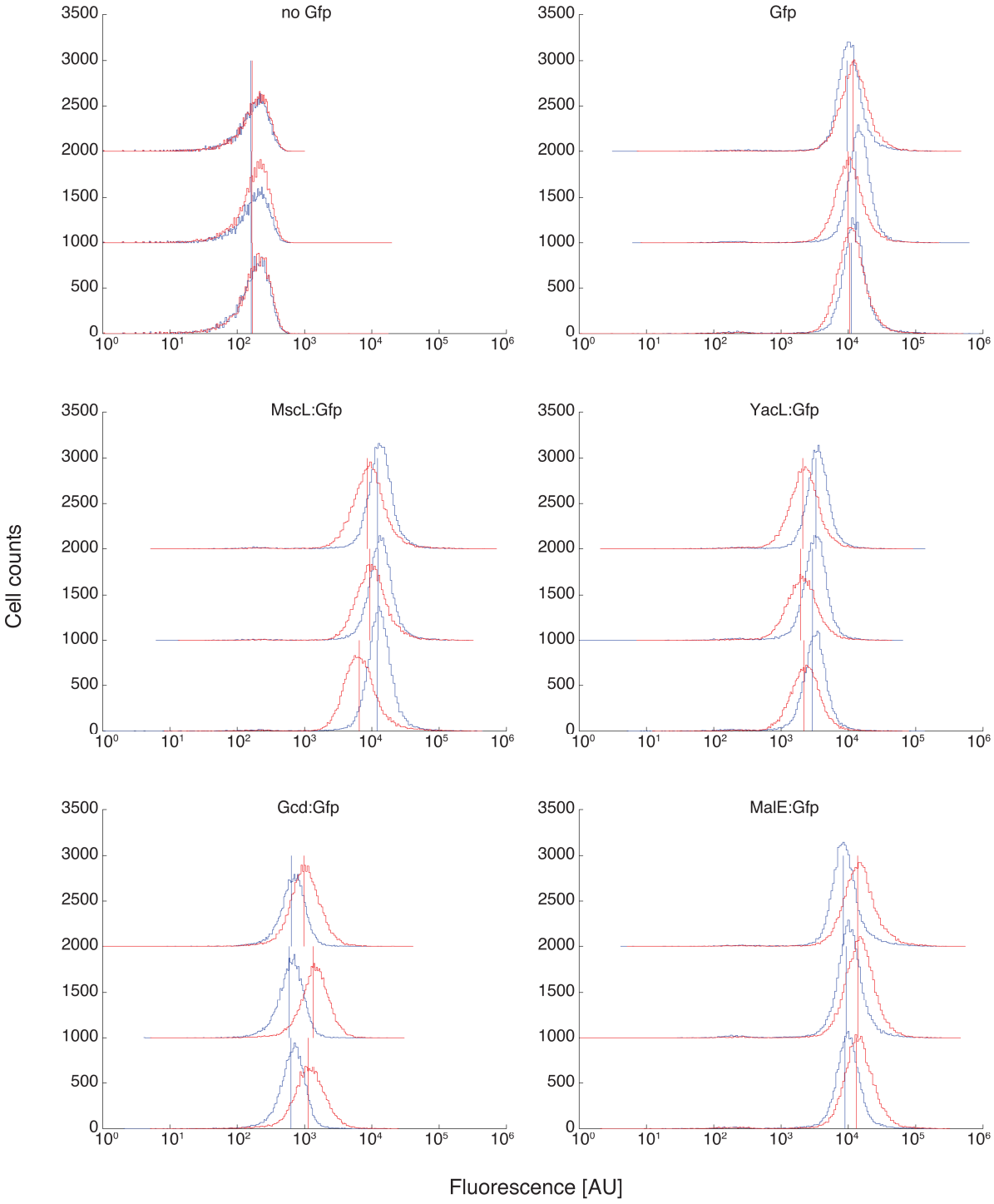


Figure 3.S4. Raw flow cytometry data for measuring direct CyaR regulation of target proteins. Red and blue lines show histograms of cells expressing CyaR and a control RNA, respectively.

Vertical lines mark the mean of each distribution. Histograms from replicate experiments are stacked vertically for visualization.

Table 3.S1. Protein quantifications from CyaR BONCAT experiments. This table is available as a downloadable excel file in the online version of this thesis.

Table 3.S2. Strains used in this study.

Strain	Relevant genotype	Source
NRD359	MG1655 Δ*cyaR*::*cat*	De Lay and Gottesman 2009
JDB1	MG1655 Δ*cyaR*	this study
JDB2	MG1655 Δ*cyaR ΔlysA*::*cat*	this study
JDB3	MG1655 Δ*cyaR ΔlysA*	this study
DH10B	F– *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *end*A1 *ara*D139 Δ(*ara leu*) 7697 *gal*U *gal*K *rps*L *nup*G λ–	

Table 3.S3. Plasmids used in this study.

Plasmid name	Description	Source
pBRplac	Amp^r; *araBAD* promoter-based expression vector having a pBR322 origin	De Lay and Gottesman 2009
pNRD405	AatII-EcoRI *cyaR*-containing fragment cloned into the same sites in pBRplac	De Lay and Gottesman 2009
pKD3	template for Cm^R mutant construction	Datsenko and Wanner 2000
pKD46	temperature-sensitive λ Red recombinase expression plasmid	Datsenko and Wanner 2000
pCP20	temperature-sensitive FLP recombinase expression plasmid	Datsenko and Wanner 2000
pXG0	non-fluorescent control plasmid (no $_{sf}$*gfp**)	Urban and Vogel 2007
pXG1	fluorescent control plasmid with constitutive $_{sf}$*gfp* expression	Urban and Vogel 2007
pJDB1	pXG1 with *gfp* replaced by $_{sf}$*gfp* and with the N-terminal linker region removed	this study
pXG10sf	plasmid for construction of $_{sf}$*gfp* fusion; pSC101* origin and Cm^R cassette	Corcoran et al. 2012
pJV300	control plasmid that expresses a ~50 nt nonsense RNA derived from the *rrnB* terminator region	Sitka et al. 2007
pZE12-luc	backbone vector for producing sRNA plasmids with ColE1 origin, Amp^r cassette, and PLlacO promoter driving luciferace expression	Urban and Vogel 2007
pJDB2	pZE12-luc with *cyaR* insert replacing luciferase	this study
pJDB3	pXG10sf with *mscL*(-23 to +30)-$_{sf}$*gfp* (NheI and NsiI)	this study
pJDB4	pXG10sf with *yacL*(-49 to +30)-$_{sf}$*gfp* (NheI and NsiI)	this study
pJDB5	pXG10sf with *gcd*(-38 to +30)-$_{sf}$*gfp* (NheI and NsiI)	this study
pJDB6	pXG10sf with *malE*(-45 to +30)-$_{sf}$*gfp* (NheI and NsiI)	this study

* $_{sf}$*gfp* refers to the superfolder variant of *gfp*

Table 3.S4. Oligos used in this study.

Oligo name	Sequence	Use
JDB_galR-cat-for	CAACTCCGTCGCTGGAGGCAAGTCATCATGCAACCAGCG ACTAACCGCAGGTGTAGGCTGGAGCTGCTTC	*lysA*::cam λ Red template
JDB_lysR-cat-rev	TTTTTATGATGTGGCGTAATCATAAAAAAGCACTTATCTG GAGTTTGTTAATGGGAATTAGCCATGGTCC	*lysA*::cam λ Red template
JDB_cyaR-for	(5' Phos)-GCTGAAAAACATAACCCATAAAATG	pJDB2 (insert)
JDB_cyaR-rev	GTTTTTTCTAGACCTTTTATTTCATTGTATTACGCG	pJDB2 (insert)
PLlacoB	CGCACTGACCGAATTCATTAA	pJDB2 (vector)
PLlacoD	GTGCTCAGTATCTTGTTATCCG	pJDB2 (vector)
JDB_pXG10sf-for	CCTCGCTAGCGGATC	pJDB3–6 (vector)
JDB_pXG10sf-rev	CACAATGCATGTGCTCA	pJDB3–6 (vector)
JDB_mscL-for	GTTTTTATGCATGTCGGCTTCATAGGG	pJDB3 (insert)
JDB_mscL-rev	GTTTTTGCTAGCAAATTCGCGAAATTCTT	pJDB3 (insert)
JDB_yacL-for	GTTTTTATGCATAATGGCGCAATGC	pJDB4 (insert)
JDB_yacL-rev	GTTTTTGCTAGCGGTAATATCGCGCAGA	pJDB4 (insert)
JDB_gcd-for	GTTTTTATGCATGCTACTAAAATATTAATGAATTGAAATG GTGTC	pJDB5 (insert)
JDB_gcd-rev	GTTTTTGCTAGCTCGTCGCGAGCCTGTATTG	pJDB5 (insert)
JDB_malE-for	GTTTTTATGCATGTTTAGGTGTTTTCACGAGCAC	pJDB6 (insert)
JDB_malE-rev	GTTTTTGCTAGCGAGGATGCGTGCACC	pJDB6 (insert)
JDB_cyaR-probe	(5' 488 AlexaFluor)-TGGTTCCTGGTACAGCTAGCATTTT ATGGGTTATG	CyaR RNA detection
JDB_5SrRNA-probe	(5' 488 AlexaFluor)-TACTCTCGCATGAGGAGACCCCACACTACCATC	5S rRNA detection
JDB_idnT-qPCR-for	ATTGCCCTCGTTCTGGTAGC	*idnT* qPCR
JDB_idnT-qPCR-rev	AACCGAGGATCATTGCCAGC	*idnT* qPCR
JDB_rrsH-qPCR-for	CAGCCACACTGGAACTGAGA	*rrsH* qPCR

JDB_rrsH-qPCR-rev	CTTCATACACGCGGCATGG	*rrsH* qPCR
JDB_ompX-qPCR-for	ATTGCATGTCTTTCAGCACTGG	*ompX* qPCR
JDB_ompX-qPCR-rev	TCATTTGGCCCTGAGCGTC	*ompX* qPCR
JDB_nadE-for	AACCGCAGATTAATGCTGAAGAGG	*nadE* qPCR
JDB_nadE-qPCR-rev	GATCCCGAGCACCAGTGATTTAATG	*nadE* qPCR
JDB_luxS-qPCR-for	AATTACCGGAGGTGGCTAAATGC	*luxS* qPCR
JDB_luxS-qPCR-rev	AACACGGTGATTGCGTCG	*luxS* qPCR
JDB_yqaE-qPCR-for	TCGTCATCACCATCATTCTGCC	*yqaE* qPCR
JDB_yqaE-qPCR-rev	CCAGAACGCGTGAATCAAACC	*yqaE* qPCR
JDB_malE-qPCR-for	AACAGGTGCACGCATCCTC	*malE* qPCR
JDB_malE-qPCR-rev	CGCCGTTAATCCAGATTACCAG	*malE* qPCR
JDB_gcd-qPCR-for	AATTAACAATACAGGCTCGCGACG	*gcd* qPCR
JDB_gcd-qPCR-rev	CGATAGGGTAGTACCAGGAGC	*gcd* qPCR
JDB_mscL-qPCR-for	TCGCGAATTTGCGATGCG	*mscL* qPCR
JDB_mscL-qPCR-rev	GGAGGCATGATGATATCGGCA	*mscL* qPCR
JDB_yacL-qPCR-for	GGCCTTGCTTGATGAAGTGG	*yacL* qPCR
JDB_yacL-qPCR-rev	CGAACCATCACCTCTTCACCG	*yacL* qPCR

References

(1) Gottesman, S., and Storz, G. (2011) Bacterial Small RNA Regulators: Versatile Roles and Rapidly Evolving Variations. *Cold Spring Harb. Perspect. Biol.* 3:a003798.

(2) Papenfort, K., and Vogel, J. (2009) Multiple target regulation by small noncoding RNAs rewires gene expression at the post-transcriptional level. *Res. Microbiol. 160*, 278–87.

(3) Feng, L., Rutherford, S. T., Papenfort, K., Bagert, J. D., van Kessel, J. C., Tirrell, D. A., Wingreen, N. S., and Bassler, B. L. (2015) A Qrr Noncoding RNA Deploys Four Different Regulatory Mechanisms to Optimize Quorum-Sensing Dynamics. *Cell 160*, 228–40.

(4) Møller, T., Franch, T., Udesen, C., Gerdes, K., and Valentin-Hansen, P. (2002) Spot 42 RNA mediates discoordinate expression of the E. coli galactose operon. *Genes Dev. 16*, 1696–706.

(5) Dieterich, D. C., Link, A. J., Graumann, J., Tirrell, D. A., and Schuman, E. M. (2006) Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT). *Proc. Natl. Acad. Sci. U. S. A. 103*, 9482–7.

(6) Dieterich, D. C., Lee, J. J., Link, A. J., Graumann, J., Tirrell, D. A., and Schuman, E. M. (2007) Labeling, detection and identification of newly synthesized proteomes with bioorthogonal non-canonical amino-acid tagging. *Nat. Protoc. 2*, 532–40.

(7) Kiick, K. L., Saxon, E., Tirrell, D. A., and Bertozzi, C. R. (2002) Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation. *Proc. Natl. Acad. Sci. U. S. A. 99*, 19–24.

(8) Bagert, J. D., Xie, Y. J., Sweredoski, M. J., Qi, Y., Hess, S., Schuman, E. M., and Tirrell, D. A. (2014) 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 13*, 1352–8.

(9) De Lay, N., and Gottesman, S. (2009) The Crp-activated small noncoding regulatory RNA CyaR (RyeE) links nutritional status to group behavior. *J. Bacteriol. 191*, 461–76.

(10) Wassarman, K. M., Repoila, F., Rosenow, C., Storz, G., and Gottesman, S. (2001) Identification of novel small RNAs using comparative genomics and microarrays. *Genes Dev. 15*, 1637–51.

(11) Papenfort, K., Pfeiffer, V., Lucchini, S., Sonawane, A., Hinton, J. C. D., and Vogel, J. (2008) Systematic deletion of Salmonella small RNA genes identifies CyaR, a conserved CRP-dependent riboregulator of OmpX synthesis. *Mol. Microbiol. 68*, 890–906.

(12) Johansen, J., Eriksen, M., Kallipolitis, B., and Valentin-Hansen, P. (2008) Down-regulation of outer membrane proteins by noncoding RNAs: unraveling the cAMP-CRP- and sigmaE-dependent CyaR-ompX regulatory case. *J. Mol. Biol. 383*, 1–9.

(13) Wright, P. R., Richter, A. S., Papenfort, K., Mann, M., Vogel, J., Hess, W. R., Backofen, R., and Georg, J. (2013) Comparative genomics boosts target prediction for bacterial small RNAs. *Proc. Natl. Acad. Sci. U. S. A. 110*, E3487–96.

(14) Szychowski, J., Mahdavi, A., Hodas, J. J. L., Bagert, J. D., Ngo, J. T., Landgraf, P., Dieterich, D. C., Schuman, E. M., and Tirrell, D. A. (2010) Cleavable Biotin Probes for Labeling of Biomolecules via Azide-Alkyne Cycloaddition. *J. Am. Chem. Soc. 132*, 18351–60.

(15) Gosset, G., Zhang, Z., Nayyar, S., Cuevas, W. A., and Saier, M. H. (2004) Transcriptome Analysis of Crp-Dependent Catabolite Control of Gene Expression in Escherichia coli. *J. Bacteriol. 186*, 3516–24.

(16) Zhang, Z., Gosset, G., Barabote, R., Gonzalez, C. S., Cuevas, W. A., and Saier, M. H. (2005) Functional interactions between the carbon and iron utilization regulators, Crp and Fur, in Escherichia coli. *J. Bacteriol. 187*, 980–90.

(17) Zheng, D., Constantinidou, C., Hobman, J. L., and Minchin, S. D. (2004) Identification of the CRP regulon using in vitro and in vivo transcriptional profiling. *Nucleic Acids Res. 32*, 5874–93.

(18) Wang, L., Hashimoto, Y., Tsao, C.-Y., Valdes, J. J., and Bentley, W. E. (2005) Cyclic AMP (cAMP) and cAMP receptor protein influence both synthesis and uptake of extracellular autoinducer 2 in Escherichia coli. *J. Bacteriol. 187*, 2066–76.

(19) Xiao, Q., Zhang, F., Nacev, B. A., Liu, J. O., and Pei, D. (2010) Protein N-terminal processing: Substrate specificity of escherichia coli and human methionine aminopeptidases. *Biochemistry 49*, 5588–99.

(20) Wang, A., Nairn, N. W., Johnson, R. S., Tirrell, D. A., and Grabstein, K. (2008) Processing of N-terminal unnatural amino acids in recombinant human interferon-beta in Escherichia coli. *ChemBioChem 9*, 324–30.

(21) Mandin, P., and Gottesman, S. (2010) Integrating anaerobic/aerobic sensing and the general stress response through the ArcZ small RNA. *EMBO J. 29*, 3094–107.

(22) Urban, J. H., and Vogel, J. (2007) Translational control and target recognition by Escherichia coli small RNAs in vivo. *Nucleic Acids Res. 35*, 1018–37.

(23) Corcoran, C. P., Podkaminski, D., Papenfort, K., Urban, J. H., Hinton, J. C. D., and Vogel, J. (2012) Superfolder GFP reporters validate diverse new mRNA targets of the classic porin regulator, MicF RNA. *Mol. Microbiol. 84*, 428–45.

(24) Yamada, M., Sumi, K., Matsushita, K., Adachi, O., and Yamada, Y. (1993) Topological Analysis of Quinoprotein Glucose Dehydrogenase in Escherichia coli and Its Ubiquinone-binding Site. *J. Biol. Chem. 268*, 12812–7.

(25) Duine, J. A., Frank, J., and van Zeeland, J. K. (1979) Glucose dehydrogenase from Acinetobacter calcoaceticus: A “quinoprotein.” *FEBS Lett. 108*, 443–6.

(26) Matsushita, K., Arents, J. C., Bader, R., Yamada, M., Adachi, O., and Postma, P. W. (1997) Escherichia coli is unable to produce pyrroloquinoline quinone (PQQ). *Microbiology 143*, 3149–56.

(27) De Jonge, R., De Mattos, M. J. T., Stock, J. B., and Neijssel, O. M. (1996) Pyrroloquinoline Quinone, a Chemotactic Attractant for Escherichia coli. *J. Bacteriol. 178*, 1224–6.

(28) Beisel, C. L., and Storz, G. (2010) Base pairing small RNAs and their roles in global regulatory networks. *FEMS Microbiol. Rev. 34*, 866–82.

(29) Mank, N. N., Berghoff, B. A., and Klug, G. (2013) A mixed incoherent feed-forward loop contributes to the regulation of bacterial photosynthesis genes. *RNA Biol. 10:3*, 347–52.

(30) Merlin, C., McAteer, S., and Masters, M. (2002) Tools for Characterization of Escherichia coli Genes of Unknown Function. *J. Bacteriol. 184*, 4573–81.

(31) Butland, G., Peregrin-Alvarez, J. M., Li, J., Yang, W., Yang, X., Canadien, V., Starostine, A., Richards, D., Beattie, B., Krogan, N., Davey, M., Parkinson, J., Greenblatt, J., and Emili, A. (2005) Interaction network containing conserved and essential protein complexes in Escherichia coli. *Nature 433*, 531–7.

(32) Lee, D. J., Busby, S. J. W., Westblade, L. F., and Chait, B. T. (2008) Affinity Isolation and I-DIRT Mass Spectrometric Analysis of the Escherichia coli O157:H7 Sakai RNA Polymerase Complex. *J. Bacteriol. 190*, 1284–9.

(33) Butland, G., Babu, M., Diaz-Mejia, J. J., Bohdana, F., Phanse, S., Gold, B., Yang, W., Li, J., Gagarinova, A. G., Pogoutse, O., Mori, H., Wanner, B. L., Lo, H., Wasniewski, J., Christopoulos, C., Ali, M., Venn, P., Safavi-Naini, A., Sourour, N., Caron, S., Choi, J.-Y., Laigle, L., Nazarians-Armavil, A., Deshpande, A., Joe, S., Datsenko, K. A., Yamamoto, N., Andrews, B. J., Boone, C., Ding, H., Sheikh, B., Moreno-Hagelsieb, G., Greenblatt, J. F., and Emili, A. (2008) eSGA: E . coli synthetic genetic array analysis. *Nat. Methods 5*, 789–5.

(34) Edwards, M. D., Black, S., Rasmussen, T., Rasmussen, A., Stokes, N. R., Stephen, T.-L., Miller, S., and Booth, I. R. (2012) Characterization of three novel mechanosensitive channel activities in Escherichia coli. *Channels 6*, 272–81.

(35) Levina, N., Tötemeyer, S., Stokes, N. R., Louis, P., Jones, M. A., and Booth, I. R. (1999) Protection of Escherichia coli cells against extreme turgor by activation of MscS and MscL mechanosensitive channels: identification of genes required for MscS activity. *EMBO J. 18*, 1730–7.

(36) Link, A. J., Vink, M. K. S., and Tirrell, D. A. (2007) Synthesis of the functionalizable methionine surrogate azidohomoalanine using Boc-homoserine as precursor. *Nat. Protoc. 2*, 1884–7.

(37) Datsenko, K. A., and Wanner, B. L. (2000) One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A. 97*, 6640–5.

(38) Hong, V., Presolski, S. I., Ma, C., and Finn, M. G. (2009) Analysis and Optimization of Copper-Catalyzed Azide-Alkyne Cycloaddition for Bioconjugation. *Angew. Chemie 48*, 9879–83.

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

(40) Benjamini, Y., and Hochberg, Y. (1995) Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. *J. Roy. Stat. Soc. B Met. 57*, 289–300.

(41) Pfaffl, M. W. (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res. 29*, 2002–7.

CHAPTER 4

TIME-RESOLVED PROTEOMIC ANALYSIS OF QUORUM SENSING IN *VIBRIO HARVEYI*

Abstract

Bacteria use a process of chemical communication called quorum sensing to assess their population density and to change their behavior in response to fluctuations in the cell number and species composition of the community. In this work, we identified the quorum-sensing-regulated proteome in the model organism *Vibrio harveyi* by bio-orthogonal non-canonical amino acid tagging (BONCAT). BONCAT enables measurement of proteome dynamics with temporal resolution on the order of minutes. We deployed BONCAT to characterize the time-dependent transition of *V. harveyi* from individual- to group-behaviors. We identified over 170 quorum-sensing-regulated proteins at early, intermediate, and late stages of the transition, and we mapped the temporal changes in quorum-sensing proteins controlled by both transcriptional and post-transcriptional mechanisms. Our analysis also revealed new quorum-sensing-regulated proteins with diverse functions, including transcription factors, chemotaxis proteins, transport proteins, and proteins involved in iron homeostasis.

Significance

This work provides the first proteome-wide, time-resolved analysis of the quorum-sensing transition of a bacterium. Many of the identified proteins have not previously been associated with quorum sensing in *V. harveyi*. Our results demonstrate the utility of bio-orthogonal non-canonical amino acid tagging (BONCAT) as a broadly applicable method for measuring changes in protein synthesis on a time scale of minutes.

Introduction

Bacteria assess their cell numbers and the species complexity of the community of neighboring cells using a chemical communication process called quorum sensing. Quorum sensing relies on the production, release, accumulation and group-wide detection of signal molecules called autoinducers. Quorum sensing controls genes underpinning collective behaviors including bioluminescence, secretion of virulence factors, and biofilm formation (1–3). The model quorum-sensing bacterium *Vibrio harveyi* integrates population-density information encoded in three autoinducers AI-1, CAI-1, and AI-2, which function as intraspecies, intragenus, and interspecies communication signals, respectively (4–6). *V. harveyi* detects the three autoinducers using the cognate membrane-bound receptors LuxN, CqsS, and LuxPQ, respectively (7–9). At low cell density (LCD), autoinducer concentrations are low, and the unliganded receptors act as kinases, funneling phosphate to the phosphorelay protein LuxU (10). LuxU transfers the phosphoryl group to the response regulator protein LuxO, which activates transcription of genes encoding five homologous quorum regulatory small RNAs (Qrr sRNAs) (11, 12). The Qrr sRNAs post-transcriptionally activate production of the transcription factor AphA and repress production of the transcription factor LuxR. AphA and LuxR are the two master quorum-sensing regulators that promote global changes in gene expression in response to population density changes (12–15). At high cell density (HCD), autoinducer binding to the cognate receptors switches the receptors from kinases to phosphatases, removing phosphate from LuxU and, indirectly, from LuxO. Dephosphorylated LuxO is inactive so transcription of the *qrr* sRNA genes ceases. This event results in production of LuxR and repression of AphA (12). Thus, the circuitry ensures that AphA is made at LCD, and it controls the regulon required for life as an individual, whereas LuxR is made at HCD, and it directs the program underpinning collective behaviors.

Previous microarray studies examined the transcriptomic response during quorum-sensing transitions. That work showed that AphA and LuxR control over 150 and 600 genes, respectively and ~70 of these genes are regulated by both transcription factors (15). Both AphA and LuxR act as activators and as repressors, and thus the precise pattern of quorum-sensing target gene expression is exquisitely sensitive to fluctuating levels of AphA and LuxR as cells transition between LCD and HCD modes. Developing a comparable understanding of the quorum-sensing-controlled proteome requires measurement of dynamic changes in protein abundance throughout the transition from individual to collective behavior.

In this work, we used the bio-orthogonal non-canonical amino acid tagging (BONCAT) method to track the proteome-wide quorum-sensing response in *V. harveyi* with temporal precision. BONCAT enabled us to identify 176 proteins that are regulated during the transition from individual to collective behavior; 90 of these proteins are in addition to those identified in earlier studies. We show that a broad range of protein functional groups, including those involved in metabolism, transport, and virulence, change during the transition to group behavior. We demonstrate how particular temporal patterns of protein production are linked to particular tiers of the regulatory cascade by comparing the proteomic profiles of the regulon controlled by the post-transcriptional Qrr sRNAs to the regulon controlled by the transcriptional regulator LuxR. Using this approach, we, for example, determined that the *V. harveyi* type VI secretion system is LuxR-regulated.

Results

The BONCAT method was developed to provide time-resolved analyses of the cellular proteome (16, 17). In a BONCAT experiment, the non-canonical amino acid L-azidohomoalanine (Aha) is provided to cells and, subsequently, incorporated into proteins in competition with methionine

(Fig. 4.S1*A*) (18). Aha-labeled proteins are chemically distinct from the remainder of the protein pool and thus, labeled proteins can be selectively conjugated to affinity tags for enrichment and mass spectrometry analysis (Fig. 4.1*A*). Because Aha can be introduced into cells in a well-defined pulse, BONCAT offers excellent temporal resolution and high sensitivity to changes in protein synthesis in response to biological stimuli (19).

Our goal was to identify time-dependent changes in protein production associated with quorum sensing. We chose to monitor the transition from individual to group behavior in *V. harveyi* because the core transcriptional regulon is well-established, providing a solid foundation for comparisons between transcriptional and translational outputs (15). To experimentally manipulate the transition from LCD to HCD, we used *V. harveyi* strain TL25 in which the genes encoding the autoinducer receptors for CAI-1 (*cqsS*) and AI-2 (*luxPQ*) and the AI-1 synthase (*luxM*) have been deleted (15) . Thus, *V. harveyi* TL25 responds exclusively to exogenously supplied AI-1, which enables precise control over the activation of quorum sensing.

The hallmark phenotypic response controlled by quorum sensing in *V. harveyi* is bioluminescence, which is activated by LuxR during the transition from LCD to HCD (20). Thus, we reasoned that light production could serve as a proxy for activation of quorum sensing (20). Upon treatment of a culture of *V. harveyi* TL25 with AI-1, bioluminescence increases sharply after 30 min and plateaus at 400-fold higher than the pre-addition level at near 90 min (Fig. 4.1*B*). Detection of Aha incorporation in *V. harveyi* cultures by in-gel fluorescence showed that BONCAT experiments could be performed in this system with a temporal resolution of ten minutes (Fig. 4.1*C*, 4.S1*B*). Using the bioluminescence profile as a guide, we combined two techniques, BONCAT and stable isotope labeling with amino acids in cell culture (SILAC), to monitor both increases and decreases in protein synthesis in ten-minute intervals between 0 and 90 min following addition of AI-1 (Fig. 4.1*B*, 4.S1*C*,*D*) (19, 21). *V. harveyi* cultures that were

not treated with AI-1 served as references for relative quantification. As expected, the production of the luciferase subunits LuxA and LuxB tracked with the bioluminescence profile in cultures treated with AI-1 (Fig. 4.1*B*). We detect LuxB at 30 min, slightly before we can detect LuxA. The LuxB measurement is coincident with the first increase in bioluminescence. Between 40 and 50 min, bioluminescence and LuxA and LuxB levels exhibited sharp increases, after which, both continued to climb at slower rates. Between 60 and 90 min, the production rates of LuxA and LuxB remained nearly constant while bioluminescence continued to increase. LuxA and LuxB increased about 8-fold total in response to autoinducer supplementation. This result highlights the fact that BONCAT measures protein synthesis rates during individual time intervals (not total protein abundance), whereas bioluminescence output reports on the total accumulated LuxAB activity.

LuxA and LuxB are encoded by the *lux* operon, which also encodes LuxC, an acyl-CoA reductase, LuxD, an acyl transferase, and LuxE, a long-chain fatty-acid ligase. LuxCDE synthesize the substrate required by the LuxAB luciferase enzyme. All five proteins exhibited large, concurrent increases in translation at 50 min (Fig. 4.1*D*). The increase in bioluminescence precedes production of LuxCDE, which suggests some basal level of luciferase substrate is present. The coincidence of the production of LuxA and LuxB with the onset of bioluminescence, and the simultaneous upregulation of all of the proteins in the *lux* operon validate the BONCAT technique as a reliable method for time-resolved analysis of the quorum-sensing response.

Detection of quorum-sensing regulators. At the core of the quorum-sensing circuit are the transcriptional regulators LuxO, AphA, and LuxR, which drive quorum-sensing transitions. Expression of *luxO*, *aphA*, and *luxR* are themselves controlled by multiple regulatory feedback

loops (13, 15, 22–24). To assess the consequences of addition of AI-1 to *V. harveyi* TL25 on these core regulators, we monitored both mRNA and protein synthesis using qRT-PCR and BONCAT, respectively. LuxO, AphA, and LuxR all showed rapid changes in protein production within 20 min of AI-1 treatment (Fig. 4.2). AphA and LuxR reached near-maximal differences in translation at the 30 min point; AphA protein production decreased 4-fold and LuxR protein production increased 16-fold. The mRNA levels of *aphA* and *luxR* tracked with those of AphA and LuxR protein changes, with the exception that *luxR* mRNA decreased in abundance between 60 and 90 min while the protein level remained constant. LuxO protein exhibited a consistent 2-fold increase in abundance throughout the time-course, whereas the corresponding mRNA levels slightly decreased. This pattern is consistent with the recent finding that the Qrr sRNAs control *luxO* mRNA through a sequestration mechanism such that the Qrr sRNAs repress LuxO protein production while not significantly altering mRNA abundance (25).

Quorum sensing causes global changes in protein synthesis. Using the above protocol for induction of quorum sensing in *V. harveyi* TL25, we next examined the quorum-sensing-controlled proteome using BONCAT to monitor protein synthesis in ten-minute time intervals immediately following addition of AI-1. We collected a total of 700,174 MS/MS spectra and identified 9238 peptides and 1564 unique protein groups (Fig. 4.S3*A*,*B*, Table 4.S1). Proteins were identified with an average of 6 peptides (median = 4); 88% of proteins were identified by 2 or more peptides (Fig. 4.S3*C*). Relative protein abundances at each time point were calculated with an average of 49 unique quantifications (median = 17) (Fig. 4.S3*D*). By comparing evidence counts, MS-MS counts, and MS intensities of Met and Aha-containing peptides, we estimated the extent of replacement of Met by Aha to be roughly 15% (Table 4.S2). Proteins with

differences greater than 1.5-fold with false discovery rate-adjusted *p*-values less than 0.05 were considered significant.

Induction of quorum sensing altered production of 176 proteins (Fig. 4.3*A*). Unsupervised hierarchical clustering partitioned the regulated proteins into 10 groups based on their temporal production profiles (Fig. 4.3*A,B*). Proteins from the *lux* operon clustered closely (group F), and LuxR and AphA, which exhibited distinct production profiles, were assigned to very small clusters. Several clusters showed differences in protein production at early time points (groups D, E, I), whereas other clusters changed more abruptly at the 50-min time point (groups B, D, F, H) (Fig. 4.3*B*). Differences in protein production between AI-1-treated and control cultures were modest within the first 20 min, with only 7 and 19 significant protein changes at 0–10 min and 10–20 min, respectively. The number of autoinducer-regulated proteins increased with time after induction, with 42–119 proteins altered between 40–90 min after AI-1 treatment (Fig. 4.3*C*,*D*). 90 of the AI-1-regulated proteins are newly associated with quorum sensing in *V. harveyi* (Fig. 4.3*E*, Table 4.2). In total, our analysis identified 278 proteins that are members of the previously established *aphA*, *luxR*, or quorum-sensing regulons (15). Interestingly, only 86 of these proteins exhibited significant up- or down-regulation by BONCAT (Fig. 4.S4).

Bioinformatic analysis reveals regulation of functionally related protein groups. To identify major shifts in protein production in response to induction of quorum sensing, we used principal component analysis (PCA) to simplify the dataset by reducing the dimensionality from 9 time points to 2 principal components. Weighting vectors showing the contribution of each time point to the principal components highlighted three distinct proteomic states: 1) an early period in which few proteins changed (10–30 min), 2) a transitional period that included rapid changes

in protein production (40–50 min), and 3) a late period in which many proteins exhibited large differences in translation (60–90 min) (Fig. 4.4*A*, Table 4.S3). As confirmation of these states, proteins with principal component coordinates near the 1st, 2nd, and 3rd sets of vectors exhibited time-course production profiles with punctuated changes at early, middle, and late stages (Fig. 4.4*B*). Gene ontology analysis identified 13 protein groups regulated by quorum sensing (Fig. 4.4*C*, 4.S5). Several of these groups were involved in transport, including iron, oligopeptide, and dicarboxylic acid transport. A set of 50 proteins with functional annotations for transporter activity was the largest of enriched ontology groups. Other groups of biological processes included bioluminescence, type VI secretion, siderophore synthesis, thiamine metabolism, and chemotaxis.

To identify groups of functionally related proteins with similar patterns of protein production, we mapped protein interactions from the STRING database onto the PCA plot and scanned for protein networks that localized via their principal components (Fig. 4.4*D*). Consistent with our gene ontology analysis, we identified interacting protein groups associated with regulation of bioluminescence, type VI secretion, chemotaxis, iron homeostasis, oligopeptide transport, and thiamine metabolism in the quorum-sensing response (Fig. 4.4*D*). For example, regarding peptide transport, synthesis of the substrate binding protein of the oligopeptide permease complex, OppA, decreased two-fold between 50–90 min (26). Also, a large group of proteins (16) involved in iron transport exhibited decreased production profiles late in the experiment, and a group of iron-regulatory proteins (6) increased in levels. With respect to chemotaxis, we observed both increases and decreases in protein levels: homologs of methyl-accepting chemotaxis proteins and the CheA and CheY signaling proteins decreased, whereas putative methyl-accepting chemotaxis proteins increased in abundance. Taken

together, these results suggest an overall quorum-sensing-driven remodeling of iron homeostasis and chemotactic behavior.

Defining the temporal order of protein regulation in response to quorum sensing. The Qrr sRNAs play a central role in dictating the transition between LCD and HCD states by controlling expression of the quorum-sensing transcriptional regulators, AphA, LuxR, and LuxO (Fig. 4.5*A*) (22, 23). The Qrr sRNAs directly regulate 16 additional targets outside of the quorum-sensing cascade with functions in virulence, metabolism, polysaccharide export, and chemotaxis (27). The direct Qrr targets constitute the set of "first-response" genes and also trigger the later, broader changes in downstream gene expression. With respect to the second wave of quorum-sensing gene expression changes, LuxR plays the major role. Therefore, we compared the temporal patterns of regulation of proteins known to be direct targets of either the Qrr sRNAs or LuxR (27, 28). We detected regulation of production of seven proteins known to be encoded by Qrr-regulated genes, all of which exhibited significant differences in expression within 20 minutes of AI-1 treatment (Fig. 4.5*B*). Conversely, 20 of the 21 LuxR-regulated proteins identified by BONCAT showed differences in production only after at least 30 minutes of AI-1 induction. Thus, the differences in timing between Qrr- and LuxR-regulated genes reflect the underlying structure of the quorum-sensing circuitry. We investigated the protein production profiles of the newly identified proteins to pinpoint additional candidates for regulation by the Qrr sRNAs. We found 19 additional proteins that are regulated within 20 minutes of AI-1 treatment, suggesting that the corresponding mRNAs may be targeted by the Qrr sRNAs (Table 4.1). The candidates include two putative chemotaxis proteins, the serine protease inhibitor ecotin, the type III secretion protein chaperone SycT, a chitinase, and several other proteins involved in metabolism. Strikingly, the mRNA and protein production of VIBHAR_02788 (a

predicted chemotaxis protein) increased 4- and 12-fold, respectively, within the first 10 minutes after AI-1 treatment, suggesting that VIBHAR_02788 is a good candidate for post-transcriptional regulation by the Qrr sRNAs (Fig. 4.5*C*).

The mechanisms that control production of quorum-sensing-regulated proteins undoubtedly become more complex as the response progresses. We identified proteins that were regulated at all stages (early (0–20 min), intermediate (20–60 min), and late (60–90 min)) following AI-1 treatment (Fig. 4.5*D*, Table 4.S1). Differences in the timing of quorum-sensing-regulated proteins suggest that additional regulatory components or mechanisms orchestrate the transition from individual to group behavior. For example, direct LuxR targets were regulated in both the intermediate and late phases, despite the fact that LuxR reaches its peak production at 30 min (Fig. 4.2*A*). This result suggests that accumulation of LuxR or additional transcriptional regulators contribute to control of LuxR-regulated genes.

Quorum sensing regulates type VI secretion proteins in *V. harveyi*. Components of the type VI secretion system (TSSS) were among the proteins most strongly up-regulated in response to AI treatment (Fig. 4.6*A*). Identified TSSS proteins included the haemolysin co-regulated effector protein (Hcp; VIBHAR_05871), and two additional proteins whose homologs have been implicated in TSSS regulation and Hcp secretion (VIBHAR_05854 and VIBHAR_05858) (29, 30). TSSS proteins exhibited a coordinated increase in production at 50 min, a profile similar to that of LuxCDABE.

In *V. harveyi*, the TSSS homologs are encoded by five putative operons: *VIBHAR_05855–05851, VIBHAR_05856–05858*, *VIBHAR_05865–05859*, *VIBHAR_05871–05866,* and *VIBHAR_05872–05873* (Fig. 4.S6*A*). Analysis of the mRNA levels of the operons confirmed the increase in expression of TSSS components between 50 to 60 min after AI-induction; timing

consistent with second-tier regulation (Fig. 4.6*B*). Previous microarray data comparing wild-type, Δ*luxR*, Δ*aphA*, and Δ*luxR* Δ*aphA V. harveyi* strains showed that TSSS gene expression was reduced in Δ*luxR* strains, but expression was not altered in the Δ*aphA* strain, providing evidence that expression of TSSS genes is LuxR-dependent and AphA-independent (Fig. 4.S6*B*) (13). Consistent with this notion, ChIP-seq data identified a LuxR binding site in the bi-directional promoter region of *VIBHAR_05855–05856* (28). Using electrophoretic mobility shift assays, we confirmed the presence of this LuxR binding site and determined that LuxR binds to two additional promoter regions in the TSSS locus (Fig. 4.S6*C*). This result shows that, unlike *Vibrio cholerae* which deploys the Qrr sRNAs to post-transcriptionally regulate TSSS*, V. harveyi* uses LuxR to control TSSS production (31). This finding suggests that although both organisms have TSSS under quorum-sensing control, they employ different regulatory strategies to achieve distinct timing of TSSS protein production.

Discussion

Global transcriptomic studies of *V. harveyi* have uncovered a continuum of changes in gene expression during the transition from LCD to HCD. As *V. harveyi* responds to changes in concentrations of autoinducers, shifts in the levels of the regulatory components AphA, LuxR, and the Qrr sRNAs occur, which in turn alter the expression of the downstream genes in the quorum-sensing regulon. Here we used the BONCAT method to measure changes in the quorum-sensing-regulated proteome during the transition from LCD to HCD, with a time-resolution of 10 min. We found correlated changes in production of the LuxCDABE enzymes and in the intensity of bioluminescence produced by the culture, and we observed regulation of the core regulatory components AphA, LuxR, and LuxO. Notably, the increase in LuxO upon

induction of quorum sensing occurred at the level of the protein, but not the mRNA, consistent with the hypothesis that the *luxO* mRNA is regulated by sequestration by the Qrr sRNAs (25).

The time resolution of the BONCAT method allowed us to identify proteins whose rates of synthesis were altered during the early, intermediate, and late stages of the LCD to HCD transition. The proteins found to be regulated within the first 20 min of autoinducer treatment included seven of the 20 known Qrr sRNA targets along with 19 other proteins not previously associated with Qrr regulation. No known Qrr targets were regulated at later times. In contrast, changes in the known LuxR targets occurred between 30 and 90 min following induction. Notably, proteins in the TSSS were upregulated between 40 and 50 min following autoinducer treatment, suggesting LuxR regulation of type VI secretion in *V. harveyi*; this conclusion was confirmed by electrophoretic mobility shift assays. Several LuxR-regulated genes exhibited changes in protein production only very late in the BONCAT experiment, which suggests either that they are responsive to accumulating LuxR levels, that they are regulated by another transcription factor downstream of LuxR, or that they are co-regulated by other factors.

We found quorum-sensing-dependent changes in 176 proteins that span a broad range of functional groups, including those related to iron homeostasis, molecular transport, metabolism, and chemotaxis. Ninety of these proteins are newly associated with quorum sensing in *V. harveyi*; the remainder are members of the previously established quorum-sensing, AphA, and/or LuxR regulons. Interestingly, nearly 200 other proteins from these regulons were identified by BONCAT but were not significantly up- or down-regulated. For example, the quorum-sensing regulon, which was defined by differences in gene expression between a mutant *V. harveyi* strain locked at LCD and a strain locked at HCD, contains 365 regulated genes as determined by microarray analysis (15). We quantified protein expression levels of 127 (35 %) of these genes, 45 (35 %) of which were significantly regulated. The differences between the

genetic and proteomic results may arise, at least in part, from differences in the growth media used in the two experiments (rich (LM) medium in the genetic study (13, 15) *vs*. minimal (AB) medium here). Support for this explanation comes from the finding that in *V. harveyi* cultures grown in LM medium, *aphA* mRNA levels decrease nearly 50-fold following addition of AI-1 (data not shown); a much smaller (~4-fold) decrease occurs in AB medium. Furthermore, we would not expect the rapid addition of saturating amounts of AI-1 to a *V. harveyi* culture to exactly recapitulate the effects of locking the strain into either the LCD or the HCD state, or of the gradual changes in the concentrations of autoinducers or regulators that accompany the growth of wild-type cultures. The complementary nature of the genetic and proteomic experiments underscores the importance of measuring mRNA and protein levels under a variety of conditions relevant to quorum sensing.

The BONCAT method has allowed us to identify a diverse set of proteins that respond to the induction of quorum sensing in *V. harveyi*. The method facilitates monitoring of changes in protein synthesis on a time scale of minutes, and enables correlation of those changes with the underlying temporal pattern of regulation of the quorum-sensing response. The approach described here should prove useful in studies of a wide variety of time-dependent cellular processes.

Materials and Methods

Cell culture – For each set of experiments, overnight cultures of *V. harveyi* strain TL25 (Δ*luxM* Δ*luxPQ* Δ*cqsS*) was used to inoculate 625 ml of AB minimal medium containing 18 amino acids (−Met, −Lys) at an OD_{600} of 0.003 (15). The culture was divided into six 100-ml aliquots. Three aliquots were supplemented with “light” Lys and three were supplemented with “heavy” Lys (U-$^{13}C_6$ U-$^{15}N_2$ L-lysine, Cambridge Isotope Laboratories). When the aliquotted cultures reached an

OD_{600} of 0.1 (~5 doublings), two "heavy" cultures (replicates 1 and 2) and one "light" culture (replicate 3) were treated with AI-1 at a final concentration of 10 µM ('AI-1 added'); the other three cultures were left untreated ('no AI-1 added'). At the specified time intervals, Aha was pulsed into all six cultures at a final concentration of 1 mM. After 10 min of Aha treatment, protein synthesis was halted by the addition of 100 µg/ml chloramphenicol (Sigma). Cells were pelleted, frozen at -80 °C, and stored for downstream processing. Aha was synthesized as described previously (32). Cultures were grown at 30 °C in a shaking incubator at 250 rpm.

Molecular methods – To measure changes in gene expression following induction of quorum sensing in *V. harveyi* TL25, cultures were grown as described above, divided in half, and AI-1 was added to one of the aliquots. Samples were collected every 10 min and RNA was isolated as described previously (13). cDNA synthesis and qRT-PCR were performed as described previously (22). The levels of gene expression were normalized to the internal standard *hfq* using either the $\Delta\Delta C_T$ method or the standard curve method. At least two replicates were collected for each sample ('AI-1 added' or 'no AI-1 added'). The graphs show the average of those measurements and are calculated as 'AI-1 added' divided by 'no AI-1 added'. Electrophoretic mobility shift assays were performed as previously described (15). PCR products were generated using oligonucleotides (Integrated DNA Technologies) listed in Table 4.S4.

BONCAT – Cells were lysed by heating in 1% SDS in PBS at 90 °C for 10 min and lysates were cleared by centrifugation. Protein concentrations were determined with the BCA protein quantitation kit (Thermo Scientific), and paired 'light' and 'heavy' cultures were mixed at equal quantities of total protein. Azide-alkyne click chemistry was performed as described in Hong et al. with a 0.1 mM alkyne-DADPS tag and allowed to proceed for 4 hr at room temperature (Fig.

4.S1*E*) (33). The DADPS tag was synthesized as described previously (34). Proteins were concentrated by acetone precipitation and solubilized in 2% SDS in PBS. Solutions were diluted to 0.15% SDS in PBS, and tagged proteins were captured by incubating with Streptavidin UltraLink Resin (Thermo Scientific) for 30 min at room temperature. Resin was washed with 35 column volumes of 1% SDS in PBS and 10 column volumes of 0.1% SDS in ddH_2O. The DADPS tag was cleaved by incubating the resin in 5% formic acid in 0.1% SDS in ddH_2O for 1 hr. Columns were washed with 5 column volumes of 0.1% SDS in H_2O, during which proteins remained bound, and proteins were subsequently eluted in 15 column volumes of 1% SDS in PBS. Protein enrichment was confirmed by SDS-PAGE, and eluted proteins were concentrated on 3kDa MWCO spin filters (Amicon).

In-gel digestion – Concentrated proteins were separated on precast 4–12% polyacrylamide gels (Life Technologies) and visualized with colloidal blue stain (Life Technologies). Lanes were cut into 8 slices and proteins were destained, reduced, alkylated, digested with LysC (Mako), and extracted as described in Bagert et al. (19). Extracted peptides were desalted with custom-packed C_{18} columns as described in Rappsilber et al. (35), lyophilized, and resuspended in 0.1% formic acid (Sigma).

Liquid Chromatography-Mass spectrometric analyses – Liquid chromatography-mass spectrometry and data analyses were carried out on an EASY-nLC-Orbitrap mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) as previously described with the following modifications (36). For the EASY-nLC II system, solvent A consisted of 97.8% H_2O, 2% ACN, and 0.2% formic acid and solvent B consisted of 19.8% H_2O, 80% ACN, and 0.2% formic acid. For the LC-MS/MS experiments, samples were loaded at a flow rate of 500 nL/min onto a 16-cm

analytical HPLC column (75 µm ID) packed in-house with ReproSil-Pur C_{18}AQ 3 µm resin (120 Å pore size, Dr. Maisch, Ammerbuch, Germany). The column was enclosed in a column heater operating at 30 °C. After ca. 20 min of loading time, the peptides were separated with a 60 min gradient at a flow rate of 350 nL/min. The gradient was as follows: 0–30% Solvent B (50 min), 30–100% B (1 min), and 100% B (8 min). The Orbitrap was operated in data-dependent acquisition mode to alternate automatically between a full scan (*m*/*z*=300–1700) in the Orbitrap and subsequent 10 CID MS/MS scans in the linear ion trap. CID was performed with helium as collision gas at a normalized collision energy of 35% and 30 msec of activation time.

Protein Quantification and Ratio Statistics – Thermo RAW files were processed with MaxQuant (v. 1.4.1.2) using default parameters and LysC/P as the enzyme. Additional variable modifications for Met were Aha (-4.9863), L-2,4-diaminobutanoate (-30.9768), a product of Aha reduction, alkyne-DADPS (+835.4300), and 5-hexyn-1-ol (+93.0868), a product of alkyne-DADPS cleavage. Multiplicity was set to 2, and light and heavy (+8.0142) lysine labels were specified for all experiments. Aha and 5-hexyn-1-ol modifications were included in protein quantification. We required protein quantifications to be calculated with at least two evidences for each set of experiments.

Both pooled variances and bootstrap statistical methods were employed as previously described to estimate the individual protein ratio standard errors (19, 37). First, pooled estimates of peptide variation were calculated separately for peptides with well-characterized ratios and those based on requantification in MaxQuant. Second, standard errors of the overall protein ratios were calculated by generating 1000 bootstrap iterations. These iterations were generated by resampling the replicates and peptides and adding a small amount of random variation to each measurement based on the pooled variance estimates. Once the bootstrapped

samples were generated for each protein, the standard error of the protein ratio was calculated from the standard deviation of the bootstrapped iterations. Using the standard error, proteins with ratios significantly different from 1:1 were identified using a Z-test and *p*-values were adjusted to account for multiple hypothesis testing using the Benjamini and Hochberg method (38).

Bioinformatic Analysis – Hierarchical clustering was performed with R (v. 3.1.1) using Ward's method (39). Confidence intervals (95^{th} percentile) for cluster time-series data were calculated by a bootstrapping approach using the tsplot function from the Python (v. 2.7) module seaborn (v. 0.4.0). Singular value decomposition was computed for PCA with the Python module matplotlib.mlab (v. 1.4.0). Gene ontology analysis was performed using a combination of GO terms and Kegg orthology and module terms. Group scores were defined as the mean of protein distances from the origin of the PCA biplot (PC1 vs. PC2). Statistical cutoffs (p-value < 0.05) were calculated using a bootstrapping approach that calculates scores for 100,000 groups randomly selected from the total pool of quantified proteins. Cutoffs were calculated individually for each group size (n = 4, 5, etc.) and groups with fewer than 4 members were excluded. Version 9.1 of the STRING database was used for identifying protein interactions, and interacting networks were identified by manual inspection (40).

Acknowledgements

Michael J. Sweredoski and Sonja Hess assisted with proteomics experiments and analysis. Lihui Feng performed qRT-PCR experiments. Julia C. van Kessel performed qRT-PCR and EMSA experiments. Julia C. van Kessel and Bonnie L. Bassler helped design experiments and assisted in composing the manuscript.

This work was supported by the Howard Hughes Medical Institute, National Institutes of Health (NIH) Grant 5R01GM065859 and National Science Foundation (NSF) Grant MCB-0343821 to BLB, and by NIH Grant R01GM062523 and the Institute for Collaborative Biotechnologies through grant W911NF-09-0001 from the U.S. Army Research Office to DAT. The content of the paper does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

Figures and Tables

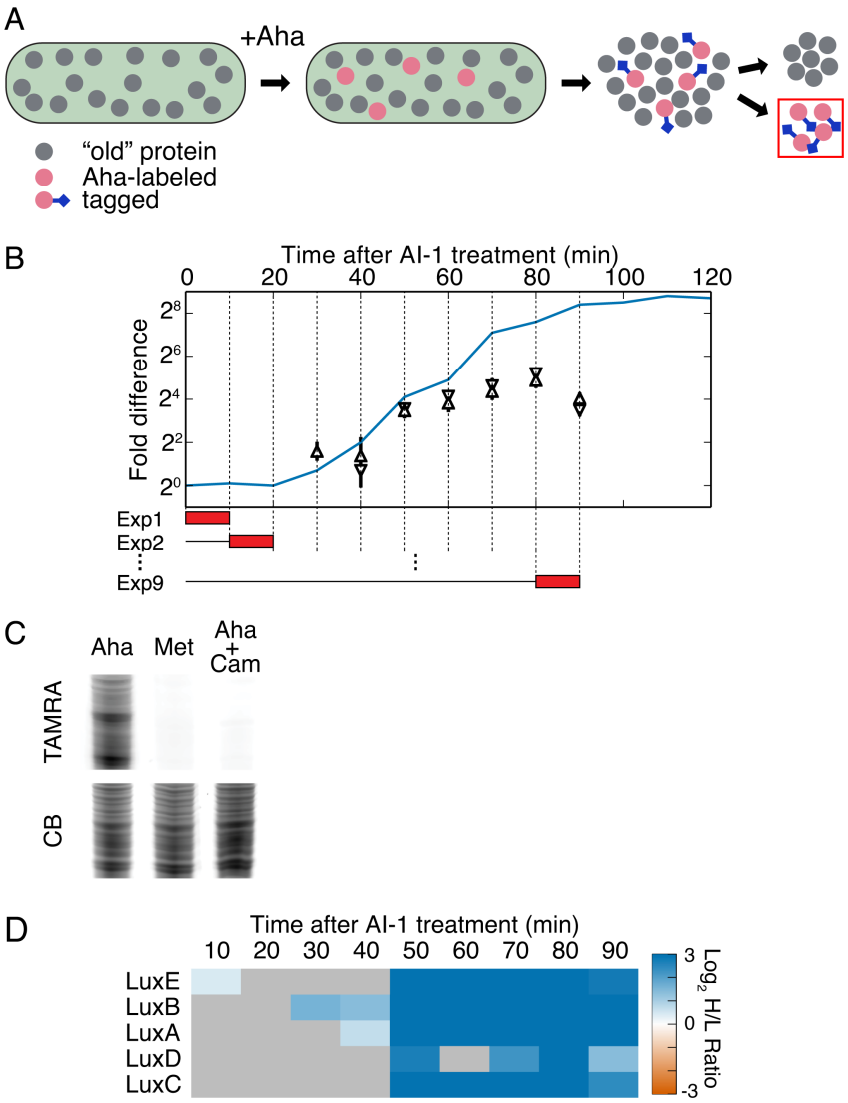


Figure 4.1. BONCAT analysis of quorum sensing. (A) Treatment of *V. harveyi* TL25 with Aha allows selective tagging and enrichment of newly synthesized proteins. (B) Schematic of BONCAT experiments. The blue line shows bioluminescence emission after AI-1 treatment. Red boxes represent the duration of Aha pulses in separate experiments (Exp1, Exp2, etc.). BONCAT quantification of the luciferase subunits LuxA (pyramid) and LuxB (reverse pyramid). Error bars

denote standard errors of the mean. All proteomic experiments were performed in triplicate. (C) Labeled proteins from a 10-min Aha pulse were conjugated to an alkyne-TAMRA dye and visualized by in-gel fluorescence. Cultures were treated with Aha, Met, or with Aha together with the protein synthesis inhibitor Cam. CB denotes Colloidal Blue staining. (D) Heat map showing measured protein production from the *lux* operon. Gray boxes denote samples in which the protein could not be reliably quantified. Chloramphenicol, Cam; TAMRA, tetramethylrhodamine.

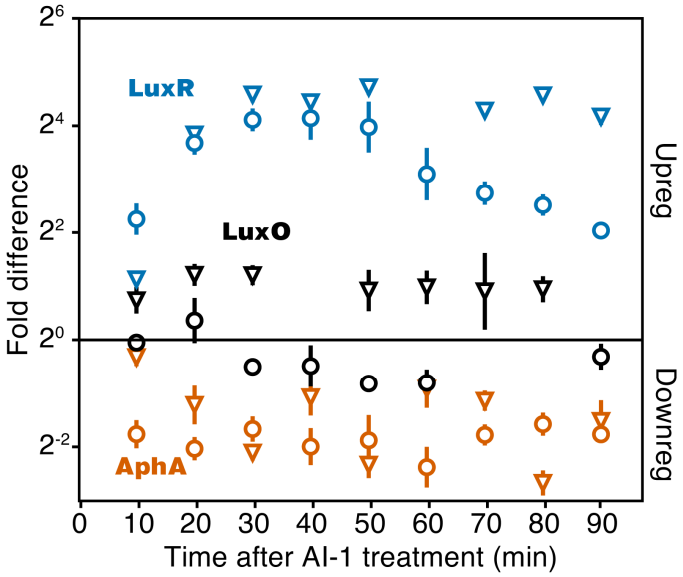


Figure 4.2. Detection of major quorum-sensing components. Quantitation of LuxR, AphA, and LuxO protein (reverse pyramids) and mRNA (circles). Fold changes are calculated as the difference between cultures treated and not treated with AI-1. LuxR protein quantification was confirmed by manual inspection of MS-MS spectra and calculated peptide retention times (Fig. 4.S2). Error bars show sample standard errors of the mean.

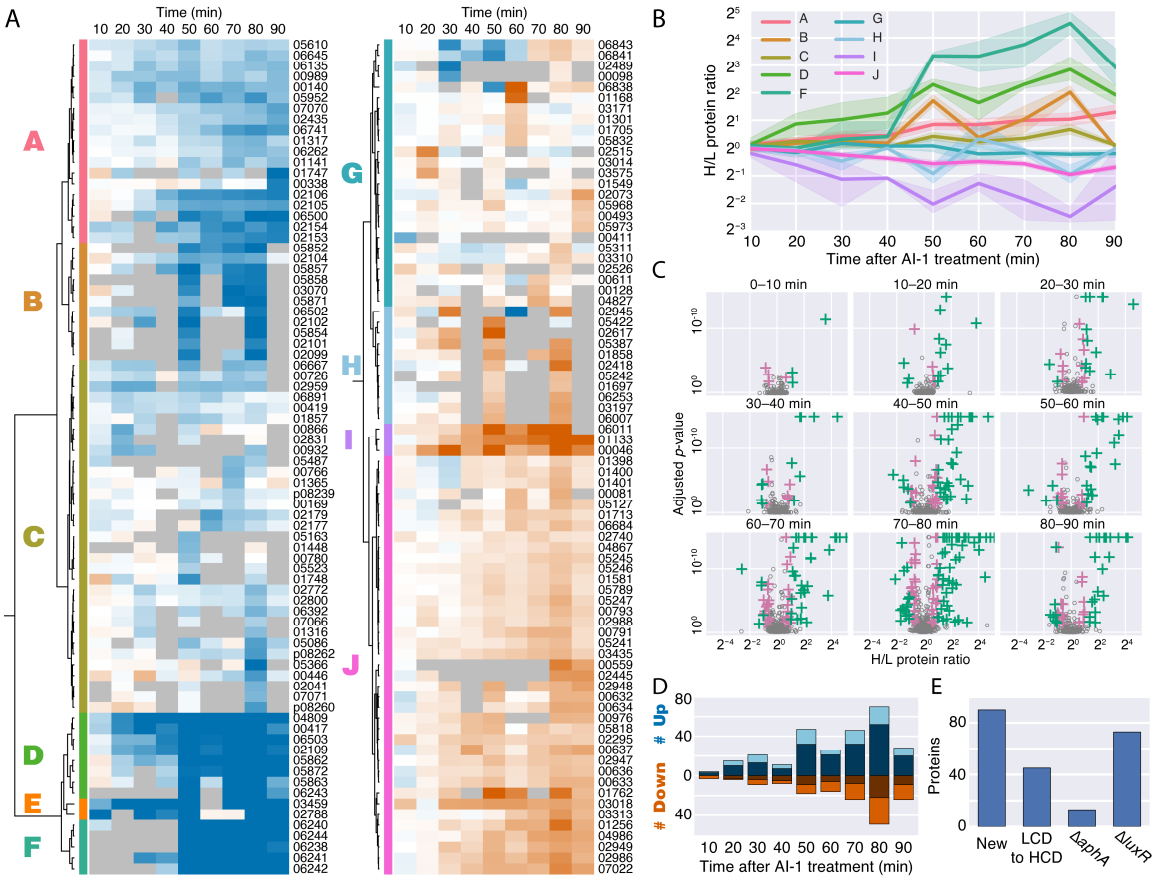


Figure 4.3. Identification of the quorum-sensing-regulated proteome. (A) Heatmap showing calculated abundances of significantly AI-1-regulated proteins, organized by unsupervised hierarchical clustering. Blue denotes up-regulation and orange denotes down-regulation. Missing ratios are denoted by gray boxes. Proteins with more than 6 missing time points were omitted from the clustering analysis. (B) Temporal behavior of protein clusters. Shaded regions denote 95th percentile confidence intervals. (C) Volcano plots showing outlier proteins for each time point. Proteins with adjusted *p*-values less than 0.05 (Benjamini-Hochberg FDR) are marked by plus signs. Significant proteins with H/L ratios greater than 2-fold and between 1.5- and 2-fold are designated by green and pink markers, respectively. (D) Total numbers of up-regulated (blue) and down-regulated (orange) proteins at each time point. Dark blue and dark orange

portions of bars represent proteins up- and down-regulated more than 2-fold. (E) Identification of new proteins controlled by quorum sensing, and the numbers of outlier proteins identified that belong to the previously established *aphA*, *luxR*, and quorum-sensing (LCD to HCD) regulons.

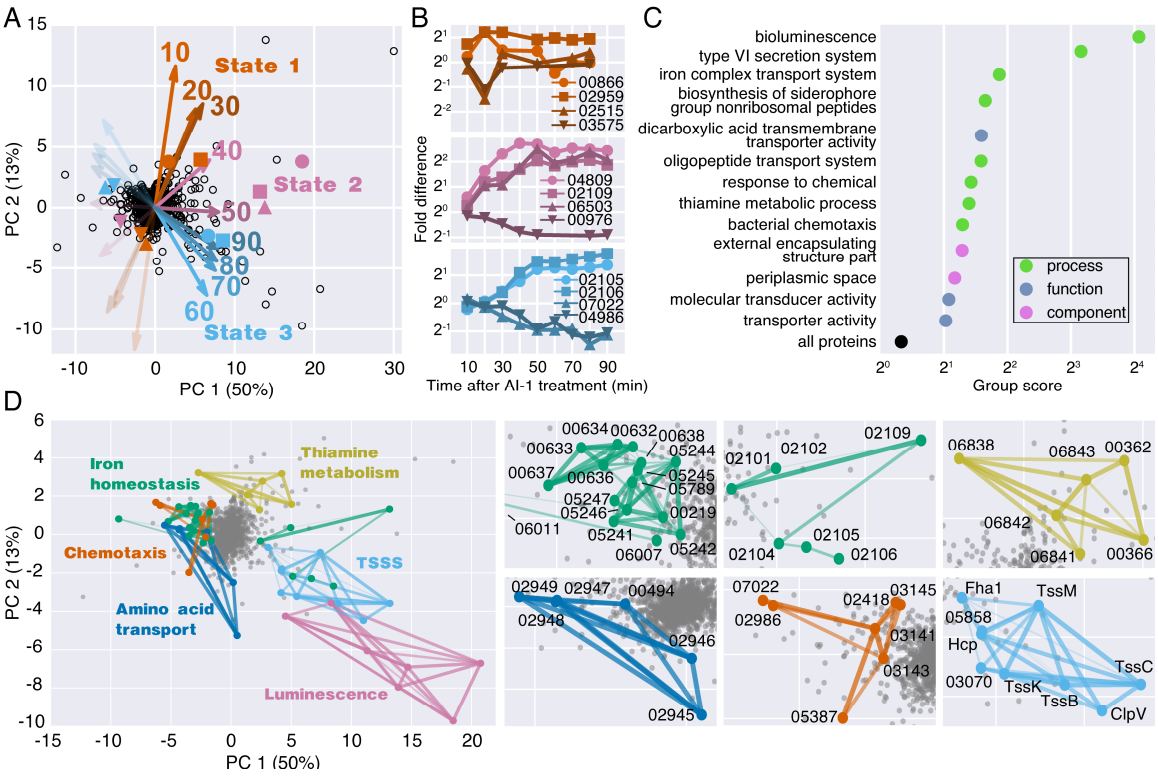


Figure 4.4. Bioinformatic analysis of the quorum-sensing-regulated proteome. (A) Principal component (PC) analysis of the time-course proteomics data. Percentages show the fractions of the variance for each PC. Vectors show the weights for each time point used to project protein ratios onto the PCs, providing a visual representation of the contribution of each time point to a protein's position in PC space. Weight vectors are positively scaled by a factor of 20 for visibility. Orange, pink, and blue vectors highlight distinct, time-dependent proteomic states after AI-1 treatment. (B) Time-course production profiles of select proteins, represented by colored markers in the PCA biplot, show different timings of protein regulation in response to AI-1 treatment. (C) Gene ontology groups controlled by quorum sensing. Groups were assigned a score based on their member positions on the PCA plot. Significantly AI-1-regulated, non-redundant groups are shown (p-value < 0.05). (D) Identification of functionally related, and similarly AI-1-regulated protein groups. A select set of STRING interacting networks were

mapped onto the PCA plot with strength and confidence of interactions represented by line thickness and opacity.

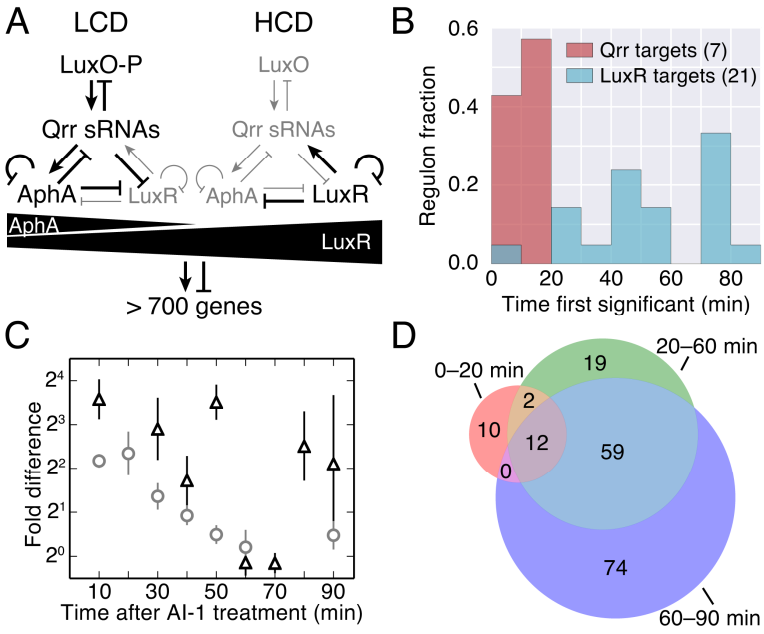


Figure 4.5. Analysis of the timing of quorum-sensing-regulated protein changes. (A) Diagram of the cytoplasmic portion of the quorum-sensing signal transduction pathway in *V. harveyi*. The horizontal black triangles represent concentration gradients of AphA and LuxR. Panel adapted from van Kessel et al. (15). (B) Timing of changes of proteins whose genes are direct targets of the Qrr sRNAs or LuxR. Proteins regulated by both the Qrr sRNAs and LuxR were excluded from the list of LuxR targets. (C) Protein (triangles) and mRNA (circles) measurements of VIBHAR_02788 following AI-1 treatment. Error bars designate standard error of the mean. (D) Venn diagram showing the numbers of proteins regulated at early, intermediate, and late times after AI-1 treatment.

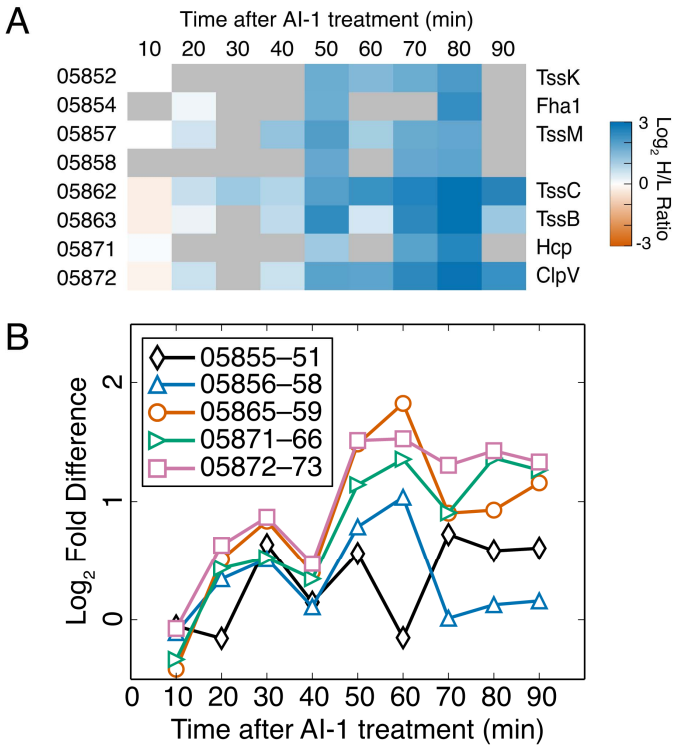


Figure 4.6. Type VI secretion is controlled by quorum sensing in *V. harveyi*. (A) Heatmap of all quantified TSSS proteins. Values of protein quantification are represented by the color bar. Gray boxes denote values that could not be reliably quantified. (B) RT-PCR of the five TSSS gene clusters after AI-1 treatment.

Table 4.1. Proteins regulated between 0 and 20 min after AI-1 treatment.

Gene locus	Peptides	Protein description	0–10 min	10–20 min
VIBHAR_00411	2	acetolactate synthase	2.22	ND
VIBHAR_00419	7	bifunctional protein GlmU	1.08	1.53
VIBHAR_00866	3	chemotaxis protein	1.17	2.33
VIBHAR_00932	2	DNA polymerase III subunit beta	1.38	2.17
VIBHAR_00989	4	MurNAc-6-P etherase	1.23	1.73
VIBHAR_02109	12	non-ribosomal peptide synthetase	1.20	2.28
VIBHAR_02295	5	fumarate/nitrate reduction transcriptional regulator	-1.55	-1.12
VIBHAR_02515	1	uncharacterized protein	-1.22	-2.80
VIBHAR_02526	3	uncharacterized protein	-1.51	-1.08
VIBHAR_02788	11	chemotaxis protein	11.94	ND
VIBHAR_02831	2	type III secretion protein	1.21	2.04
VIBHAR_03014	4	superoxide dismutase	-1.04	-1.63
VIBHAR_03256	2	uncharacterized protein	-1.79	ND
VIBHAR_03575	2	putative Holliday junction resolvase	-1.03	-2.17
VIBHAR_04809	11	uncharacterized protein	1.51	3.11
VIBHAR_05607	3	chitinase	ND	3.37
VIBHAR_06502	17	ATPase	1.36	1.65
VIBHAR_06503	30	peptidase	1.12	2.07
VIBHAR_06891	3	ecotin	1.30	1.59

Table 4.2. Proteins newly associated with quorum sensing.

Functional association	Proteins
Transcription factor	6
Secretion	8
Metabolism	19
Iron homeostasis	8
Chemotaxis	4
Molecular transport	11
Kinase	4
Protease	3
Other	8
Unknown	19

Supplementary Figures and Tables

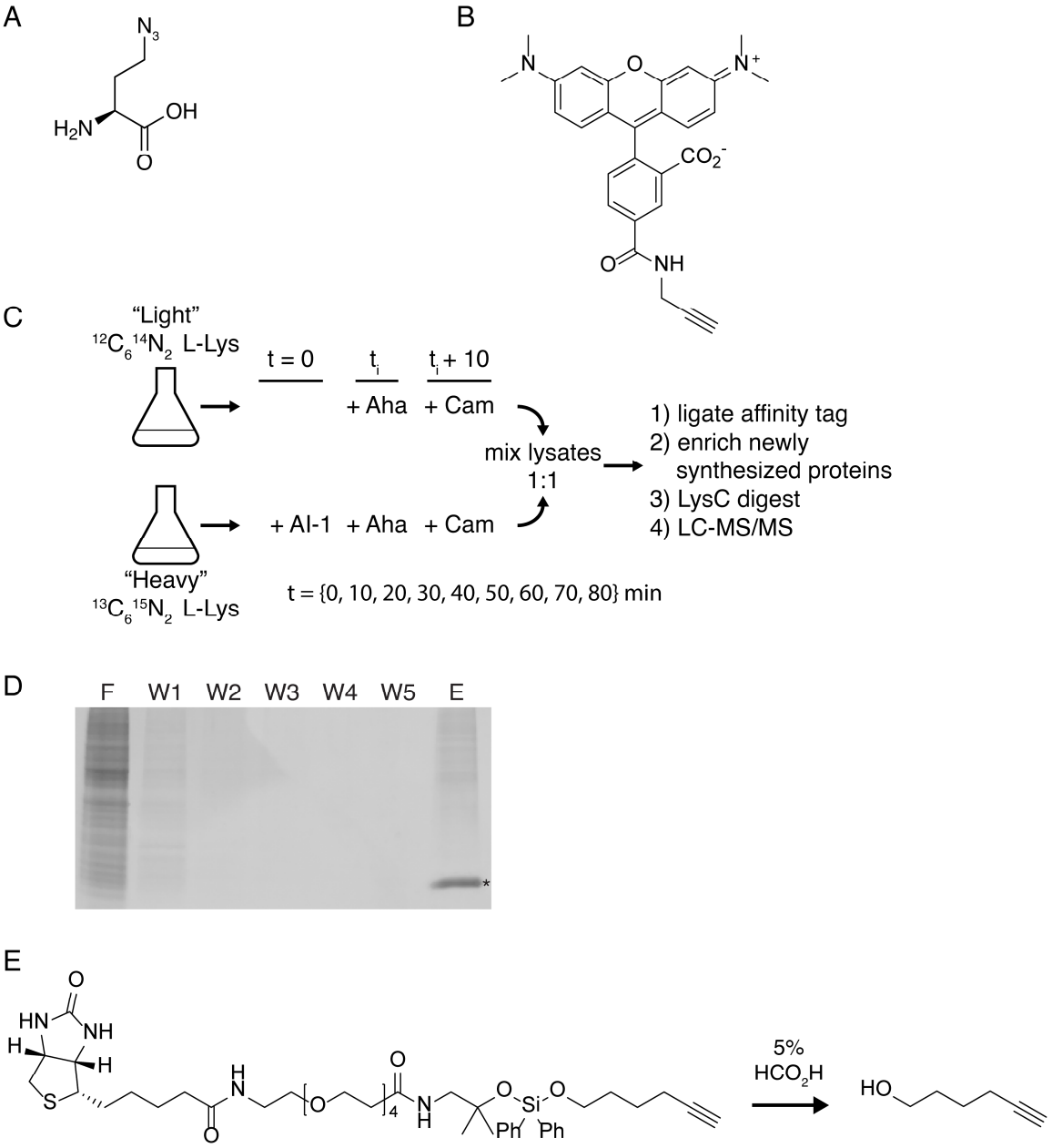
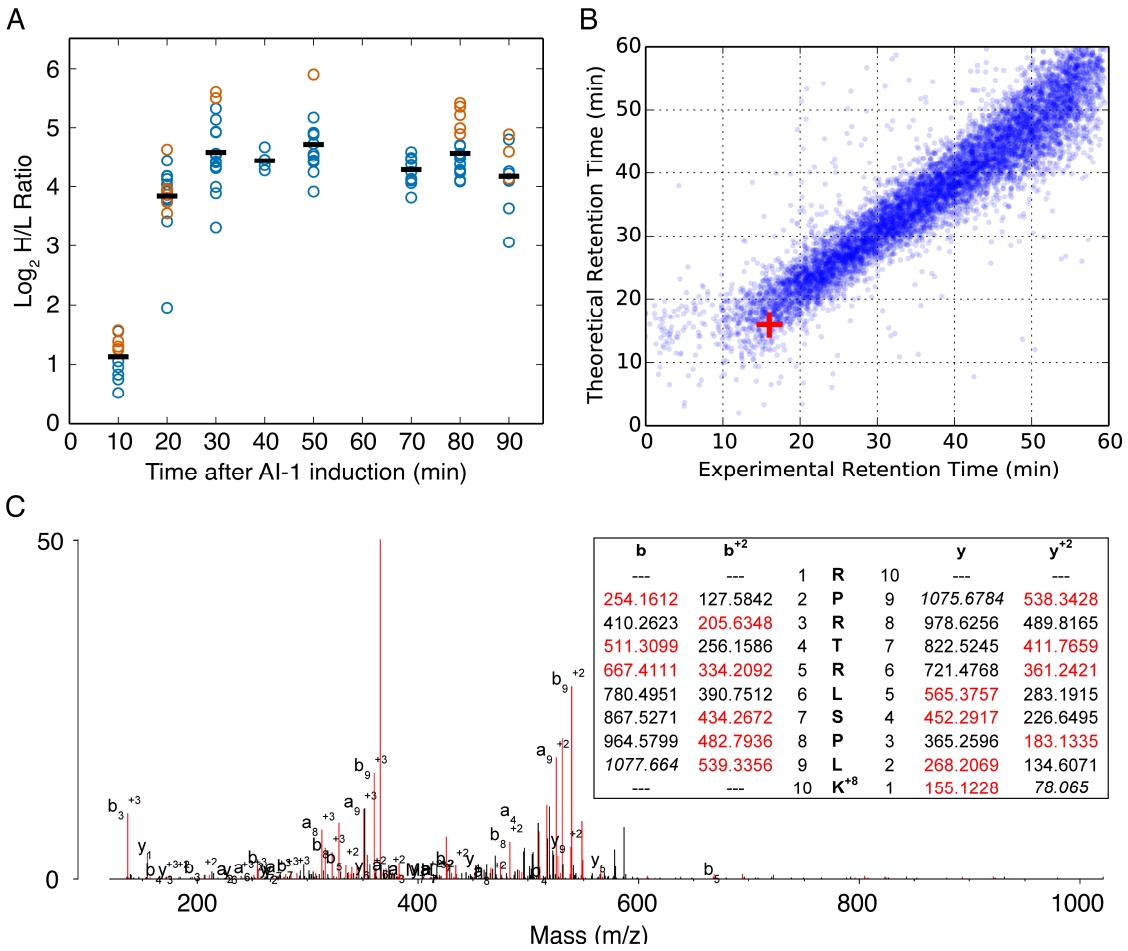


Figure 4.S1. Chemical structures of reagents used and diagram of BONCAT-SILAC experiments. (A) The structure of L-azidohomoalanine (Aha). (B) The structure of TAMRA-alkyne. (C) A combined BONCAT-SILAC approach for quantifying differences in protein translation. Reference

cultures were not treated with AI-1 at time 0 min. Otherwise, they were treated identically to experimental cultures. Experiments were performed in triplicate with one isotope label swap experiment. (D) Gel showing enrichment of Aha-labeled proteins using the DADPS tag. F – flow-through, W1-5 – washes, E – elution. The band marked by * is monomeric avidin. (E) The structures of the alkyne DADPS tag and the alkyne fragment released upon cleavage.



b	b^{+2}				y	y^{+2}
---	---	1	R	10	---	---
254.1612	127.5842	2	P	9	*1075.6784*	538.3428
410.2623	205.6348	3	R	8	978.6256	489.8165
511.3099	256.1586	4	T	7	822.5245	411.7659
667.4111	334.2092	5	R	6	721.4768	361.2421
780.4951	390.7512	6	L	5	565.3757	283.1915
867.5271	434.2672	7	S	4	452.2917	226.6495
964.5799	482.7936	8	P	3	365.2596	183.1335
1077.664	539.3356	9	L	2	268.2069	134.6071
---	---	10	K^{+8}	1	155.1228	*78.065*

Figure 4.S2. Confirmation of LuxR peptide (RPRTRLSPLK) quantitation. (A) Masses in the range of 1222.7622 ± 2 ppm, the predicted mass of the RPRTRLSPLK peptide. Orange and blue markers represent normalized ratios of peptides from label swap experiments. (B) An additive model of polypeptide chromatography accurately predicts the retention time of the RPRTRLSPLK peptide. The measured and calculated retention times were 16.02 min and 15.09 min, respectively. (C) Fragmentation spectra of candidate masses in the 1222.7622 ± 2 ppm range were matched to the RPRTRLSPLK peptide by ProteinProspector (v 5.12.4). Red text and lines denote matched fragmentation spectra of the RPRTRLSPLK peptide.

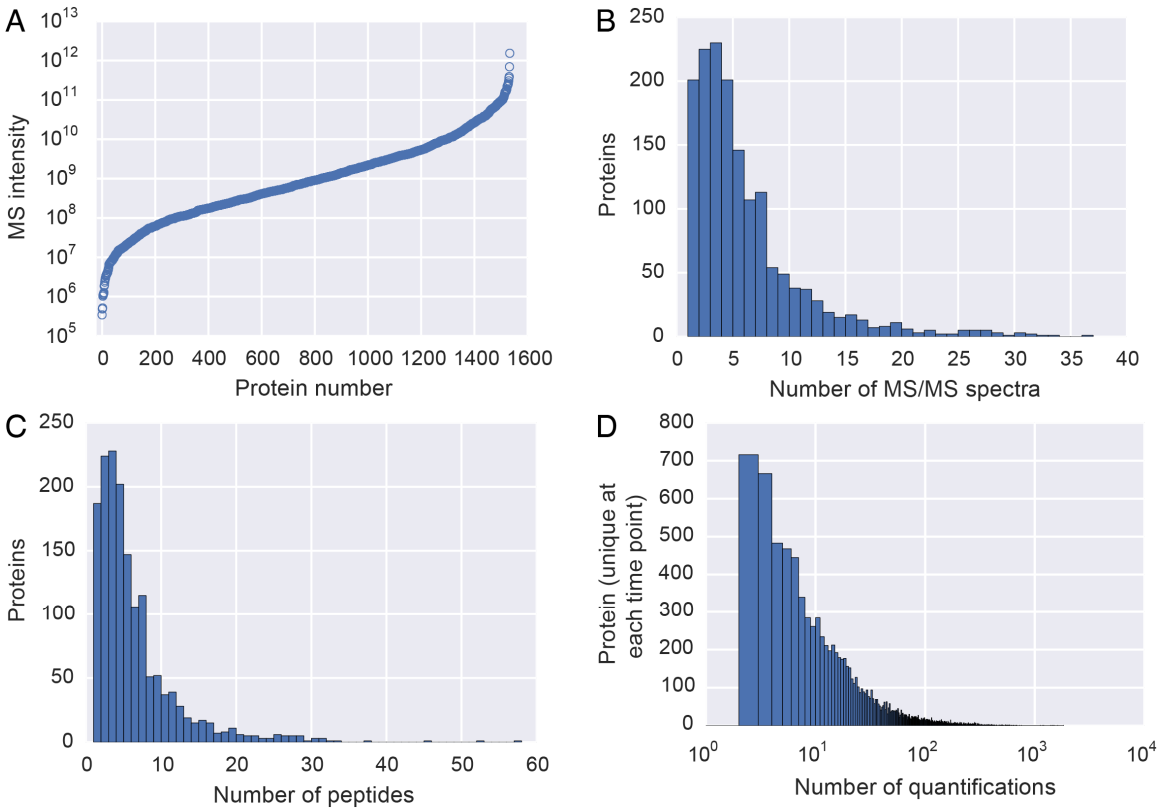


Figure 4.S3. Summary of measurements from proteomics experiments. (A) Sorted list of MS intensities for all quantified proteins. (B) MS/MS spectra per protein. Number of peptides (C) and quantifications (D) for each protein, calculated separately for each time point.

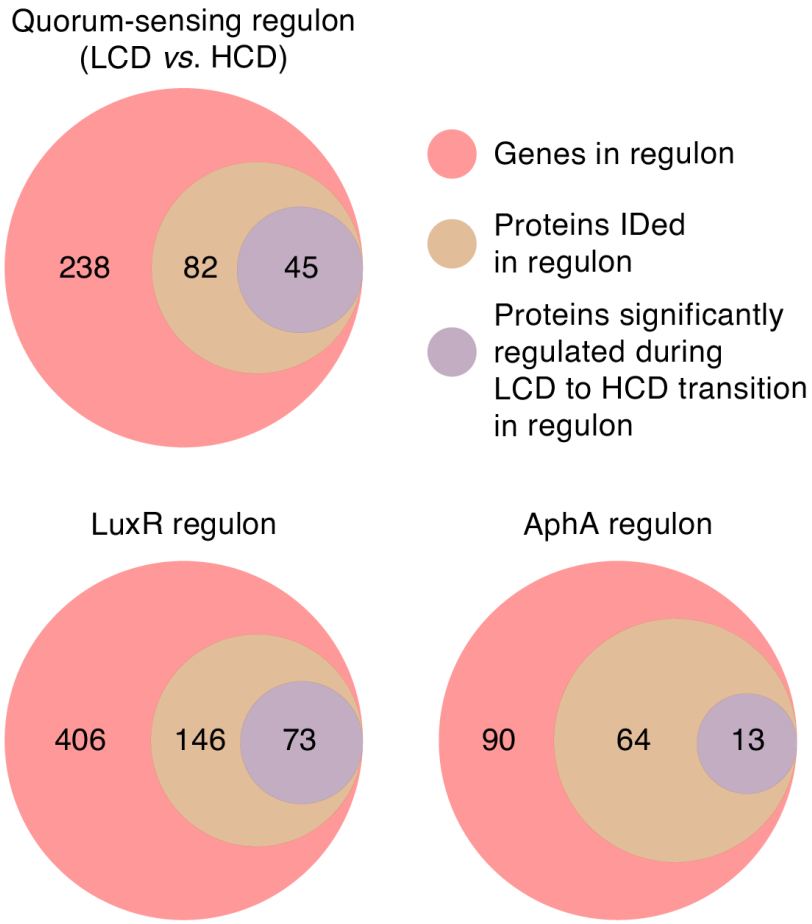


Figure 4.S4. Comparison of BONCAT proteomics data with the previously measured LuxR, AphA, and quorum-sensing regulons. For each regulon, the subset of genes for which proteins were identified with and without significant regulation is designated.

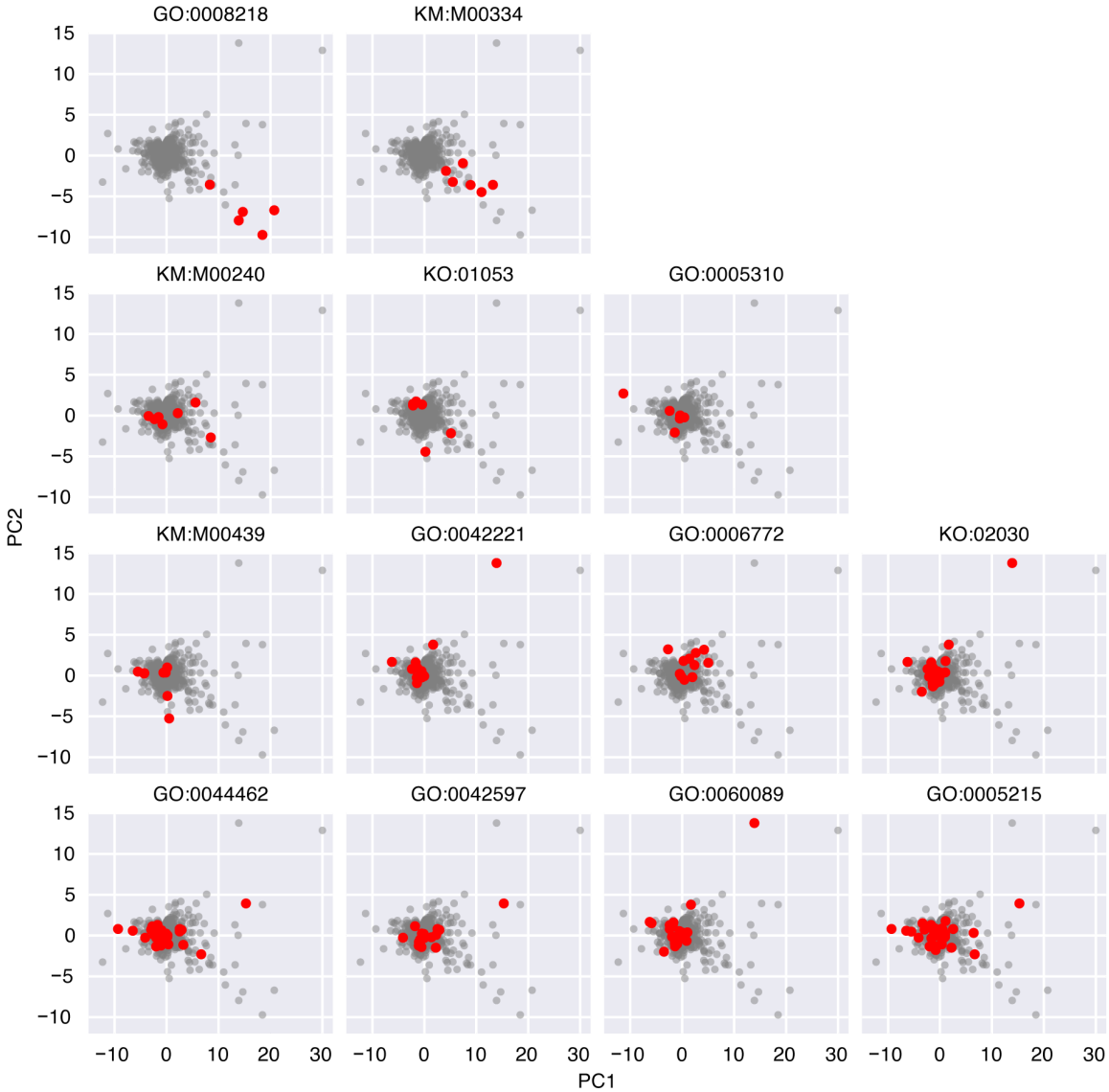


Figure 4.S5. Gene ontology analysis. Proteins from the significantly regulated gene ontology groups (Fig. 4.6*B*) are shown on the PCA plot. Groups were scored based on the average

distance of proteins from the origin, and groups with fewer than 4 members were excluded. Ontology analysis used a combination of groups from the Gene Ontology (GO) database, and the Kegg Orthology (KO) and Kegg Module (KM) databases.

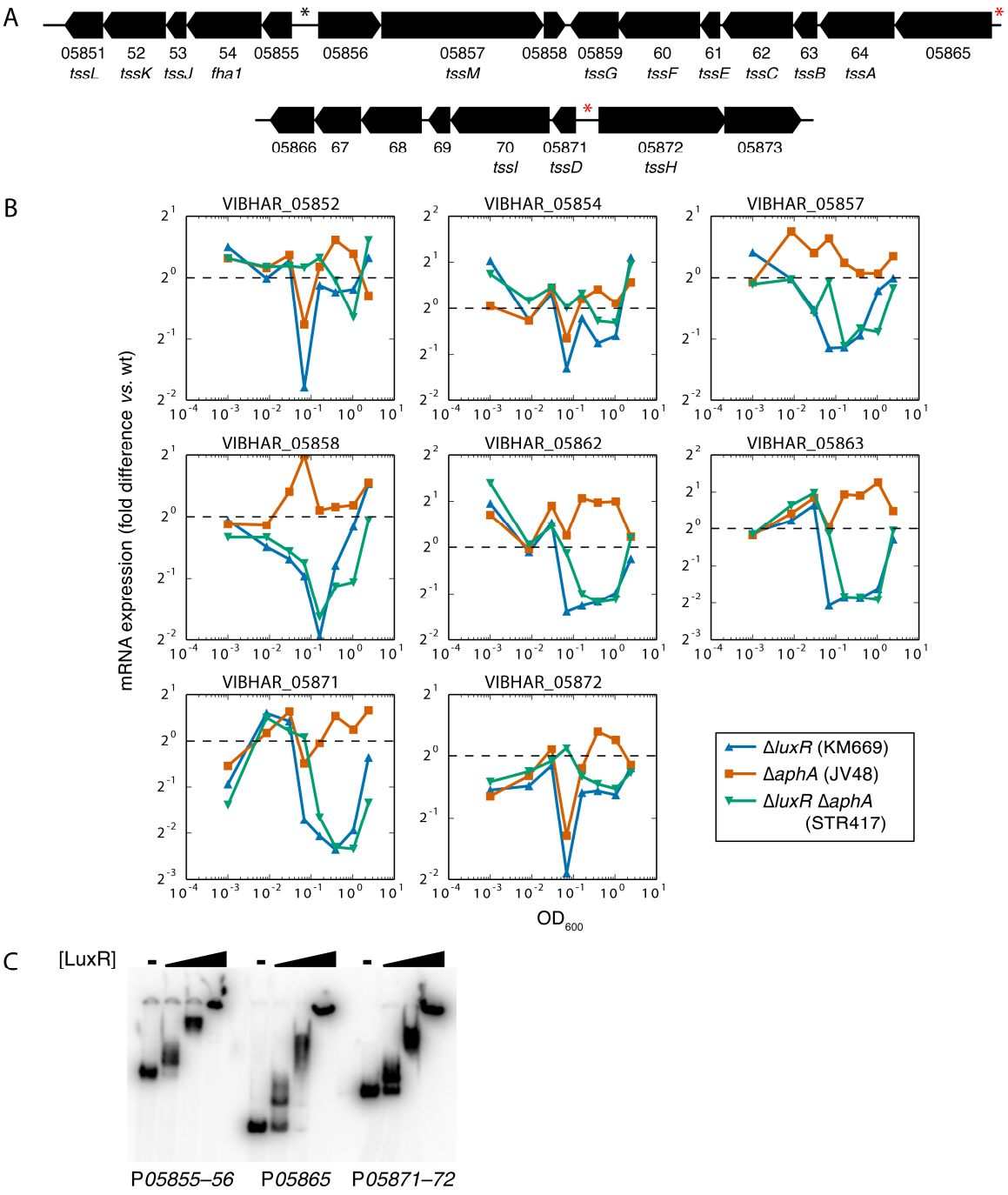


Figure 4.S6. Type VI secretion is LuxR-regulated. (A) The type VI secretion genes in *V. harveyi* are organized into five putative operons. The black asterisk symbol marks the location of the LuxR binding site previously identified by ChIP. Red askterisk symbols mark newly identified LuxR binding sites. (B) Upregulation of type VI secretion operons at HCD is LuxR-dependent and

AphA-independent. Results are from van Kessel et al. (2013). These data show relative gene expression values from Δ*aphA* (JV48), Δ*luxR* (KM669), and Δ*aphA* Δ*luxR* (STR417) *V. harveyi* strains relative to wild-type (BB120; wt) at varying cell densities. (C) EMSAs for reaction mixtures containing 0, 10, 100, or 1000 nM LuxR incubated with 1 nM radiolabeled DNA substrate corresponding to the three TSSS promoter regions for *VIBHAR_05855–56*, *VIBHAR_05865*, or *VIBHAR_05871–72*.

Table 4.S1. Quantified proteins from BONCAT experiments in *V. harveyi*. This table is available as a downloadable excel file in the online version of this thesis.

Table 4.S2. Calculation of Aha incorporation.

Measure of Abundance	Aha Peptides	Met Peptides	All Peptides	Calculated Aha Incorporation
Evidence Counts	26,496	131,808	158,304	16.7%
MS-MS Counts	23,285	147,918	171,203	13.6%
MS Intensity	5.09×10^{11}	2.72×10^{12}	3.23×10^{12}	15.8%

Table 4.S3. The weights used to transform protein ratios into principal component space. The mean (μ) and standard deviation (σ) of each sample were used to standardize the original variables prior to multivariate analysis.

	10 min	20 min	30 min	40 min	50 min	60 min	70 min	80 min	90 min	Variance accounted for
PC1	0.128	0.254	0.300	0.343	0.399	0.325	0.378	0.387	0.393	50%
PC2	0.583	0.407	0.428	0.196	-0.018	-0.360	-0.259	-0.220	-0.168	13%
PC3	0.782	-0.491	-0.258	-0.127	0.023	0.224	0.109	0.038	0.018	9%
PC4	-0.162	-0.652	0.385	0.526	0.092	0.026	-0.257	-0.212	0.064	7%
PC5	0.042	0.296	-0.473	0.523	-0.273	0.392	-0.119	-0.386	0.147	6%
PC6	-0.020	0.088	0.446	-0.335	-0.181	0.736	-0.202	-0.091	-0.239	5%
PC7	0.029	-0.080	0.294	-0.100	-0.684	-0.135	0.406	-0.152	0.472	4%
PC8	-0.031	-0.047	0.042	0.321	-0.175	0.018	0.593	0.032	-0.713	3%
PC9	0.045	-0.022	-0.039	0.228	-0.475	-0.015	-0.375	0.757	-0.061	3%
μ	0.001	-0.004	0.028	0.003	0.050	0.025	0.052	0.047	0.021	-
σ	0.203	0.250	0.311	0.259	0.451	0.381	0.444	0.575	0.375	-

Table 4.S4. Oligonucleotides used in this study.

Name	Sequence
P05855-56	GGGCGAAAGATATCAAGTCTCTCTT
P05855-56	ATTTTCCAATTCCAACTGATTATATGAAGG
P05865	GTTGCTCTTCACTAGCGCTCTTG
P05865	CCTTGTTTCAAGGCTGGTATTTAAA
P05871-22	ATATGCTGGAGTTGGCATCGTTATT
P05871-22	TTTATTCTTTAGAGGAAAAGAGGTGGTC
aphA qRT-PCR	ATCCATCAACTCTAGGTGATAAAC
aphA qRT-PCR	CGTCGCGAGTGCTAAGTACA
luxO qRT-PCR	GCATTCCTGATCTTATTCTGCTCG
luxO qRT-PCR	TCCATCCCCGTCATATCAGGTA
luxR qRT-PCR	GCAAAGAGACCTCGTACTAGG
luxR qRT-PCR	GCGACGAGCAAACACTTC
02788 qRT-PCR	TGTTTAACAGTATACGTACTCGAATCG
02788 qRT-PCR	TCAGTAAATGCATCGGTAGTCAT
05853 qRT-PCR	CAACAGGCATTACGCCAG
05853 qRT-PCR	CGCAAAATAACTGGAGAGGATTG
05857 qRT-PCR	CGATTTTGGTTTCTACCATAGTGG
05857 qRT-PCR	CAATCCATAGATATAGCTTATCTGCATCT
05861 qRT-PCR	GTTATCGTCTTTTAGAGCGGATTG
05861 qRT-PCR	AGGTGAGAATGAATGGATTCGAT
05864 qRT-PCR	GATTGATATCGAACGTTTGCTTACG
05864 qRT-PCR	CTTCACTTCGGATTCCATCATTTC
05871 qRT-PCR	GTGAAACTCAAGGTCACATCAC
05871 qRT-PCR	AGTTCTTGAACTAGGAACTCATCAA
05872 qRT-PCR	GTCTCTGAAAAAGCAAACGAAGT
05872 qRT-PCR	GATTGTCTTAATAAGTTCTCAGAACAATCA

References

1. Miller MB, Bassler BL (2001) Quorum sensing in bacteria. *Annu Rev Microbiol* 55:165–99.

2. Zhu J, Mekalanos JJ (2003) Quorum Sensing-Dependent Biofilms Enhance Colonization in Vibrio cholerae. *Dev Cell* 5:647–56.

3. Jayaraman A, Wood TK (2008) Bacterial Quorum Sensing: Signals, Circuits, and Implications for Biofilms and Disease. *Annu Rev Biomed Eng* 10:145–67.

4. Cao J-G, Meighen EA (1989) Purification and structural identification of an autoinducer for the luminescence system of Vibrio harveyi. *J Biol Chem* 264:21670–6.

5. Ng W-L et al. (2011) Signal production and detection specificity in Vibrio CqsA/CqsS quorum-sensing systems. *Mol Microbiol* 79:1407–17.

6. Chen X et al. (2002) Structural identification of a bacterial quorum-sensing signal containing boron. *Nature* 415:545–9.

7. Freeman JA, Lilley BN, Bassler BL (2000) A genetic analysis of the functions of LuxN: a two-component hybrid sensor kinase that regulates quorum sensing in Vibrio harveyi. *Mol Microbiol* 35:139–49.

8. Henke JM, Bassler BL (2004) Three Parallel Quorum-Sensing Systems Regulate Gene Expression in Vibrio harveyi. *J Bacteriol* 186:6902–14.

9. Bassler BL, Wright M, Silverman MR (1994) Multiple signalling systems controlling expression of luminescence in Vibrio harveyi: sequence and function of genes encoding a second sensory pathway. *Mol Microbiol* 13:273–86.

10. Freeman JA, Bassler BL (1999) Sequence and Function of LuxU: a Two-Component Phosphorelay Protein That Regulates Quorum Sensing in Vibrio harveyi. *J Bacteriol* 181:899–906.

11. Freeman JA, Bassler BL (1999) A genetic analysis of the function of LuxO, a two-component response regulator involved in quorum sensing in Vibrio harveyi. *Mol Microbiol* 31:665–77.

12. Tu KC, Bassler BL (2007) Multiple small RNAs act additively to integrate sensory information and control quorum sensing in Vibrio harveyi. *Genes Dev* 21:221–33.

13. Rutherford ST, van Kessel JC, Shao Y, Bassler BL (2011) AphA and LuxR/HapR reciprocally control quorum sensing in vibrios. *Genes Dev* 25:397–408.

14. Shao Y, Bassler BL (2012) Quorum-sensing non-coding small RNAs use unique pairing regions to differentially control mRNA targets. *Mol Microbiol* 83:599–611.

15. Van Kessel JC, Rutherford ST, Shao Y, Utria AF, Bassler BL (2013) Individual and Combined Roles of the Master Regulators AphA and LuxR in Control of the Vibrio harveyi Quorum-Sensing Regulon. *J Bacteriol* 195:436–43.

16. Dieterich DC, Link AJ, Graumann J, Tirrell DA, Schuman EM (2006) Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT). *Proc Natl Acad Sci U S A* 103:9482–7.

17. Dieterich DC et al. (2007) Labeling, detection and identification of newly synthesized proteomes with bioorthogonal non-canonical amino-acid tagging. *Nat Protoc* 2:532–40.

18. Kiick KL, Saxon E, Tirrell DA, Bertozzi CR (2002) Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation. *Proc Natl Acad Sci U S A* 99:19–24.

19. Bagert JD et al. (2014) 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* 13:1352–8.

20. Swartzman E, Silverman M, Meighen EA (1992) The luxR Gene Product of Vibrio harveyi Is a Transcriptional Activator of the lux Promoter. *J Bacteriol* 174:7490–3.

21. Ong S-E et al. (2002) Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. *Mol Cell Proteomics* 1:376–86.

22. Tu KC, Long T, Svenningsen SL, Wingreen NS, Bassler BL (2010) Negative Feedback Loops Involving Small Regulatory RNAs Precisely Control the Vibrio harveyi Quorum-Sensing Response. *Mol Cell* 37:567–79.

23. Teng S-W et al. (2011) Active regulation of receptor ratios controls integration of quorum-sensing signals in Vibrio harveyi. *Mol Syst Biol* 7:491.

24. Tu KC, Waters CM, Svenningsen SL, Bassler BL (2008) A small-RNA-mediated negative feedback loop controls quorum-sensing dynamics in Vibrio harveyi. *Mol Microbiol* 70:896–907.

25. Feng L et al. (2015) A Qrr Noncoding RNA Deploys Four Different Regulatory Mechanisms to Optimize Quorum-Sensing Dynamics. *Cell* 160:228–40.

26. He Q, Chen J, Li C (2011) An Extracellular Oligopeptide Permease May Be a Potential Virulence Factor of Vibrio harveyi. *J Ocean Univ China* 10:343–50.

27. Shao Y, Feng L, Rutherford ST, Papenfort K, Bassler BL (2013) Functional determinants of the quorum-sensing non-coding RNAs and their roles in target regulation. *EMBO J* 32:2158–71.

28. Van Kessel JC, Ulrich LE, Zhulin IB, Bassler BL (2013) Analysis of Activator and Repressor Functions Reveals the Requirements for Transcriptional Control by LuxR , the Master Regulator of Quorum Sensing in Vibrio harveyi. *MBio* 4:e00378–13.

29. Mougous JD, Gifford CA, Ramsdell TL, Mekalanos JJ (2007) Threonine phosphorylation post-translationally regulates protein secretion in Pseudomonas aeruginosa. *Nat Cell Biol* 9:797–803.

30. Salomon D, Gonzalez H, Updegraff BL, Orth K (2013) Vibrio parahaemolyticus Type VI Secretion System 1 Is Activated in Marine Conditions to Target Bacteria, and Is Differentially Regulated from System 2. *PLoS One* 8:e61086.

31. Shao Y, Bassler BL (2014) Quorum regulatory small RNAs repress type VI secretion in Vibrio cholerae. *Mol Microbiol* 92:921–30.

32. Link AJ, Vink MKS, Tirrell DA (2007) Preparation of the functionalizable methionine surrogate azidohomoalanine via copper-catalyzed diazo transfer. *Nat Protoc* 2:1879–83.

33. Hong V, Presolski SI, Ma C, Finn MG (2009) Analysis and Optimization of Copper-Catalyzed Azide-Alkyne Cycloaddition for Bioconjugation. *Angew Chemie* 48:9879–83.

34. Szychowski J et al. (2010) Cleavable Biotin Probes for Labeling of Biomolecules via Azide-Alkyne Cycloaddition. *J Am Chem Soc* 132:18351–60.

35. Rappsilber J, Mann M, Ishihama Y (2007) Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat Protoc* 2:1896–906.

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

37. Pierce NW et al. (2013) Cand1 Promotes Assembly of New SCF Complexes through Dynamic Exchange of F Box Proteins. *Cell* 153:206–15.

38. Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. *J Roy Stat Soc B Met* 57:289–300.

39. Ward JH (1963) Hierarchical Grouping to Optimize an Objective Function. *J Am Stat Assoc* 58:236–44.

40. Jensen LJ et al. (2009) STRING 8 – a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res* 37:D412–6.

APPENDIX A

CELL-SELECTIVE IDENTIFICATION OF NEWLY SYNTHESIZED PROTEINS IN MAMMALIAN CELLS USING AN ENGINEERED AMINOACYL TRNA SYNTHETASE

Abstract

Cell-selective identification of newly synthesized proteins will facilitate studies of biological processes in multicellular organisms. We have previously developed methods to identify newly synthesized proteins in a cell-selective manner in bacterial cells; however, comparable strategies for proteomic analysis in mammalian cells are not available. Here we introduce a mutant mammalian methionyl tRNA synthetase (AnlRS) enzyme that incorporates the non-canonical amino acid azidonorleucine (Anl) into newly synthesized proteins. Controlled expression of this enzyme restricts protein labeling to cells of interest and facilitates identification by mass spectrometry. This approach does not require expression of orthogonal tRNAs or the removal of canonical amino acids. Successful Anl incorporation in a variety of mammalian cells demonstrates the potential utility of this enzyme for a wide range of proteomic studies.

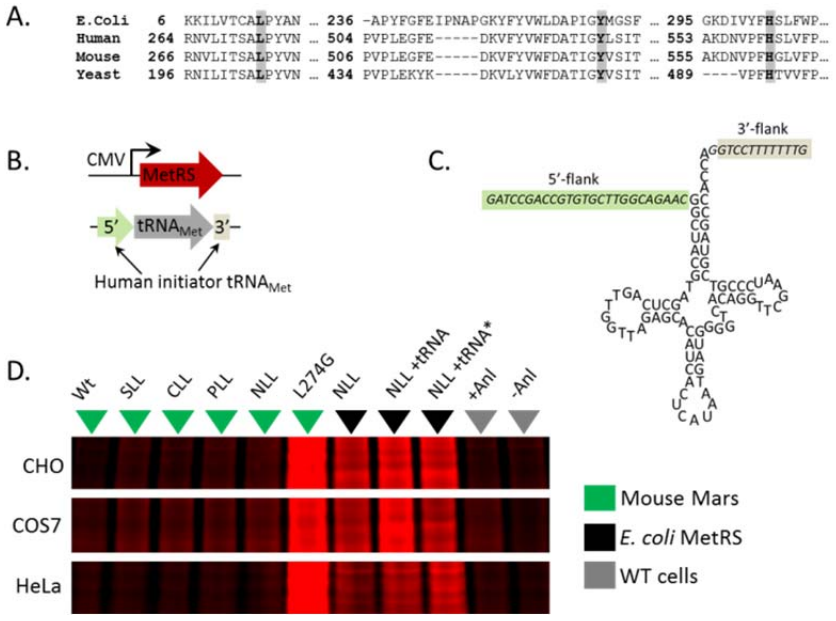


Figure A.1. Identification of MetRS variants for the metabolic incorporation of Anl in mammalian cells. (A) Sequence alignment of various MetRS enzymes with positions 13/260/301 of the *E. coli* MetRS highlighted. (B) Constructs for the expression of MetRS and *E. Coli* $tRNA^{Met}$ under the CMV promoter and an internal promoter of eukaryotic tRNAs, respectively. (C) *E. coli* $tRNA^{Met}$ coding sequence containing 5' and 3' flanking sequences from human initiator $tRNA^{Met}$. (D) In-gel fluorescence image shows TAMRA labeling, which indicates Anl incorporation in cellular proteins. tRNA* designates the presence of a 3' CCA tail in the tRNA coding sequence. The wild type mouse MetRS enzyme is indicated by WT and the ±Anl conditions correspond to empty vector labeling controls.

This work is reported in a manuscript currently under preparation:

Mahdavi, A., Jindal, G. A., Bagert, J. D., Dong, D., Sweredoski, M. J., Hess, S., Tirrell, D. A. Cell-selective identification of newly synthesized proteins in mammalian cells using an engineered aminoacyl tRNA synthetase.

APPENDIX B

A QRR NONCODING RNA DEPLOYS FOUR DIFFERENT REGULATORY MECHANISMS TO OPTIMIZE QUORUM-SENSING DYNAMICS

Abstract

Quorum sensing is a cell-cell communication process that bacteria use to transition between individual and social lifestyles. In vibrios, homologous small RNAs called the Qrr sRNAs function at the center of quorum-sensing pathways. The Qrr sRNAs regulate multiple mRNA targets including those encoding the quorum-sensing regulatory components luxR, luxO, luxM, and aphA. We show that a representative Qrr, Qrr3, uses four distinct mechanisms to control its particular targets: the Qrr3 sRNA represses luxR through catalytic degradation, represses luxM through coupled degradation, represses luxO through sequestration, and activates aphA by revealing the ribosome binding site while the sRNA itself is degraded. Qrr3 forms different base-pairing interactions with each mRNA target, and the particular pairing strategy determines which regulatory mechanism occurs.Combined mathematicalmodeling and experiments show that the specific Qrr regulatory mechanism employed governs the potency, dynamics, and competition of target mRNA regulation, which in turn, defines the overall quorum-sensing response.

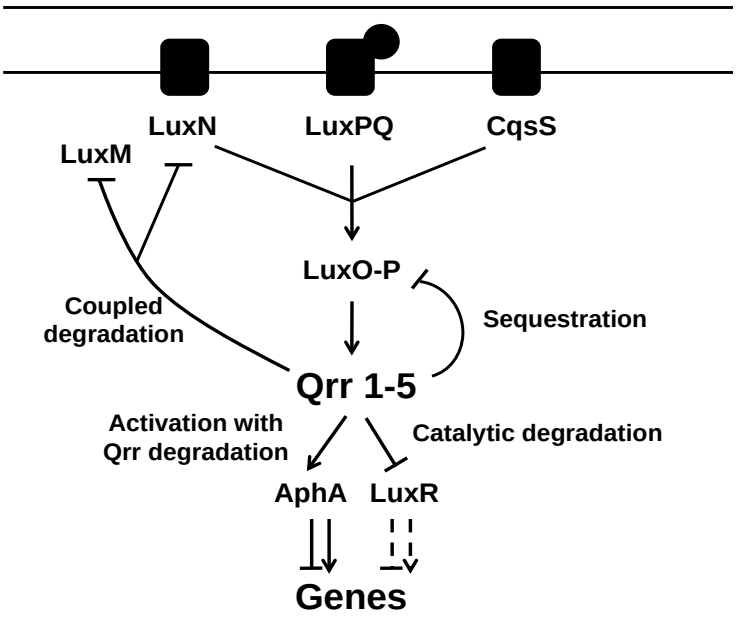


Figure B.1. Schematic for how a *V. harveyi* quorum-sensing Qrr sRNA uses four regulatory mechanisms to control target mRNAs. At low cell density, the three quorum-sensing receptors LuxN, LuxPQ, and CqsS transfer phosphate through LuxU (not shown) to LuxO. Phosphorylated LuxO activates transcription of genes encoding five sRNAs called Qrr1–Qrr5. Using Qrr3 as the representative quorum-sensing regulatory sRNA, we show that the Qrr sRNA catalytically represses the high-cell-density master regulator *luxR*. The Qrr sRNA represses *luxO* through sequestration. The Qrr sRNA represses the *luxMN* operon through coupled degradation (*luxM* encodes the synthase that produces the ligand for LuxN). The Qrr sRNA also activates translation of the low-cell-density master regulator *aphA*; base-pairing with the *aphA* mRNA leads to Qrr degradation.

This work was reported in:

Feng, L., Rutherford, S. T., Papenfort, K., Bagert, J. D., van Kessel, J. C., Tirrell, D. A., Wingreen, N. S., and Bassler, B. L. (2015) A Qrr Noncoding RNA Deploys Four Different Regulatory Mechanisms to Optimize Quorum-Sensing Dynamics. *Cell 160*, 228–40.

APPENDIX C

COMPARISON OF FAMILY 9 CELLULASES FROM MESOPHILIC AND THERMOPHILIC BACTERIA

Abstract

Cellulases containing a family 9 catalytic domain and a family 3c cellulose binding module (CBM3c) are important components of bacterial cellulolytic systems. We measured the temperature dependence of the activities of three homologs: *Clostridium cellulolyticum* Cel9G, *Thermobifida fusca* Cel9A, and *C. thermocellum* Cel9I. To directly compare their catalytic activities, we constructed six new versions of the enzymes in which the three GH9-CBM3c domains were fused to a dockerin both with and without a *T. fusca* fibronectin type 3 homology module (Fn3). We studied the activities of these enzymes on crystalline cellulose alone and in complex with a miniscaffoldin containing a cohesin and a CBM3a. The presence of Fn3 had no measurable effect on thermostability or cellulase activity. The GH9-CBM3c domains of Cel9A and Cel9I, however, were more active than the wild type when fused to a dockerin complexed to scaffoldin. The three cellulases in complex have similar activities on crystalline cellulose up to 60°C, but *C. thermocellum* Cel9I, the most thermostable of the three, remains highly active up to 80°C, where its activity is 1.9 times higher than at 60°C. We also compared the temperature-dependent activities of different versions of Cel9I (wild type or in complex with a miniscaffoldin) and found that the thermostable CBM is necessary for activity on crystalline cellulose at high temperatures. These results illustrate the significant benefits of working with thermostable enzymes at high temperatures, as well as the importance of retaining the stability of all modules involved in cellulose degradation.

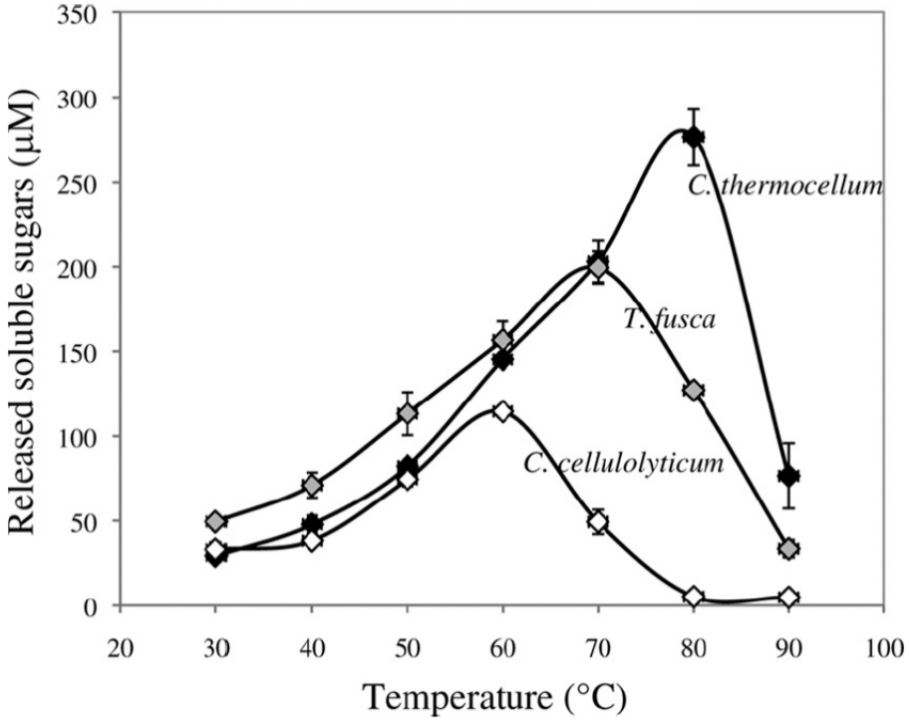


Figure C.1. Temperature dependence of the activities of three homologous Cel9 cellulases in complex with a miniscaffoldin on Avicel. Curves are labeled as follows: white for *C*. *cellulolyticum* Cel9G-do/ScafC3, gray for *T*. *fusca* Cel9A-do/ScafC3, and black for *C*. *thermocellum* Cel9I-do/ScafC3. Bars indicate the standard deviations of two independent experiments. Released soluble sugars were assayed after 20 min with 100 nM complex and on 10 g/liter Avicel.

This work was reported in:

Mingardon, F., Bagert, J. D., Maisonnier, C., Trudeau, D. L., and Arnold, F. H. (2011) Comparison of Family 9 Cellulases from Mesophilic and Thermophilic Bacteria. *Appl*. *Environ*. *Microb*. *77*, 1436–42.

APPENDIX D

CLEAVABLE BIOTIN PROBES FOR LABELING OF BIOMOLECULES VIA AZIDE-ALKYNE CYCLOADDITION

Abstract

The azide-alkyne cycloaddition provides a powerful tool for bio-orthogonal labeling of proteins, nucleic acids, glycans, and lipids. In some labeling experiments, e.g., in proteomic studies involving affinity purification and mass spectrometry, it is convenient to use cleavable probes that allow release of labeled biomolecules under mild conditions. Five cleavable biotin probes are described for use in labeling of proteins and other biomolecules via azide-alkyne cycloaddition. Subsequent to conjugation with metabolically labeled protein, these probes are subject to cleavage with either 50 mM $Na_2S_2O_4$, 2% $HOCH_2CH_2SH$, 10% HCO_2H, 95% CF_3CO_2H, or irradiation at 365 nm. Most strikingly, a probe constructed around a dialkoxydiphenylsilane (DADPS) linker was found to be cleaved efficiently when treated with 10% HCO_2H for 0.5 h. A model green fluorescent protein was used to demonstrate that the DADPS probe undergoes highly selective conjugation and leaves a small (143 Da) mass tag on the labeled protein after cleavage. These features make the DADPS probe especially attractive for use in biomolecular labeling and proteomic studies.

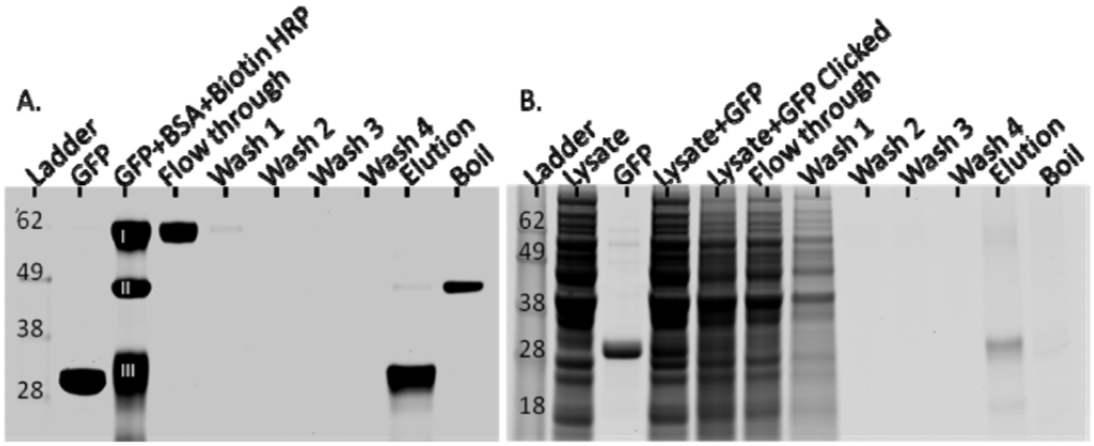


Figure D.1. Polyacrylamide gel analysis of affinity purification of 7-Hpg-GFP. (A) After click reaction with probe **12**, affinity purification from a solution containing BSA (band I), biotinylated HRP (band II), and 7-Hpg-GFP (band III), was performed on a streptavidin agarose resin. Wash 1 was performed with 1% SDS in PBS, wash 2 was performed with 6 M urea in 250 mM ammonium bicarbonate, wash 3 was performed with 1 M NaCl in PBS, and wash 4 was performed with 0.1% SDS in water to remove PBS. Elution with 5% formic acid was followed by boiling of resin in 2% SDS in PBS to remove remaining proteins. (B) Affinity purification of 7-Hpg-GFP from a DH10B lysate. Wash and elution conditions were identical to those in (A). Gels were stained with Coomassie Brilliant Blue.

This work was reported in:

Szychowski, J., Mahdavi, A., Hodas, J. J. L., Bagert, J. D., Ngo, J. T., Landgraf, P., Dieterich, D. C., Schuman, E. M., and Tirrell, D. A. (2010) Cleavable Biotin Probes for Labeling of Biomolecules via Azide-Alkyne Cycloaddition. *J. Am. Chem. Soc. 132*, 18351–60.