THE ROLE OF ACETYLATION IN MODULATION OF SALMONELLA ENTERICA PHYSIOLOGY

by

ANASTACIA ROSE PARKS

(Under the Direction of Jorge C. Escalante-Semerena)

ABSTRACT

The post-translational modification of acetylation is a form of chemical modification of proteins or small molecules that can rapidly affect a variety of cellular functions and is present in cells from all domains of life. Enzymes called acetyltransferases perform the acetylation reaction through the addition of an acyl group from the small molecule acetyl-CoA and adding it to an amine of a small molecule or protein. Two types of acetylation of amines can occur; first being N^{ε}-acetylation, where the epsilon associated amine of a lysyl side chain is acetylated. The other form of acetylation being N^{α}-acetylation, or addition of an acyl moiety to the primary amine of a small molecule or the N-terminal amine of a peptide chain. N^{ε}-acetylation can be reversed through the enzymatic activity of NAD^{+}-dependent sirtuin protein lysine deacetylases, that remove the acyl group from the acetyllysine using NAD^{+} as a co-substrate to generate *O*-acetyl-ADP-ribose and nicotinamide as byproducts.

This work provides new developments in the role acetylation or deacetylation of three specific areas: i) The role of reversible lysine acetylation (RLA) in modulation

of a toxin-antitoxin system in *S. enterica* that contributes to bacterial persistence. ii) Discovery and characterization of a novel N^{α}-acetyltransferase, NatA, that modulates the deacetylase activity of the NAD^{+}-dependent sirtuin protein lysine deacetylase CobB in *S. enterica*. iii) Characterization of the physiological significance of possessing two isoforms of CobB deacetylase in *S. enterica*, and lastly iv) The discovery and characterization of NatA-mediated N-terminal acetylation of the nucleoid-associated protein HU in *S. enterica*.

INDEX WORDS: *Salmonella*, Sirtuins, Deacetylase, Acetylation, N-terminal, persistence, Bacteria

THE ROLE OF ACETYLATION IN MODULATION OF SALMONELLA ENTERICA PHYSIOLOGY

by

ANASTACIA ROSE PARKS

BS, University of South Florida, 2014

MS, University of South Florida, 2015

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2021

THE ROLE OF ACETYLATION IN MODULATION OF SALMONELLA ENTERICA PHYSIOLOGY

by

ANASTACIA ROSE PARKS

Major Professor: Jorge C. Escalante- Semerena

Committee: Janet Westpheling
Diana Downs
Anna Karls

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
December 2021

DEDICATION

First, I would like to dedicate this dissertation to myself. I know that sounds selfish, but I worked really hard for this body of work. I love science, I love being a scientist, I love this. I could not ask for a better way to live my life. I fought so hard to get here, so I'm going to pat myself on the back this time.

I next would like to dedicate this dissertation to all of the strong, supportive women in my life. They made this dissertation possible. I needed all of them to keep me going, even when I felt everything in the world was going wrong. Thank you Libby, Chelsey, Jessica, Tori, Sophia, Abbie, Rachel, Shayla, and many, many other wonderful women in my life, this is to you.

Finally, I would like to dedicate this dissertation to my step-father, Larry, that is the greatest parent I have ever known. He is the most selfless human I have ever met and is a "model citizen". I could not have asked for a better role model and support in my life. Thank you so much, Larry, this is to you.

ACKNOWLEDGEMENTS

I would like to thank all the wonderful friends I have made in graduate school, and I really hope that I can maintain these friendships life-long.

I would like to thank my pets for being my biggest loves and helping me the only way they know how, to just love me. Thank you, Io, Ms. Pee, Sandi, and Buci. I love you all.

I would like to acknowledge my grandmother, Dorothy Edwards, for being my mother and for being the strong, caring woman that I loved and do love with all my heart. I would also like to acknowledge my father Carl, and thank him for his encouragement in my love of science.

I would like to thank Jorge for being the best mentor I could ever ask for. I would not be able to get this level of training anywhere else in the entire world. Jorge is a wild man, but his passion really shines through. I loved working on my dissertation work under his guidance, and I felt that we make a really great team. Thank you so much Jorge, I appreciate you noticing me past my low GRE scores, and seeing the passion for research that I have, and taking a chance on me. I know you hate to be acknowledged, but I don’t care. Thank you, Jorge! I will miss you and cherish the fun times.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION

1.1 OVERVIEW OF ACETYLATION AND DEACETYLATION

Cells possess a variety of regulatory techniques in order to adapt cellular function according to external stimuli. Among these regulatory mechanisms are transcriptional regulation, or the regulation of the production of messenger RNA (mRNA), translational regulation, being the regulation of the production of proteins from mRNA, and post-translational modification (PTM), which is typically the addition or removal of chemical moieties to proteins or small molecules already present in the cell [1]. A specific type of post-translational modification is that of acetylation, where the addition or removal of an acetyl moiety from acetyl-CoA to an amine group of a protein or small molecule and has shown to serve as a critical regulatory mechanism of many different cellular processes in all domains of life [2, 3]. Two forms of acetylation exist; i) reversible acetylation of the epsilon amino moiety of lysyl residues of proteins (N^{ε}-acetylation) to generate acetyllysine; ii) irreversible N^{α}-acetylation of primary amines, such as those on small molecules or the *N*-terminal amine of a peptide chain. Acetylation can be a non-enzymatic process via acetyl-phosphate, or via acetyl-CoA where alkaline pH can affect the lability of the thioester bond on acetyl-CoA to hydroxyl anions. When pH is >8, lysyl residues are deprotonated. The combination of these factors is conducive to allow

nonenzymatic transfer of the acetyl group to the epsilon amino moiety of lysine [4]. The body of this work focuses on enzymatically conducted N^{ε}- and N^{α}-acetylation of proteins, which is performed through the action of a family of enzymes called GCN5-related N-acetyltransferases (GNATs, PF00583) [5].

In cases reported this far, lysine acetylation modifies a protein's function and in most cases this modification can be reversed by a deacetylase, which conducts the removal of the acetyl moiety from lysine by two different mechanisms. The first mechanism of acetylation occurs through zinc-dependent lysine deacetylases, that uses a water ion, Zn(II) ion and an active site histidinyl residue to coordinate a nucleophilic attack against the carbonyl of acetyllysine, generating acetate as a reaction product (Pfam 08295) [6, 7] . The second type of deacetylase, often referred to as sirtuins (Pfam 02416), require nicotinamide-adenine-dinucleotide (NAD^+) as a co-substrate to remove the acetyl moiety from lysine, which generates *O*-acetyl-ADP-ribose *O*-AADPR) and nicotinamide (NM) as reaction products [8, 9].

Early work on acetylation began in yeast, where the NAD^+-dependent deacetylase, ySir2p (sirtuin) regulates transcriptional silencing by modulating the acetylation state of lysine-rich histone tails [9]. In eukaryotes, sirtuins participate in the regulation of gene expression, metabolism, stress response, and genome integrity maintenance [10-13]. The focus of this body of work revolves around the physiological contributions of protein N^{ε}- and N^{α}-acetylation in *S. enterica*. In *S. enterica*, sirtuin activity is needed to regulate the activities of adenine-monophosphate-forming acyl-CoA synthetases (*e.g.,* acetyl- and propionyl-CoA

synthetase, Acs and PrpE, respectively). Bacterial sirtuin CobB works in opposition to the protein acetyltransferase (Pat), that acylates (deactivates) Acs and PrpE, while CobB deacylates (reactivates) [14-16].

In regards to N^{α}-acetylation acetylation of proteins, not only in *S. enterica* but also in all bacteria, the only identified systems of acetyltransferase-mediated N^{α}-acetylation acetylation of proteins are i) that of N-terminal acetylation of protein ribosomal subunits by the rim acetyltransferases[17-19] and ii) N-terminal acetylation of the NAD^{+}-dependent sirtuin deacetylase CobB in *S. enterica* [20] described in this chapter 4 of this work and ii) The *N*-terminal acetylation of the nucleoid associated proteins HupA and HupB in *S. enterica*, which also is described in chapter 6 of this work.

1.2 DISSERTATION OUTLINE

The dissertation that follows outlines my efforts to increase our understanding of the field of acetylation and deacetylation in *S. enterica* through characterization and analysis of several different acetyltransferase enzymes and the sole NAD^{+}-dependent sirtuin deacetylase, CobB. Chapter 2 will provide a review on current literature of N-terminal acetylation in prokaryotes, specifically touching on topics such as the role of N-terminal acetylation of proteins in eukaryotes vs prokaryotes; the details outlining the known protein targets of N-terminal acetylation and the enzymes that acetylate; and finally recent advancements in this field and currently used methods and limitations to study N-terminal protein acetylation.

Chapter 3 details how reversible lysine acetylation is utilized by a toxin acetyltransferase called TacT, acting on its partner antitoxin, TacA, to allow the survival of *S. enterica* during unfavorable conditions via self-toxicity through acetylation of tRNA. This system is reversible through the deacetylation of the antitoxin protein by NAD^+-dependent sirtuin deacetylase, CobB.

Chapter 4 will discuss the characterization of a novel protein N-terminal acetyltransferase, named NatA, that acetylates the N-terminal amine of the long form of NAD^+-dependent sirtuin deacetylase, CobB. This modulates the enzymatic activity of $CobB_{Long}$, thus adding an additional layer of regulation to the modulation of enzymes involved in growth of *S. enterica* using low concentrations of acetate as a carbon and energy source.

The genome of the enteric bacterium *Salmonella enterica* contains one sirtuin-encoding gene, *cobB.* Interestingly, this bacterium synthesizes two CobB isoforms, which differ in size by a 37-residue *N*-terminal extension. Other enterobacteria (*e,g., Escherichia coli, Klebsiella pneumoniae, Yersinia enterocolitica*) also encode two isoforms of CobB, but others only encode the long isoform (*e.g., Erwinia amylovora*). Chapter 5 goes into further detail on characterization of the physiological implications of having two biologically active isoforms of CobB. The long isoform of CobB can bind to a variety of nucleic acid substrates, while the short isoform cannot. Chapter 5 also outlines a difference in growth utilization of the carbon source ethanolamine is observed depending on the concentration of different isoforms of CobB available to the cell. Finally, $CobB_L$ can

bind and float with liposomes, suggesting that perhaps the N-terminal extension of CobB is used to bind/interact with lipids at the cell membrane.

Chapter 6 goes into further characterization of physiologically relevant substrates for the novel N-terminal acetyltransferase, NatA. Chapter 6 demonstrates that NatA can acetylate the N-terminal amine of the nucleoid associated proteins HupA and HupB, which changes the way HupA and HupB interact with nucleic acids and modulate gene expression.

Chapter 7 summarizes the contributions of this work as it pertains to the effect acetylation has on regulation of the physiology of the bacterium *S. enterica*, as well as more broad aspects of the contribution of this work to novel protein N-terminal acetylation in prokaryotes.

Chapter 8 concludes the findings of this work as well as discusses future experiments to be done and hypotheses surrounding these regulatory systems in *S. enterica*.

The appendices of this dissertation describe a separate project from my work in the Escalante-Semerena laboratory that contribute to regulation of iron, thiamine, reversible lysine acetylation, acetate metabolism, and curli fiber production in *S. enterica*.

1.3 REFERENCES

1. Walsh, C.T., S. Garneau-Tsodikova, and G.J. Gatto, Jr., *Protein posttranslational modifications: the chemistry of proteome diversifications.* Angew Chem Int Ed Engl, 2005. **44**(45): p. 7342-72.

2. Glozak, M.A., et al., *Acetylation and deacetylation of non-histone proteins.* Gene, 2005. **363**: p. 15-23.

3. Soppa, J., *Protein acetylation in archaea, bacteria, and eukaryotes.* Archaea, 2010. **2010**.

4. VanDrisse, C.M. and J.C. Escalante-Semerena, *Protein Acetylation in Bacteria.* Annu Rev Microbiol, 2019. **73**: p. 111-132.

5. Vetting, M.W., et al., *Structure and functions of the GNAT superfamily of acetyltransferases.* Arch Biochem Biophys, 2005. **433**(1): p. 212-26.

6. Hodawadekar, S.C. and R. Marmorstein, *Chemistry of acetyl transfer by histone modifying enzymes: structure, mechanism and implications for effector design.* Oncogene, 2007. **26**(37): p. 5528-40.

7. Marmorstein, R., *Structure of histone deacetylases: insights into substrate recognition and catalysis.* Structure, 2001. **9**(12): p. 1127-33.

8. Denu, J.M., *Linking chromatin function with metabolic networks: Sir2 family of NAD(+)-dependent deacetylases.* Trends Biochem Sci, 2003. **28**(1): p. 41-8.

9. Moazed, D., *Enzymatic activities of Sir2 and chromatin silencing.* Curr Opin Cell Biol, 2001. **13**(2): p. 232-8.

10. Brachmann, C.B., et al., *The SIR2 gene family, conserved from bacteria to humans, functions in silencing, cell cycle progression, and chromosome stability.* Genes Dev., 1995. **9**: p. 2888-2902.

11. Guarente, L., *Sirtuins in aging and disease.* Cold Spring Harb. Symp. Quant. Biol., 2007. **72**: p. 483-488.

12. Starai, V.J., et al., *Short-chain fatty acid activation by acyl-coenzyme A synthetases requires SIR2 protein function in Salmonella enterica and Saccharomyces cerevisiae.* Genetics, 2003. **163**: p. 545-555.

13. Kwon, H.S. and M. Ott, *The ups and downs of SIRT1.* Trends Biochem. Sci., 2008. **33**: p. 517-525.

14. Garrity, J., et al., *N-lysine propionylation controls the activity of propionyl-CoA synthetase.* J Biol Chem, 2007. **282**(41): p. 30239-45.

15. Starai, V.J., et al., *Sir2-dependent activation of acetyl-CoA synthetase by deacetylation of active lysine.* Science, 2002. **298**(5602): p. 2390-2.

16. Starai, V.J. and J.C. Escalante-Semerena, *Identification of the protein acetyltransferase (Pat) enzyme that acetylates acetyl-CoA synthetase in Salmonella enterica.* J Mol Biol, 2004. **340**(5): p. 1005-12.

17. Isono, K. and S. Isono, *Ribosomal protein modification in Escherichia coli. II. Studies of a mutant lacking the N-terminal acetylation of protein S18.* Mol Gen Genet, 1980. **177**(4): p. 645-51.

18. Isono, S. and K. Isono, *Ribosomal protein modification in Escherichia coli. III. Studies of mutants lacking an acetylase activity specific for protein L12.* Mol Gen Genet, 1981. **183**(3): p. 473-7.

19. Yoshikawa, A., et al., *Cloning and nucleotide sequencing of the genes rimI and rimJ which encode enzymes acetylating ribosomal proteins S18 and S5 of Escherichia coli K12.* Mol Gen Genet, 1987. **209**(3): p. 481-8.

20. Parks, A.R. and J.C. Escalante-Semerena, *Modulation of the bacterial CobB sirtuin deacylase activity by N-terminal acetylation.* Proc Natl Acad Sci U S A, 2020. **117**(27): p. 15895-15901.

CHAPTER 2

LITERATURE REVIEW: PROTEIN N-TERMINAL ACYLATION: AN EMERGING FIELD IN PROKARYOTIC CELL PHYSIOLOGY

[1]Parks A.R. and Escalante-Semerena J.C. To be submitted to *Trends in Microbiology*.

2.1 ABSTRACT

N-terminal (Nt)-acylation is the irreversible addition of an acyl moiety to the terminal alpha amino group of a peptide chain. This type of modification alters the nature of the Nt, which can interfere with the function of the modified protein by disrupting protein interactions, function, localization, degradation, hydrophobicity, or charge. Nt acylation is found in all domains of life and is a highly common occurrence in eukaryotic cells. However, in prokaryotes very few cases of Nt acylation have been reported. It was once thought that Nt acylation of proteins was uncommon in prokaryotes, but recent evidence suggests that this modification may be more common than once realized. In this review, we discuss what is known about prokaryotic Nt acetylation and the acetyltransferases that are responsible, as well as recent advancements in this field and currently used methods to study Nt acetylation.

2.2 KEYWORDS - N-terminal acetylation, Nα-acetylation, post-translational modification, modulation of protein function, protein degradation

2.3 INTRODUCTION

Post-translational modifications (PTMs) allow for rapid changes in protein function in response to exogenous or endogenous stimuli. PTMs involve the addition or removal of chemical moieties to biological molecules, as well as can be proteolytic cleavage. A large variety of protein PTMs have been identified that include glycosylation, lipidation, S-nitrosylation, adenylylation, acylation,

ubiquitinylation, methylation, phosphorylation, ADP-ribosylation, SUMOylation, to name a few. Here we focus on the PTM of protein acetylation, specifically Nt acetylation.

Acyl moieties can be added to amino groups associated with the epsilon carbon of lysyl side chains of proteins, as well as the alpha associated amines on N termini of proteins (Fig. 2.1). Acetyltransferases are enzymes that catalyze the transfer of acyl moieties from acyl-CoA to amino groups of lysyl residues of proteins or small molecules. This type of acetylation is also referred to as N(epsilon) or Nε, and has been characterized in all domains of life.

Nε acetylation was first reported in eukaryotes in the context of gene silencing, where acetyl groups are added to Nε amino groups of lysyl residues of histone proteins, neutralizing the positive charge of lysine preventing interactions with the negative charge of the phosphate backbone of DNA, allowing relaxation of DNA and for transcription to occur [1].

The field of lysine acetylation in modulation of protein function in bacteria began with studies of the acetyl-CoA synthetase (Acs) enzyme, which is acetylated on its active-site lysine by the **p**rotein **a**cetyl**t**ranferase (Pat), effectively blocking the first half reaction catalyzed by Acs, *i.e.,* the conversion of acetate to acetyl-AMP [2]. The acetyl moiety on Acs^{Ac} is removed by the NAD^+-dependent CobB sirtuin deacylase, restoring Acs activity. Reversible lysine acetylation (RLA) of Acs appears to be controlled by the energy and carbon status of the cell. That is, elevated levels of NAD^+ reflect on a physiological state in which the cell that is actively growing, since reducing equivalents are readily disposed of, hence shifting

the NAD^+:NADH ratio to a value >1. The net result of this shift is a greater proton motive force (PMF) that fuels nutrient acquisition, motility, and ATP synthesis among other processes. Under such conditions, the demand for acetyl-CoA (Ac-CoA) increases for anabolic purposes hence it makes sense that acetate activation into Ac-CoA by Acs should increase. Under this scenario, it is expected that Acs remains fully active and unacetylated. Should the physiological conditions be reversed, that is, the NAD^+:NADH ratio is <1 and Ac-CoA levels drop, one expects Acs activity to drop since higher substrate amounts for Pat would result in acetylated, inactive Acs with the concomitant reduction of acetate activation. For a comprehensive review on lysine acetylation in bacteria, refer to the following references [3-5].

Some acetyltransferases can acetylate primary amino groups of small molecules, which can serve a variety of functions for the cell. The first characterizations of small molecule acetylation involved bacterial acetylation of antibiotic compounds, which resulted in antibiotic resistance [6]. Since this discovery, a wide-range of small molecules have been shown to be acetylated, which impacts a variety of cellular processes such as detoxification, toxin-antitoxin systems, anabolism of specific compounds, transcription to name a few. For a comprehensive review of small molecule acetylation in bacteria, see [7].

In this review we discuss current findings on Nt acetylation of proteins in bacteria, and address methodologies for the study of Nt acetylation and their limitations.

Nt acetylation involves the addition of an acetyl group to the alpha amino group of the N-terminal residue of a peptide chain. This type of modification is ubiquitous,

with ~60-80% of proteins in lower and higher eukaryotes (including humans) being acetylated on their N-termini. In stark contrast, Nt acetylation in bacteria has been studied mostly in the context of the ribosome, where some of the proteins are N-terminally acetylated by the Rim family acetyltransferases (discussed below). More recently, N-terminomics studies have shed light on the prevalence of Nt acetylation in bacteria, ranging from ~3-10% of the total proteins being N-terminally acetylated.

2.4 ROLE OF NT ACETYLATION IN EUKARYOTES

Hypotheses including The Ac/N-End rule

Even though the occurrence of Nt acetylation appears to be widespread, the physiological purpose it serves remains debated. Some studies suggest that N-terminal acetylation of proteins in eukaryotes provides stability to proteins and prevents protein degradation. In contrast, other studies characterized the Ac/N-End rule pathway, suggest that Nt acetylation acts as a degradation signal to target Nt acetylated proteins to be labeled with ubiquitin by ubiquitin-ligase Doa10, which commits the ubiquitin-labeled proteins for proteasomal degradation. In *Sacchromyces cerevisiae*, some proteins that have P2'-residues of Ala, Val, Ser, Thr, and Cys can be Nt acetylated and committed to the Ac/N-End rule degradation pathway, while proteins acetylated in the initiator methionine can either be committed to the Ac/N-end rule pathway or, in mammal cells, the acetyl-iMet residue can be cleaved from the protein and the P2'-residue is arginylated (Karakozva science). A large caveat with these hypotheses is that only specific

protein examples support these systems, begging the question of whether the Ac/N-End rule pathway is relevant to most proteins or to just the few that have been characterized by this system. Two recent studies suggest that Nt acetylation neither negatively or positively affects protein stability. A second caveat to the Ac/N-end rule is that it has been fully characterized in yeast, but has not been expanded to other eukaryotic systems.

2.5 OTHER N-TERMINAL MODIFICATIONS TO CONSIDER

Formylation, deformylation, and initiator methionine cleavage

As mentioned above, Nt modification occurs in all domains of life and is one of the most common protein modifications, especially in eukaryotic cells. In eukaryotes, mRNA translation in the cytosol uses free methionyl-tRNA (Met-tRNA) to begin translation, whilst in mitochondria, eukaryotic plastids, and bacteria N-formylated methionyl-tRNA (fMet-tRNA) is used to initiate translation. In these organelles and cells, the methionyl-$tRNA_f^{Met}$ formyltransferase esterifies methionine to the 3' end of the initiator tRNA [8, 9]. The N-formyl group on iMet can be retained after translation or can be removed rapidly after translation begins by the essential enzyme peptide deformylase (PDF) [10-13]. After deformylation of iMet, the iMet residue usually, but not always, is cleaved from mature proteins through catalysis by methionine aminopeptidase (MAP), via a process called N-terminal methionine excision or NME, which occurs in eukaryotes and prokaryotes and is thought to process 55-70% of the all proteins depending on the organism [14].The ability of the PDF enzyme to deformylate iMet is not influenced by the

nature of amino acid residues occurring after iMet [15]. In contrast, substrate preference of the MAP enzyme seems influenced by the size and nature of the penultimate residue, with a preference for residues containing small side chains, such as Gly, Ala, Ser, Thr, Cys, Pro, or Val [16-18]. Notably, the MAP enzyme cannot excise iMet that retains its formyl group [12], and it has decreased catalytic activity *in vitro* on methionine residues that have been oxidized to methionine sulfoxide [16]. Most importantly, the physiological function of NME is controversial, with some groups arguing that NME provides protein stability, while others favor the idea of NME triggering protein degradation by providing substrates for the N-End Rule (NER), which is a protein degradation process that uses specific N-terminal amino acid residues as signals called N-degrons [19-21]. This hypothesis however, is less applicable to bacteria where the NER system differs considerably as compared to eukaryotes (For a review on N-end rule refer to reference Varshavsky 2011.

In the case of Rim acetyltransferases, NME is required for the acetylation of Ser or Ala if they are the second residue (P2) of the protein, as discussed under section *Rim acetyltransferases*.

A ground-breaking study for N-terminal proteomes was conducted by Bonissone *et al* [22]. This group of investigators analyzed proteomics data that included mass spectrometric data and a comparative genomics approach of yeast, mammals and 112 million spectra from 57 bacterial species. In this study, the authors challenged assumptions in the field surrounding the physiological role and specificity of the first three amino acids of proteins and their involvement in NME by comparing

datasets across all domains of life. Methionine aminopeptidases (MAPs) are known to remove the initiator methionine from proteins with a set of seven specific amino acid residues (*i.e.,* Gly, Ala, Ser, Pro, Val, Thr, Cys) in the P2 position to be the new N-terminus of a protein after cleavage. Bonissone *et al* found elevated levels of conservation of only two of the seven MAP-targeted position 2 (P2) residues, Ser or Ala, in a diverse array of prokaryotic and eukaryotic species, suggesting an evolutionary pressure for Ser or Ala residues to be conserved in the P2 position. This finding contrasts with the previous observation by Arfin and Bradshaw [23], who first observed that the specific target residues of NME directly correlates with the exact residues that are considered “stabilizing” amino-terminal residues” (*i.e.,* Gly, Ala, Ser, Cys, Thr, Pro, and Val), which were found on proteins with longer half-lives than proteins containing terminal residues of “destabilizing” residues (*i.e.,* Glu, Gln, Asp, Asn, Ile, Leu, Phe, Trp, Tyr, His, Lys, Arg) [21, 24]. Bonissone *et al* argued that their data set suggested that only Ser and Ala were important for NME and the remaining target residues were evolutionarily inconsequential for the NME system [22].

An alternative hypothesis for the function of NME is that it primes the N-termini of proteins so that other post-translational modifications can occur, such as N-terminal acetylation [25]. Recent studies have data to support a connection between Nt acetylation of proteins and NME and NER in the yeast *S. cerevisiae*, where N-terminally acetylated proteins (either on the iMet or the remaining P2 residue) are able to be targeted as N-degrons for degradation by the Doa10 E3 ligase and the Ubc6 or Ubc E2 enzymes is able to act as a novel N-recognin [25].

This system is limited to a select proteins and has yet to be characterized in higher eukaryotic cells. Further discussion on NME and N-terminomics studies are continued under N-TERMINAL ACETYLOMES IN PROKARYOTES.

2.6 RIM N-ACETYLTRANSFERASESS (NATs)

Some of the first examples of N(alpha) (also referred to as Nα) acetylation of bacterial proteins implicated ribosomal proteins modified by the Rim (**ri**bosomal **m**odification) acetyltransferases in *Escherichia coli*. In the 1970s, two-dimensional polyacrylamide gel electrophoresis was used to characterize the ribosomal proteins of more than 1500 thermosensitive *E. coli* mutant strains, which led to the discovery of mutations outside of ribosomal structural genes that altered the physical characteristics of some ribosomal proteins. Utilizing the method of mapping by the gradient of transmission [26], the mutations identified were in the genes encoding the acetyltransferases RimJ, RimI, and RimL [27-30].

The acetyltransferase RimJ acetylates the N-terminal alanine of ribosomal protein S5 in *E. coli* [30-32]. To date, the function of this acetylation event is unknown. *rimJ* overexpression was found to suppress the growth defect of an *E. coli* strain with a mutation in ribosomal protein S5 that resulted in cold sensitivity and ribosome biogenesis [33]. Not only has RimJ been implicated in the thermal stability of the ribosome, RimJ has been found to be involved in thermoregulation of the pyelonephritis-associated pili or *pap* operon, an operon that is repressed at 23 ºC [34]. In mutants that can transcribe the *pap* operon at non-permissive temperature of 23 ºC, overexpression of *rimJ* restores thermoregulation of the *pap*

operon, and prevents transcription at 23 ºC. It is suggested that RimJ controls *pap* transcription in a temperature-dependent manner by preventing phase variation to turn the genes into the “ON” state, however, the exact mechanism by which RimJ regulates *pap* is unknown [35]. RimJ has also been shown to N-terminally acetylate the recombinantly expressed proteins of thymosin alpha1 and the Z-domain of staphylococcal protein A during co-overexpression in *E. coli* [36-38], however, the lack of physiological relevance of these protein substrates may just indicate N-terminal amino acid preferences of RimJ.

The second ribosomal modification acetyltransferase is RimI, which acetylates the N-terminal alanine of the 30S ribosomal subunit protein S18, following cleavage of the initiator formyl-methionine [31, 39]. Even though protein S18 has been shown to be acetylated by RimI, the physiological impact of this acetylation event is still unknown. There is speculation that since the S18 protein is located near the E-site of the ribosome, acetylation of S18 could influence the translation initiation process by changing the structure of S18 near the mRNA landing site [40]. Vetting and co-workers performed the biochemical characterization studies of the *Salmonella typhimurium* RimI acetyltranferase in complex with acetyl-CoA, a peptide substrate, and solved the crystal structure of RimI in order to understand the reaction mechanism of RimI [41]. RimI acts on acetyl-CoA through a nucleophilic addition-elimination mechanism, where Glu103 acts as the catalytic base and accepts a proton from ribosomal protein S18. This causes a nucleophilic attack on the carbonyl group of acetyl-CoA by the N-terminal nitrogen of alanine in protein S18 which causes the formation of a tetrahedral intermediate. Lastly,

Tyr115 of RimI serves as the acid by providing a proton and causing decomposition of the tetrahedral intermediate [41].

A study in *E. coli* attempted to characterize the potential for RimI to be a Nε-lysine acetyltransferase through the construction of a strain of *E. coli* where the primary protein lysine acetyltransferase YfiQ (Pat homologue from *S. enterica*) gene was deleted from the chromosome, as well as the gene encoding Pta, which enables *E. coli* to generate acetyl-phosphate. Proteins can be non-enzymatically acetylated on lysyl residues by acetyl-phosphate [42]. The sole lysine deacetylase, CobB, was also deleted, along with the gene *acs*, encoding acetyl-CoA synthetase, which in the absence of Pat, it causes detrimental effects on cell growth. The authors call this strain the “gutted strain”, and provide the *E. coli rimI* gene *in trans*, then assay cell lysates in anti-acetyllysine antibody western blots to look for changes in acetylated protein patterns [43]. The lysates were then subjected to mass spectrometry and enriched for acetyllysine were subjected to LC-MS/MS in an attempt to identify protein targets of RimI. The authors found 11 unique proteins with acetylation in the RimI-overexpressed lysate with 14 unique lysine sites acetylated [43]. However, no further validation of acetylation of the identified proteins by RimI was conducted.

In an attempt to confirm acetylation of ribosomal protein S18 and expand the number of potential acetylation targets of RimI, the biochemical characterization of the *Mycobacterium tuberculosis* homologue of RimI (~32% sequence identity to *S. typhimurium* RimI) was performed [44]. The $RimI^{Mtb}$ enzyme was purified to homogeneity and tested for its ability to acetylate small peptide substrates that

represent the first six residues of ribosomal protein S18, as well as a small-scale tryptic peptide library (STPL) was a tested for acetylation by RimIMtb by using MS/MS, MALDI-TOFF, or the DTNB assay. A caveat of this study was that even though the STPL assay demonstrated that RimI^{Mtb} was able to acetylate peptides identical to the N-terminus of proteins whose genes were located near *rimI* in the chromosome (GroES, GroEL1, TsaD), purification of the full length and folded proteins of GroES and TsaB-TsaD were not acetylated by RimI^{Mtb} in an *in vitro* [^{14}C]-acetyl-CoA radiolabel transfer assay [44]. The authors argue that RimI^{Mtb} has more relaxed substrate specificity as compared to Rim acetyltransferases commonly being considered as highly substrate specific. However, this flexibility in N-terminal substrates is only seen using the peptide library and not with full length protein targets [44]. The authors provide *in vivo* support of protein-protein interactions between RimI^{Mtb} and genes that are located near *rimI* in the genome through Mycobacterial Protein Fragment Complementation (MPFC) assay, where successful protein-protein interactions *in vivo* allowed reconstitution of mDHFR (dihydrofolate reductase), which aids in survival of cells exposed to trimethoprim, an antibiotic that blocks the reduction of dihydrofolate. Survival of cells where co-transformants yielded trimethoprim resistance occurred in pairings with RimI^{Mtb} and GroES chaperone proteins, and RimI^{Mtb} with TsaD (a neighboring tRNA-A_{37}-t^{6}A transferase) [44].

The RimI^{Mtb} protein was further investigated using biophysical approaches to characterize variants of RimI^{Mtb} to generate a more stable protein. The most stable variant of RimI^{Mtb}, called MtRimIC21A4-153 (a variant containing a cysteine to

alanine point mutation with both N-terminal and C-terminal truncations) was compared to the full-length native $RimI^{Mtb}$ using 2D ^{1}H-^{15}N heteronuclear single quantum coherence (HCQC) NMR spectroscopy and circular dichroism (CD) and it was found that the truncations and mutation had very little effect on the enzymatic activity of this enzyme as compared to the parental protein [45].

The third known prokaryotic Rim acetyltransferase is RimL, which acetylates the N-terminal serine residue in ribosomal protein L12 in *E. coli* [46]. This acetylation event results in the formation of L12's acetylated form, thus called L7 [47]. Since the amino acid sequence of these proteins are identical other than the N-terminal acetylation of L7, this protein is referred to as ribosomal protein L7/L12. *E. coli* RimL can post-translationally acetylate L12 *in vitro* and acetylate L12 *in vivo*, when ribosomal protein L12 is in complex with the 50s ribosomal subunit [29, 39, 48]. The *Salmonella typhimurium* RimL protein can also acetylate ribosomal protein L12 [49].

The population of ribosomal proteins S5 and S18 are always found in their acetylated state, but protein L12 is unique in that the ratio of L7/L12 protein (acetylated vs unacetylated) differs depending on the rate and growth phase of the cells, suggesting a role of acetylation in modulation during the growth cycle. During growth in rich medium, the L12 (unacetylated form) is present in increasing concentrations over the L7 form, from early logarithmic to mid-logarithmic phase, 85% of L12 is unacetylated [50]. However, as cells enter stationary phase the acetylated form, L7, becomes the dominant form over the unacetylated L12 [50]. Another important feature of the aforementioned study is that the shift in increased

acetylation seen in L7/L12 requires the production of new ribosomes and not modification of already formed ribosomes containing unacetylated L12 proteins [50]. The authors suggest that the acetylated L7 protein may aid the survival of a nongrowing culture or in the adaptation to a different condition or cell phase [50]. In contrast, when cells are grown in minimal medium, L12 protein lacking N-terminal methionine is 100% acetylated, thus in L7 form. The authors data suggest that during conditions of stress, acetylation of L12 increases the stability of the ribosomal stalk complex where acetylation/L7 protein has a tighter interaction with its associated ribosomal subunit L10 [51, 52].Even with consideration of the aforementioned experiments, further experimentation is required to elucidate the physiological relevance of RimL acetylation of L12.

The studies of *in vitro* acetylation of L12 by *E. coli* RimL further investigated several factors aiming to compare the acetylation flexibility of RimL to its eukaryotic Nα-acetyltransferase counterparts. To test the specificity of RimL for L12's N-terminal amino acid sequence, the ability of the enzyme to acetylate L12 protein whose N-terminal serine was changed to alanine was conducted. First, the authors determined that RimL acetylated the $L12^{ISA}$ variant, albeit at a slower rate than the wildtype L12 protein. Second, the ability of RimL to acetylate a L12 variant that has the second residue changed from isoleucine to aspartic acid ($L12^{ISD}$) was tested, and the authors showed that RimL could acetylate the $L12^{I2D}$ variant, but to a much lesser degree than the wildtype L12 protein. The authors note this is in stark contrast to eukaryotic Nα-acetyltransferases that typically acetylate N-termini with acidic residues such as aspartic acid, in the secondary position [53]. The final

comparison noted is that in contrast to eukaryotic Nα-acetyltransferases that require accessory subunits for acetylation activity, RimL homodimers are sufficient for the acetylation reaction to occur [49].

2.7 NON-RIBOSOMAL N-TERMINALLY ACETYLATED PROTEINS

Salmonella NatA and CobB

The *Salmonella enterica* genome possesses ~26 GCN5-related acetyltranserase genes, of which the *in vivo* function of half of their gene products has been elucidated. Our group recently determined that the *S. enterica* putative acetyltransferase NatA (formerly YiaC) acetylates the N-terminus of the NAD^+-dependent CobB sirtuin deacylase [54]. In *S. enterica*, the CobB deacylase is present in two biologically active isoforms, which differ in size by a 37-amino acid N-terminal extension. NatA can acetylate the long isoform of CobB (hereafter $CobB_L$), resulting in a decrease in the enzymatic ability of $CobB_L$ to deacylate its *bona fide* protein substrate Acs^{Ac} [54]. The NatA and $CobB_L$ are the first non-ribosomal, Nα-acetyltransferase/protein substrate pair to be identified in bacteria, as well as being the first system that has the physiological function in vivo elucidated of known N-terminal acetylation events in bacteria.

N-terminally acetylated proteins without known acetyltranferases

Several other proteins have been identified as N-terminally acetylated in bacteria but do not have any correlating acetyltransferase, nor has the relevance of the

modification been elucidated. In *Escherichia coli*, the ribosomal elongation factor Tu (Ef-Tu) was found to be N-terminally acetylated on Ser after iMet cleavage [55]. The *mycobacterial* Esx-1 (ESAT-6 system 1) transporter is involved in the pathogenesis of *mycobacterial* species. Several Esx-1 protein substrates have been found to be N-terminally acetylated [56-58]. Two *mycobacterial* proteins transported by Esx-1, called EsxA (ESAT-6 which stands for early secreted antigenic target, 6 kDa) and EsxB (culture filtrate protein, 10 kDa, CFP-10) , have been shown to interact with each other in an acetylation-dependent manner, where N-terminal acetylation of EsxA decreases the binding interaction with EsxB, or possibly changes complex oligomerization [59]. Even though these proteins are directly linked to *mycobacterial* virulence, A genetic screen was conducted in *Mycobacterium marinum* to identify genes involved in Esx-1 export identified that deletion of a putative N-acetyltransferase called *MMAR_0039* caused hyper N-terminal acetylation of bacterial surface-associated EsxA. The authors note that seeing an increase in acetylation when an acetyltransferase is deleted is not expected, and proposes that alternative proteins in this pathway could be modified by *MMAR_0039* are yet to be identified [58].

A study analyzing changes in the proteome of the radiation-tolerant bacterium *Deinococcus deserti* when exposed to gamma radiation over time found not only an increase in the presence of GyrA (DNA gyrase subunit A), the enzyme responsible for introducing negative supercoiling to DNA but also a mixture of N-terminally acetylated and unacetylated populations of GyrA that had the iMet removed and were modified on the remaining Thr residue. The authors also note

upregulation of a putative GCN5-related acetyltransferase, deide_20140, but do not have direct evidence that deide_20140 is the correlating modifier of GyrA [60].

2.8 N-TERMINAL ACETYLOMES IN PROKARYOTES

Prokaryotic NTA is more prevalent than previously thought

In bacteria, the long-held paradigm of N-terminal acetylation occurring rarely and only in the instances of ribosomal protein acetylation has been challenged through the study of N-terminal protein modifications by mass spectrometry. In recent years, N-terminomics studies have begun to provide evidence that N-terminal protein modification is more prevalent in bacteria that previously realized. For example, work performed in the early 2010s provided insights into the conservation of NME motif recognition, which was discovered through blind PTM searches of 68 million bacterial mass spectra in 45 different bacterial species, obviating the fact that the prevalence of N-terminal protein acetylation was much higher than previously realized. Remarkably, 16% of the N-terminal modifications analyzed were N-terminal acetylations, and most of the N-terminal residues that were acetylated were Ser, Ala, or Thr, but not always, and Nt acetylation was not restricted as tightly to Ser or Ala as in the yeast data set [22].

Shortly after the Bonnissone *et al* paper was published, a dedicated N-terminomics study of *Pseudomonas aeruginosa* PA14 was published [61]. To enrich for N-terminal peptides, the authors utilized protein fractionation by isoelectric points (pI) and then analyzed digested protein fractions of nanoLC-MS/MS. It was found that 63% of the peptides analyzed had populations with

complete iMet cleavage, 31.5% of peptides fully retained their iMet, and 4.9% of peptides were found both with retained iMet and with iMet cleavage [61]. To select for N-terminally acetylated peptides, three strategies were incorporated into the work flow. The first strategy was to perform a direct analysis of extracted proteins after tryptic digestion. The second strategy involved enrichment of N-terminally acetylated peptides through CNBr-activated Sepharose resin [62]. The final strategy encompassed in-solution isoelectric focusing (IEF) fractionation using the Microrotofor IEF cell. The combination of the aforementioned strategies yielded 117 N-terminally acetylated peptides corresponding to proteins involved in a wide range of cellular functions, including transcription, translation, metabolism, transport, and also a large set of hypothetical proteins, showing that Nt acetylation may be integrated into multiple facets of *P. aeruginosa* cellular functioning. A set of N-terminally acetylated proteins whose iMet was cleaved had P2 residues that fit within the NME-specific criteria, such as Ser, Ala, Gly, Thr and Val, yet surprisingly peptides were found acetylated with and without iMet that largely contained Asn and to a smaller extent several other residues at P2 (*e.g.,* Ile, Ser, Thr, Gln, Gly, Lys Leu, Ala, Val). Finally, a subset of N-terminal sequences was found to have acetylation on the iMet that also possessed P2 residues with small radius of gyration such as Ser, Ala, Thr, and Gly, but were not limited to these P2 residues, with Asn, Leu, Gln, Ile and Lys being more abundant as P2 residues retaining iMet. The authors hypothesize that even though these N-termini should be substrates for NME, perhaps Nt acetylation blocked MAP catalysis of the target protein's iMet. A large set of proteins N-termini that were acetylated on the iMet

with P2 residues that are considered “destabilizing”, such as Glu, Asp, Gln, His, Ile, Lys, Leu, and Asn, which retainment of the iMet is expected because these P2 residues are not compatible with MAP, but they also are not compatible substrates for Rim acetyltransferases. However, *P. aeruginosa* possesses 50 genes annotated as potential acetyltransferases which could be alternative N-terminal acetyltransferases. The *S. enterica* protein deacylase CobB possesses an N-terminal Met-Glu sequence which is recognized and acetylated by the NatA acetyltransferase [54], lending support to the idea that GNATs with similar N-terminal substrates could exist in *P. aeruginosa*.

Bienvenut et. Al. used SILProNAQ (stable-isotope protein N-terminal acetylation quantification) methodology [63] to conduct a proteome-wide analysis of the effect on the N-terminome of *E. coli* proteins when iMet deformylation was inhibited by the PDF-specific antibiotic actinonin [64]. The authors’ data suggested that 10% of the proteins identified in their study were N-terminally acetylated, and they found over 100 distinctly acetylated N-termini. This dataset of Nt acetylation consisted of three subsets due to NME; a set of 27 proteins (3% of the 10% total acetylated protein) with acetylation on the iMet; acetylation on the P2 residue (75 proteins, 7% of the Nt acetylayion data set); and populations of protein where acetylation occurred on both the iMet and on the P2 residue after NME (two proteins, 0.2% of dataset) [64].

A large-scale quantitative *E. coli* proteomics study was conducted utilizing peptide mixtures from 22 different growth conditions [65. The proteins were digested with trypsin and either directly analyzed by shotgun LC-MS/MS or were

fractionated by isoelectric points using OFFGEL electrophoresis {Chenau, 2008 #31607]. To quantify copies of a protein per cell, stable isotopic dilution (SID) and selected reaction monitoring (SRM) mass spectrometry was utilized to quantify 41 control proteins to use as references in cellular abundance estimates as well as the number of cells was quantified by flow cytometry. The Intensity Based Absolute Quantification (iBAQ) approach was used for protein quantification [66]. The resulting mass spectra were analyzed for all types of protein modifications, but of interest to this review 31 proteins were found to be acetylated at the N-terminus [65]. In this report, the authors show Nt acetylation occurred primarily on N-terminal Ser, Ala, Met, and Thr. Strikingly, the authors also note that when comparing Nt acetylation across the 22 conditions tested, the abundance of Nt acetylation decreased as the average growth rate of *E. coli* in a condition increased, with the most Nt acetylation abundance seen in cells that had reached and maintained stationary phase for three days, and the least abundance of Nt acetylation was found in exponential phase in LB. Tothether, these data provide not only a glimpse into the diversity proteins modified by Nt acetylation but also the abundance of Nt acetylation under different conditions.

The authors further analyzed the N-terminal acetylation pattern of peptides isolated from three independent mutants of *E. coli*, each lacking either *rimJ*, *rimI*, or *rimL*. The only change found from the parent strain was in the *rimJ* mutant strain, where a significant decrease in Nt acetylation at Ser or Thr residues were seen [65], suggesting that RimJ was modifying proteins other than ribosomal protein S5 *in vivo*, which supported the studies on regulation of the *pap* operon by RimJ [35].

A proteomics study enriching for both Nα-acetylated and Nε-acetylated proteins in *Acinetobacter baumannii* strain ATCC 17978 was conducted using a similar work flow as the one used to study *P. aeruginosa* [61], where direct LS-MS/MS analysis was compared to IEF fractionated peptides and CNBr-activated Sepharose bead enrichment of N-terminal peptides. This study included an additional enrichment process that included the use of separating peptides by charge using solid cation exchange chromatography (SCX). This study identified 145 N-terminally acetylated proteins, where 59% of the proteins identified possessed Nt acetylation on the iMet, followed by Ser(7%), Ala (7%), and Val (7%). Just as in the *P. aeruginosa* N-terminome, a large subset of proteins possessing P2 residues with small radius of gyration were acetylated on the iMet when MAP cleavage should have occurred, again suggesting perhaps acetylation of the iMet blocked MAP cleavage of these peptides. Seven peptides identified were acetylated on both the iMet and the free P2 residue, where all but one of these peptides had Asn as the P2 residue. The authors also note that Asn was the most abundant residue at the P2 position.

Another validating note of their dataset was that 9 of the 145 proteins identified to be N-terminally acetylated were already identified as potentially acetylated in other bacteria, those proteins being ribosomal protein L10 [61], ribosomal protein L7/L12 [31], ribosomal protein S18 [22, 31], protein elongation factor EfTs [61], DapD [64], trigger factor [64], preprotein translocase subunit SecB [61, 64], and the beta subunit of nucleoid associated protein HU [61]. The categories of proteins found to be N-terminally acetylated are involved in metabolism, transcription and

translation, and carbon catabolism, but a few specific genes that differed were those of the biofilm and adhesion associated protein Bap (A1S_2724), starvation-induced protein CtsA, and the antibiotic efflux pump component AbeM, to name a few [61].

Other investigators performed a quantitative analysis of Nt acetylation in the pathogenic bacterium *Mycobacterium tuberculosis* strain Erdman and the opportunistic pathogen *Mycobacterium marinum.* These studies were performed by combining several N-terminal peptide enrichment protocols followed by nano-UHPLC-MS/MS [67]. These enrichment protocols included chemical acetylation of free amines using N-acetoxy-D_3-succinamide, so that physiologically relevant acetylation could be distinguished from chemical acetylation. For protein digestion, trypsin and GluC proteases were used to increase the sequence coverage of N-terminal peptides that have Arg or Lys rich N-termini. Finally, the use of both strong cation chromatography (SCX) and strong anion exchange chromatography (SAX) enabled greater enrichment of N-terminal peptides from internal peptides. The resulting study identified 211 proteins as N-terminally acetylated out of 1920 total protein N-termini identified in *M. tuberculosis*, and 347 N-terminally acetylated peptides out of 838 identified N-terminal peptides in *M. marinum*. Of the N-terminally acetylated peptides identified, the majority of peptides that were acetylated on the retained iMet had P2 residues that were Ala, followed by Lys, Gln, and Phe [67]. Of N-termini that had been acetylated on the remaining P2 residue, almost all were acetylated on Thr residues, followed by a few peptides

possessing acetyl-Ser. This finding is striking, and in stark contrast to the preferential acetylation of Ser or Ala residues seen in *E. coli* [64].

A large proportion of proteins identified as N-terminally acetylated were categorized as being functionally involved in intermediary metabolism and respiration in these organisms. The next largest category where proteins that are considered “conserved hypothetical” proteins, and the next largest subset belonged to cell wall and cell processes, information pathways, and lipid metabolism [67].

N-terminomics of the obligate photoautotrophic anaerobic green sulfur bacterium *Chlorobaculum tepidum* was conducted using the Combined Fractional Diagonal Chromatography (COFRADIC) method [68] on sub-fractionated proteins. The resulting analysis only found 4 proteins to be N-terminally acetylated, those proteins being: a C-terminal subunit of a transketolase (acc #: Q8KE86) acetylated on an N-terminal Gly, a glutathionine S-transferase (acc #: Q8KFS3) acetylated on the iMet residue with a P2 residue of Asn; a putative response regulator (acc #: Q8KAU3) acetylated on the iMet residue with a Lys at P2; and finally a putative uncharacterized protein (acc #: Q8KAL6) acetylated on the iMet residue with a P2 Asn residue [69].

METHODS AND LIMITATIONS IN STUDYING N-TERMINAL ACETYLATION

The field of Nt acetylation has several challenges and limitations. Quantitative N-terminomics studies suggest that the ratio of acetylated *vs* unacetylated N-termini is very low, creating difficulties in detection of acetylated N-termini. Another caveat

is that in contrast to lysine acetylation where anti-acetyllysine antibodies exist to enable detection of acetylated proteins, no antibodies exist for the detection and enrichment of N-terminally acetylated proteins. This is in part due to the variability of amino acid sequence of the N-terminus.

Mass spectrometry-based approaches

Mass spectrometry is the most frequently used method for the identification of novel, *in vivo* N-terminally acetylated proteins. Several methods have been developed to enrich for N-terminal peptides and exclude internal peptides, which typically are more abundant and crowd N-terminal peptide signals. The in-depth nuances of the different methods used in N-terminomics studies can be found reviewed elsewhere, but generally include i) chemical acetylation of free amines with isotope-labeled acetyl-groups, ii) protease digest, iii) fractionation of N-terminal peptides with SCX (strong cation exchange chromatography) and/or SAX (strong anion exchange chromatography), iv) liquid chromatography and mass spectrometry, v) Bio-informatics analysis. For reviews concerning each of the different methods, refer to these references: COFRADIC method [68, 70-74]; , SILProNAQ [63, 64]zhang; Cyanogen bromide-activated Sepharose [62, 75]; SCX [76, 77]; SAX [67]. N-terminal peptides that retained their iMet residues can be separated from N-terminally cleaved peptides by oxidation of methionine to methionine sulfoxide using a reagent such as hydrogen peroxide. Peptides containing methionine sulfoxide elute earlier during reverse-phase-HPLC separation than those without iMet residues [78].

In vitro assays

NBD-Cl

An affordable, quick method for the *in vitro* detection of N-terminally acetylated proteins compared to unmodified N-termini with detection in the micromolar range was developed utilizing fluorogenic derivatization with 4-chloro-7-nitrobenzofurazan (NBD-Cl) [79]. At pH 7, NBD-Cl reacts with free N-terminal amines of proteins, giving off a fluorescent signal at wavelength 535 nm. Proteins that possess a blocked N-terminal amino group (*i.e.*, by an acetyl group) do not react with NBD-Cl and no fluorescence is seen. This method is specific to N-terminal amino groups, where free amino groups on the side chain of lysyl residues do not react with NBD-Cl. A great aspect of this method is that proteins do not need to be modified for the assay to be conducted, however, the pH must be maintained near 7 for the NBD-Cl dye to function properly.

Radiolabeled Acetyl-CoA transfer assay

N-terminal acetylation and lysine acetylation of proteins can be visualized *in vitro* through the utilization of ^{3}H- or ^{14}C-radiolabeled acetyl-CoA, where a carbon or hydrogen associated with acetyl group has an isotope. Radiolabeled acetyl-CoA can be incubated with cellular lysates that contain acetyltransferase proteins and target substrates, then are resolved by SDS-PAGE, exposed to a phosphor screen, and imaged [80]. The appearance of bands in the phosphorimage that correlate with the molecular mass of target proteins in conjunction with control reactions show if acetylation has occurred. A major consideration when conducting

in vitro acetylation assays is to perform the assays near pH 7. At pH values 8 or greater, a higher percentage of lysyl residues are deprotonated and the thioester bond of acetyl-CoA is more labile to hydroxyl anions. These factors favor nonenzymatic acetylation to occur, which may cause false positives to occur since radiolabeled acetylation assays do not differentiate epsilon acetylation from alpha acetylation [3].

An alternative method that allows in vivo incorporation of radiolabeled acetyl groups using ^{3}H-sodium acetate has been developed called cell-based acetylation [81-83]. Cells of interest in culture media are incubated with ^{3}H-sodium acetate. Protein targets of interest are purified out of cell extracts and then resolved by SDS-PAGE and imaged using a phosphorimager. A major caveat with this assay is nonenzymatic acetylation via acetyl-phosphate and acetyl-CoA. It has been shown that in stationary phase, lysyl residues of proteins are used as “sinks” for storage of excess acetyl groups, owing to the value that acetyl-phosphate and acetyl-CoA hold as high energy molecules in the cell. Nonenzymatic incorporation of radiolabeled acetyl groups may occur to proteins and result in false positives [3, 4, 42].

Ellman’s Reagent assay

The reagent DTNB (5,5-dithio-bis-(2-nitrobenzoic acid) is commonly used for the detection and quantification of free sulfhydryl groups, and can be used to detect the free thiol generated from CoA after an acetyltransferase adds the acetyl-group from acetyl-CoA to an amino group of a protein [84, 85]. The reaction of DTNB with

the free thiols can be followed spectrophotometrically at 412 nm [86]. This in vitro assay is suitable for both epsilon and alpha-acetylation, since free CoA is the product for both reactions, and allows for the determination of kinetic parameters of enzymes. This method is, however, better suited for purified proteins and is sensitive to reducing agents that may be present in protein preps.

HPLC assays and N-terminal peptide libraries

Reverse-phase high-performance liquid chromatography (HPLC) can be used to quantify and kinetically (with radiolabeled peptides) define *in vitro* acetylation products of CoA and N-acetylated protein substrates. Alternatively, acetylated peptides, being more hydrophobic, can be separated from unacetylated counterparts. Synthetic peptide libraries have been used in order to elucidate N-terminal amino acid sequence preferences, but this method of screening peptide libraries is a more bias approach and does not consider any interactions that occur between acetyltransferases and target proteins that involves secondary or tertiary structural arrangements.

A less bias method using peptide libraries was developed that utilizes proteome-derived peptides [87]. Briefly, total proteins are isolated, alkylated, chemically acetylated, then digested with trypsin. Since lysyl residues were chemically acetylated, trypsin only cleaves C-terminal to arginyl residues. N-terminally acetylated peptides are then separated from unacetylated N-terminal peptides with by low pH using SCX [71], and peptides that contain free N-terminal amines are kept and used as substrate for purified acetyltransferase enzyme and radiolabeled

acetyl-CoA. Acetylated peptides can then be separated from unacetylated peptides by SCX, then and identified by LC-MS/MS [73, 87].

Chloroacetyl-CoA Assay

The use of the acyl-CoA derivative chloroacetyl-CoA was used in several in vitro experimental designs to identify protein substrates [79]. Chloroacetyl-CoA was synthesized by combining acidic free CoA and chloroacetic anhydride, then purified by HPLC. The synthesized chloroacetyl-CoA was used as a substrate for several acetyltransferases, including the eukaryotic Hat1 acetyltransferase that acetylates histone H4, as well as the *S. enterica* RimL acetyltransferase that acetylates ribosomal protein L12. an in vitro acetylation assay was developed where target protein(s) were incubated with chloroacetyl-CoA and the purified acetyltransferase protein. The addition of cysteamine-TAMRA (peptide bond formation generated *in situ* by combining the fluorophore TAMRA [5(6)-carboxytetramethylrhodamine] with DIPEA [N,N-diisopropylethylamine] and the peptide-coupling agents HOBt/PyBOP) where cysteamine-TAMRA's thiol eliminates the chloride of the chloroacetyl moiety on the N-terminal amine of the target protein, thus generating a TAMRA-labeled fluorescent protein. The modified protein can then be detected by fluorescent imaging of proteins resolved by SDS-PAGE. To isolate acetylated targets of an acetyltransferase, a fluorescently labeled peptide called Fluorescein-His8GGC was generated, that acted in a similar manner to the cyseamine-TAMRA to capture the chloroacetylated protein, but with the ability to isolate the chloroacetylated protein from cell extracts utilizing NTA

chromatography and visualization with the fluorescein fluorophore [79]. This method does not differentiate epsilon acetylation from alpha acetylation, as well as preparation of substrates may be technically challenging and laborious for some labs. However, like anti-acetyllysine antibodies, it can allow for isolation of acetylated proteins, whether it be on N-termini or lysyl residues.

2.10 CONCLUDING REMARKS AND OUTSTANDING QUESTIONS

The field of prokaryotic N-terminal acetylation has just begun to make advancements. There are many outstanding questions and challenges to address in this field of study. Due to the low sequence homology of GCN5-acetyltransferases, identifying functional homologues of acetyltransferases based on sequence alignments is difficult. Another difficulty in the field is the ability to discern lysine acetylation of protein targets from N-terminal acetylation. Many times non-enzymatic acetylation of lysyl residues can lead to “red herring” identifications, which requires through due-diligence on the researchers part to ensure the site of enzymatic acetylation of a protein. Obviously, the foremost outstanding question is the physiological reason for N-terminal acetylation, however, other outstanding questions in the field are as follows:

1. Why are different N-terminal populations seen?

From analysis of N-terminomics data that exist already, it seems that different populations of N-termini of the same protein can exist in the cell, ranging from iMet excision from part of a population but not all, to full iMet excision, to no iMet

excision. What is the purpose of having different residues on the N-terminus of a protein, as well as what is the purpose of chemical modifications of the N-termini? Why are some N-termini acetylated on the iMet, while others require NME-mediated iMet excision to be acetylated?

2. Why are there sequences recognized by MAP that should be cleaved by NME, yet are not, while other sequences that should not be cleaved by NME, but are?

Current N-terminomics data show populations of proteins with iMet cleavage that are for proteins that do not fit within the canonical set of residues in the P2 position of the peptide chain, such as Gly, Ala, Ser, Pro, Val, Thr, and Cys. This begs the question of whether MAP functions differently *in vivo*, or alternative iMet excision enzyme-encoding genes exist within the genome that are active against alternative P2 residues. It has also been demonstrated that lack of cleavage of iMet by NME of proteins that DO possess optimal P2 residue sequences also occurs *in vivo*. It is thought that the NME process occurs shortly after a protein is synthesized and deformylated, which begs the question as to why those proteins that are optimal for MAP-mediated iMet cleavage retain their iMet residues. Does acetylation serve a role in preventing iMet cleavage? Can iMet cleavage occur on older proteins?

3. What role does methionine oxidation/reduction play in NME and NTA, and can N-terminal acetyltransferases acetylate oxidized methionine? Free methionines and methionines within a peptide chain are able to be oxidized to

methionine sulfoxide by reactive oxygen species (ROS). Methionine sulfoxide can be reduced back to methionine through the function of methionine sulfoxide reductases as discussed earlier. It is known that MAP activity is decreased for methionine sulfoxide *in vitro*. What is the physiological hinderance of N-terminal methionine sulfoxide on MAP activity? Another question around N-terminal Met-SO is can N-terminal acetyltransferases acetylate N-terminal Met-SO?

4. Is N-terminal acetylation co-translational or post-translational?

In eukaryotes, most N-terminal acetyltransferases have been shown to interact at the ribosome in order to acetylate the N-terminus of newly synthesized polypeptides. Most N-terminal acetyltransferases even possess an auxiliary subunit that enables ribosomal attachment. Are prokaryotic N-terminal acetyltransferases also interacting with the ribosome and performing N-terminal acetylation at the ribosome, just after deformylation, or is N-terminal acetylation occurring post-translationally? How does the interplay between deformylation and N-terminal acetylation occur?

It is clear there is a dearth of knowledge in the field of prokaryotic N-terminal protein acetylation, and that new advancements in techniques and understanding the caveats of studying acetylation are needed to truly understand the complex system that is occurring *in vivo*.

2.11 REFERENCES

1. Allfrey, V.G., R. Faulkner, and A.E. Mirsky, *Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis.* Proc. Natl. Acad. Sci. U S A, 1964. **51**: p. 786-794.
2. Starai, V.J., et al., *Sir2-dependent activation of acetyl-CoA synthetase by deacetylation of active lysine.* Science, 2002. **298**: p. 2390-2392.
3. VanDrisse, C.M. and J.C. Escalante-Semerena, *Protein Acetylation in Bacteria.* Annu. Rev. Microbiol., 2019. **73**: p. 111-132.
4. Hentchel, K.L. and J.C. Escalante-Semerena, *Acylation of biomolecules in prokaryotes: a widespread strategy for the control of biological function and metabolic stress.* Microbiol. Mol. Biol. Rev., 2015. **79**: p. 321-346.
5. Christensen, D.G., et al., *Post-translational protein acetylation: An elegant mechanism for bacteria to dynamically regulate metabolic functions.* Front. Microbiol., 2019. **10**: p. 1604.
6. Shaw, W.V., *Enzymatic chloramphenicol acetylation and R factor induced antibiotic resistance in Enterobacteriaceae.* Antimicrob Agents Chemother (Bethesda), 1966. **6**: p. 221-6.
7. Burckhardt, R.M. and J.C. Escalante-Semerena, *Small-molecule acetylation by GCN5-related N-acetyltransferases in bacteria.* Microbiol. Mol. Biol. Rev., 2020. **84**: p. e00090-19.
8. Clark, B.F. and K.A. Marcker, *The role of N-formyl-methionyl-sRNA in protein biosynthesis.* J. Mol. Biol., 1966. **17**: p. 394-406.

9. Marcker, K. and F. Sanger, *N-Formyl-Methionyl-S-Rna.* J Mol Biol, 1964. **8**: p. 835-840.

10. Pine, M.J., *Kinetics of maturation of the amino termini of the cell proteins of Escherichia coli.* Biochim. Biophys. Acta, 1969. **174**: p. 359-372.

11. Meinnel, T. and S. Blanquet, *Characterization of the Thermus thermophilus locus encoding peptide deformylase and methionyl-tRNA(fMet) formyltransferase.* J Bacteriol, 1994. **176**(23): p. 7387-90.

12. Solbiati, J., et al., *Processing of the N termini of nascent polypeptide chains requires deformylation prior to methionine removal.* J. Mol. Biol., 1999. **290**: p. 607-614.

13. Mazel, D., S. Pochet, and P. Marliere, *Genetic characterization of polypeptide deformylase, a distinctive enzyme of eubacterial translation.* EMBO J., 1994. **13**: p. 914-923.

14. Meinnel, T. and C. Giglione, *Tools for analyzing and predicting N-terminal protein modifications.* Proteomics, 2008. **8**: p. 626-649.

15. Ragusa, S., et al., *Substrate recognition and selectivity of peptide deformylase. Similarities and differences with metzincins and thermolysin.* J. Mol. Biol., 1999. **289**: p. 1445-1457.

16. Frottin, F., et al., *The proteomics of N-terminal methionine cleavage.* Mol. Cell. Proteomics, 2006. **5**: p. 2336-2349.

17. Sherman, F., J.W. Stewart, and S. Tsunasawa, *Methionine or not methionine at the beginning of a protein.* Bioessays, 1985. **3**: p. 27-31.

18. Huang, S., et al., *Specificity of cotranslational amino-terminal processing of proteins in yeast.* Biochemistry, 1987. **26**: p. 8242-8246.

19. Nguyen, K.T., et al., *N-terminal methionine excision of proteins creates tertiary destabilizing N-degrons of the Arg/N-end rule pathway.* J. Biol. Chem., 2019. **294**: p. 4464-4476.

20. Tobias, J.W., et al., *The N-end rule in bacteria.* Science, 1991. **254**: p. 1374-1377.

21. Bachmair, A., D. Finley, and A. Varshavsky, *In vivo half-life of a protein is a function of its amino-terminal residue.* Science, 1986. **234**: p. 179-186.

22. Bonissone, S., et al., *N-terminal protein processing: a comparative proteogenomic analysis.* Mol. Cell. Proteomics, 2013. **12**: p. 14-28.

23. Arfin, S.M. and R.A. Bradshaw, *Cotranslational processing and protein turnover in eukaryotic cells.* Biochemistry, 1988. **27**: p. 7979-7984.

24. Varshavsky, A., *The N-end rule: functions, mysteries, uses.* Proc. Natl. Acad. Sci. U S A, 1996. **93**: p. 12142-12149.

25. Hwang, C.S., A. Shemorry, and A. Varshavsky, *N-terminal acetylation of cellular proteins creates specific degradation signals.* Science, 2010. **327**: p. 973-977.

26. De Haan, P.G., et al., *Recombination in Escherichia coli. 3. Mapping by the gradient of transmission.* Mutat. Res., 1969. **8**: p. 505-512.

27. Isono, S. and K. Isono, *Mutations affecting the structural genes and the genes coding for modifying enzymes for ribosomal proteins in Escherichia coli.* Mol. Gen. Genet., 1978. **165**: p. 15-20.

28. Isono, K. and S. Isono, *Ribosomal protein modification in Escherichia coli. II. Studies of a mutant lacking the N-terminal acetylation of protein S18.* Mol. Gen. Genet., 1980. **177**: p. 645-651.

29. Isono, S. and K. Isono, *Ribosomal protein modification in Escherichia coli. III. Studies of mutants lacking an acetylase activity specific for protein L12.* Mol. Gen. Genet., 1981. **183**: p. 473-477.

30. Cumberlidge, A.G. and K. Isono, *Ribosomal protein modification in Escherichia coli. I. A mutant lacking the N-terminal acetylation of protein S5 exhibits thermosensitivity.* J Mol Biol, 1979. **131**(2): p. 169-89.

31. Yoshikawa, A., et al., *Cloning and nucleotide sequencing of the genes riml and rimJ which encode enzymes acetylating ribosomal proteins S18 and S5 of Escherichia coli K12.* Mol. Gen. Genet., 1987. **209**: p. 481-488.

32. Janda, I., M. Kitakawa, and K. Isono, *Gene rpmF for ribosomal protein L32 and gene rimJ for a ribosomal protein acetylating enzyme are located near pyrC (23.4 min) in Escherichia coli.* Mol. Gen. Genet., 1985. **201**: p. 433-436.

33. Roy-Chaudhuri, B., et al., *Suppression of a cold-sensitive mutation in ribosomal protein S5 reveals a role for RimJ in ribosome biogenesis.* Mol Microbiol, 2008. **68**(6): p. 1547-59.

34. White-Ziegler, C.A. and D.A. Low, *Thermoregulation of the pap operon: evidence for the involvement of RimJ, the N-terminal acetylase of ribosomal protein S5.* J Bacteriol, 1992. **174**(21): p. 7003-12.

35. White-Ziegler, C.A., et al., *The N-acetyltransferase RimJ responds to environmental stimuli to repress pap fimbrial transcription in Escherichia coli.* J Bacteriol, 2002. **184**(16): p. 4334-42.

36. Bernal-Perez, L.F., et al., *RimJ-mediated context-dependent N-terminal acetylation of the recombinant Z-domain protein in Escherichia coli.* Mol Biosyst, 2012. **8**(4): p. 1128-30.

37. Fang, H., et al., *RimJ is responsible for N(alpha)-acetylation of thymosin alpha1 in Escherichia coli.* Appl Microbiol Biotechnol, 2009. **84**(1): p. 99-104.

38. Chen, J., et al., *Production of N(alpha)-acetyl Talpha1-HSA through in vitro acetylation by RimJ.* Oncotarget, 2017. **8**: p. 95247-95255.

39. Yu, M., et al., *Activity-based substrate profiling for Gcn5-related N-acetyltransferases: the use of chloroacetyl-coenzyme A to identify protein substrates.* J. Am. Chem. Soc., 2006. **128**: p. 15356-15357.

40. Nesterchuk, M.V., P.V. Sergiev, and O.A. Dontsova, *Posttranslational modifications of rfibosomal proteins in Escherichia coli.* Acta Naturae, 2011. **3**: p. 22-33.

41. Vetting, M.W., et al., *Crystal structure of RimI from Salmonella typhimurium LT2, the GNAT responsible for N(alpha)-acetylation of ribosomal protein S18.* Protein Sci, 2008. **17**(10): p. 1781-90.

42. Kuhn, M.L., et al., *Structural, kinetic and proteomic characterization of acetyl phosphate-dependent bacterial protein acetylation.* PLOS ONE, 2014. **9**: p. e94816.

43. Christensen, D.G., et al., *Identification of novel protein lysine acetyltransferases in Escherichia coli.* MBio, 2018. **9**.

44. Pathak, D., et al., *Biochemical evidence for relaxed substrate specificity of Nalpha-acetyltransferase (Rv3420c/rimI) of Mycobacterium tuberculosis.* Sci. Rep., 2016. **6**: p. 28892.

45. Hou, M., et al., *Biophysical and functional characterizations of recombinant RimI acetyltransferase from Mycobacterium tuberculosis.* Acta Biochim Biophys Sin (Shanghai), 2019. **51**(9): p. 960-968.

46. Tanaka, S., et al., *Cloning and molecular characterization of the gene rimL which encodes an enzyme acetylating ribosomal protein L12 of Escherichia coli K12.* Mol. Gen. Genet., 1989. **217**: p. 289-293.

47. Terhorst, C., et al., *The primary structure of an acidic protein from 50-S ribosomes of Escherichia coli which is involved in GTP hydrolysis dependent on elongation factors G and T.* Eur J Biochem, 1973. **34**(1): p. 138-52.

48. Miao, L., et al., *Studies of the in vitro Nalpha-acetyltransferase activities of E. coli RimL protein.* Biochem. Biophys. Res. Commun., 2007. **357**: p. 641-647.

49. Vetting, M.W., et al., *A novel dimeric structure of the RimL Nalpha-acetyltransferase from Salmonella typhimurium.* J. Biol. Chem., 2005. **280**: p. 22108-22114.

50. Ramagopal, S. and A.R. Subramanian, *Alteration in the acetylation level of ribosomal protein L12 during growth cycle of Escherichia coli.* Proc. Natl. Acad. Sci. U S A, 1974. **71**: p. 2136-2140.

51. Gordiyenko, Y., et al., *Acetylation of L12 increases interactions in the Escherichia coli ribosomal stalk complex.* J. Mol. Biol., 2008. **380**: p. 404-414.

52. Griaznova, O. and R.R. Traut, *Deletion of C-terminal residues of Escherichia coli ribosomal protein L10 causes the loss of binding of one L7/L12 dimer: ribosomes with one L7/L12 dimer are active.* Biochemistry, 2000. **39**: p. 4075-4081.

53. Polevoda, B. and F. Sherman, *N-terminal acetyltransferases and sequence requirements for N-terminal acetylation of eukaryotic proteins.* J Mol Biol, 2003. **325**: p. 595-622.

54. Parks, A.R. and J.C. Escalante-Semerena, *Modulation of the bacterial CobB sirtuin deacylase activity by N-terminal acetylation.* Proc.Natl. Acad. Sci. U S A, 2020. **117**(27): p. 15895-15901.

55. Arai, K., et al., *Primary structure of elongation factor Tu from Escherichia coli.* Proc Natl Acad Sci U S A, 1980. **77**(3): p. 1326-30.

56. Champion, P.A., et al., *ESX-1 secreted virulence factors are recognized by multiple cytosolic AAA ATPases in pathogenic mycobacteria.* Mol Microbiol, 2009. **73**(5): p. 950-62.

57. Li, Y., et al., *Capillary zone electrophoresis-electrospray ionization-tandem mass spectrometry as an alternative proteomics platform to*

ultraperformance liquid chromatography-electrospray ionization-tandem mass spectrometry for samples of intermediate complexity. Anal Chem, 2012. **84**(3): p. 1617-22.

58. Mba Medie, F., et al., *Homeostasis of N-alpha-terminal acetylation of EsxA correlates with virulence in Mycobacterium marinum.* Infect Immun, 2014.
59. Okkels, L.M., et al., *CFP10 discriminates between nonacetylated and acetylated ESAT-6 of Mycobacterium tuberculosis by differential interaction.* Proteomics, 2004. **4**: p. 2954-2960.
60. Dedieu, A., et al., *Major soluble proteome changes in Deinococcus deserti over the earliest stages following gamma-ray irradiation.* Proteome Sci, 2013. **11**(1): p. 3.
61. Ouidir, T., et al., *Characterization of N-terminal protein modifications in Pseudomonas aeruginosa PA14.* J. Proteomics, 2015. **114**: p. 214-225.
62. Zhang, X., et al., *A proteome-scale study on in vivo protein Nalpha-acetylation using an optimized method.* Proteomics, 2011. **11**(1): p. 81-93.
63. Bienvenut, W.V., C. Giglione, and T. Meinnel, *SILProNAQ: A Convenient Approach for Proteome-Wide Analysis of Protein N-Termini and N-Terminal Acetylation Quantitation.* Methods Mol. Biol., 2017. **1574**: p. 17-34.
64. Bienvenut, W.V., C. Giglione, and T. Meinnel, *Proteome-wide analysis of the amino terminal status of Escherichia coli proteins at the steady-state and upon deformylation inhibition.* Proteomics, 2015. **15**(14): p. 2503-18.
65. Schmidt, A., et al., *The quantitative and condition-dependent Escherichia coli proteome.* Nat. Biotechnol., 2016. **34**: p. 104-110.

66. Schwanhausser, B., et al., *Global quantification of mammalian gene expression control.* Nature, 2011. **473**: p. 337-42.

67. Thompson, C.R., M.M. Champion, and P.A. Champion, *Quantitative N-terminal footprinting of pathogenic mycobacteria reveals differential protein acetylation.* J. Proteome Res., 2018. **17**: p. 3246-3258.

68. Gevaert, K., et al., *Exploring proteomes and analyzing protein processing by mass spectrometric identification of sorted N-terminal peptides.* Nat. Biotechnol., 2003. **21**: p. 566-9.

69. Kouyianou, K., et al., *Proteome profiling of the green sulfur bacterium Chlorobaculum tepidum by N-terminal proteomics.* Proteomics, 2012. **12**(1): p. 63-7.

70. Staes, A., et al., *Selecting protein N-terminal peptides by combined fractional diagonal chromatography.* Nat Protoc, 2011. **6**: p. 1130-1141.

71. Staes, A., et al., *Improved recovery of proteome-informative, protein N-terminal peptides by combined fractional diagonal chromatography (COFRADIC).* Proteomics, 2008. **8**(7): p. 1362-70.

72. Ree, R., S. Varland, and T. Arnesen, *Spotlight on protein N-terminal acetylation.* Exp. Mol. Med., 2018. **50**: p. 90.

73. Van Damme, P., T. Arnesen, and K. Gevaert, *Protein alpha-N-acetylation studied by N-terminomics.* FEBS J., 2011. **278**: p. 3822-3834.

74. Van Damme, P., et al., *NatF contributes to an evolutionary shift in protein N-terminal acetylation and is important for normal chromosome segregation.* PLoS Genet., 2011. **7**: p. e1002169.

75. Zhang, X. and P. Hojrup, *Towards the N-terminal acetylome: an N-terminal acetylated peptide enrichment method using CNBr-activated sepharose resin.* Methods Mol. Biol., 2013. **981**: p. 47-56.

76. Helbig, A.O., et al., *Profiling of N-acetylated protein termini provides in-depth insights into the N-terminal nature of the proteome.* Mol. Cell. Proteomics, 2010. **9**: p. 928-939.

77. Mohammed, S. and A. Heck, Jr., *Strong cation exchange (SCX) based analytical methods for the targeted analysis of protein post-translational modifications.* Curr Opin Biotechnol, 2011. **22**(1): p. 9-16.

78. Van Damme, P., et al., *A review of COFRADIC techniques targeting protein N-terminal acetylation.* BMC Proc., 2009. **3 Suppl 6**: p. S6.

79. Bernal-Perez, L.F., L. Prokai, and Y. Ryu, *Selective N-terminal fluorescent labeling of proteins using 4-chloro-7-nitrobenzofurazan: a method to distinguish protein N-terminal acetylation.* Anal. Biochem., 2012. **428**: p. 13-15.

80. Starai, V.J. and J.C. Escalante-Semerena, *Identification of the protein acetyltransferase (Pat) enzyme that acetylates acetyl-CoA synthetase in Salmonella enterica.* J Mol Biol, 2004. **340**(5): p. 1005-12.

81. Guo, C., et al., *ATM regulates ionizing radiation-induced disruption of HDAC1:PP1:Rb complexes.* Cell Signal, 2007. **19**: p. 504-510.

82. Kraker, A.J., et al., *Modulation of histone acetylation by [4-(acetylamino)-N-(2-amino-phenyl) benzamide] in HCT-8 colon carcinoma.* Mol Cancer Ther, 2003. **2**(4): p. 401-8.

83. Verdin, E., et al., *Measurement of mammalian histone deacetylase activity.* Methods Enzymol, 2004. **377**: p. 180-96.

84. Riddles, P.W., R.L. Blakeley, and B. Zerner, *Reassessment of Ellman's reagent.* Methods Enzymol., 1983. **91**: p. 49-60.

85. Ellman, G.L., et al., *A new and rapid colorimetric determination of acetylcholinesterase activity.* Biochem. Pharmacol., 1961. **7**: p. 88-95.

86. Foyn, H., P.R. Thompson, and T. Arnesen, *DTNB-Based Quantification of In Vitro Enzymatic N-Terminal Acetyltransferase Activity.* Methods Mol Biol, 2017. **1574**: p. 9-15.

87. Van Damme, P., et al., *Proteome-derived peptide libraries allow detailed analysis of the substrate specificities of N(alpha)-acetyltransferases and point to hNaa10p as the post-translational actin N(alpha)-acetyltransferase.* Mol. Cell .Proteomics, 2011. **10**: p. M110 004580.

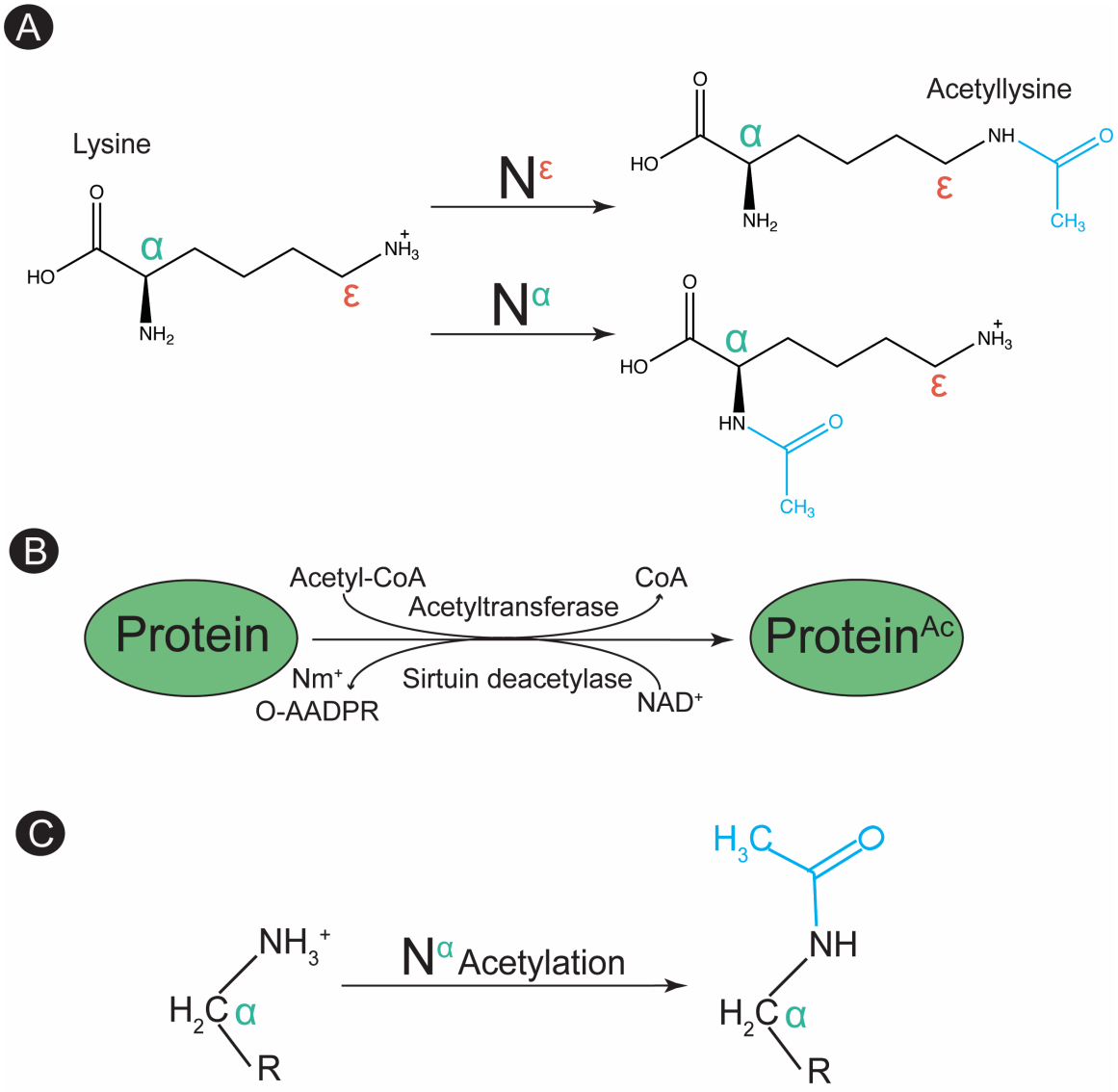

Figure 2.1. Schematic of different modes of acetylation. **A.** The amino acid lysine can be acetylated on two different amine groups; the amine associated with the alpha carbon (teal, α) or the amine associated with the epsilon carbon (orange, ε). **B**. Depiction of reversible lysine acetylation (RLA) of proteins, which involves acetylation of a protein at a lysyl residue by an acetyltransferase and deacetylation by a deacetylase. **C**. Example of alpha acetylation of an amine of a small molecule or the N-terminal amine of a peptide chain.

CHAPTER 3

A TOXIN INVOLVED IN *SALMONELLA* PERSISTENCE REGULATES ITS ACTIVITY BY ACETYLATING ITS COGNATE ANTITOXIN, A MODIFICATION REVERSED BY COBB SIRTUIN DEACETYLASE[2]

[2] Parks A.R.[a], VanDrisse C.M.[a], and Escalante-Semerena J.C. 2017. *mBio.* 8:e00708-17.

Reprinted here with permission from the publisher.

[a]Co-first authors

3.1 ABSTRACT

Bacterial toxin-antitoxin systems trigger the onset of a persister state by inhibiting essential cellular processes. The TacT toxin of *Salmonella enterica* is known to induce a persister state in macrophages through the acetylation of aminoacyl-tRNAs. Here, we show that the TacT toxin and the TacA antitoxin work as a complex that modulates TacT activity via the acetylation state of TacA. TacT acetylates TacA at residue K44, a modification that is removed by the NAD^+-dependent CobB sirtuin deacetylase. TacA acetylation increases the activity of TacT, down regulating protein synthesis. TacA acetylation altered binding to its own promoter, although this did not change *tacAT* expression levels. These claims are supported by results from *in vitro* protein synthesis experiments used to monitor TacT activity, *in vivo* growth analyses, electrophoretic mobility shift assays and qRT-PCR analysis. TacT is the first example of a Gcn-5 related *N*-acetyltransferase that modifies non-protein and protein substrates.

3.2 IMPORTANCE

During host infection, pathogenic bacteria can modulate their physiology to evade host defenses. Some pathogens use toxin-antitoxin systems to modulate a state of self-toxicity that can decrease their cellular activity, triggering the onset of a persister state. The lower metabolic activity of persister cells allows them to escape host defenses and antibiotic treatments. Hence a better understanding of the mechanisms used by pathogens to ingress and egress the persister state is of relevance to human health.

3.3 INTRODUCTION

Acetylation of proteins and small molecules is a conserved mechanism that regulates cellular processes in cells from all domains of life. Members of the Gcn5-related *N*-acetyltransferase (GNAT) protein super family (PF00583) acetylate proteins and non-protein substrates at the expense of acetyl-coenzyme A (AcCoA) [1]. In some instances, protein acetylation is reversed by class III, NAD^+-dependent deacetylases, a.k.a. sirtuins [2].

The genome of the human pathogen *Salmonella enterica* subsp. *enterica* sv Typhimurium LT2 (hereafter *S. enterica*) encodes 26 putative GNATs, three of which are annotated as toxin acetyltransferases. These putative toxins are part of type II toxin-antitoxin systems [3, 4], and the genes encoding these proteins appear to comprise separate operons for each toxin-antitoxin pair.

Toxin-antitoxin systems (TA systems) have different physiological functions, with some of them contributing to plasmid stabilization as addiction modules [5-9], while others serve as survival management systems under different stress conditions [10, 11]. TA systems have been shown to down regulate essential functions that can trigger the onset of a persister state which allows cells to survive unfavorable conditions without the need to acquire mutations [12-17]. Importantly, the slowing of essential metabolic processes by TA module expression may allow for tolerance to antimicrobials, which can lead to recalcitrant infections [18, 19].

Type II TA systems are in most cases a part of two-gene operons that encode a protein toxin and protein antitoxin [20]. Typically, the antitoxin neutralizes the toxin until a signal induces selective degradation of the antitoxin, often by stress-induced

proteases [16, 21-24]. Antitoxin degradation creates a stoichiometric imbalance that releases the toxin, thereby increasing its activity and upregulating its transcription [25, 26]. With some exceptions, antitoxins possess a DNA-binding domain that recognizes the operator site of its own promoter resulting in modulation of its own synthesis [11, 27, 28]. The *S. enterica* genome encodes three TA systems that include toxins that are homologous to Gcn5-related *N*-acetyltransferases (GNATs). All three of these systems contribute to the onset of a *S. enterica* persister state inside macrophages [29, 30]. In addition, Helaine and co-workers showed that the TacT protein of a TA type II system comprised of proteins TacT (STM3651) and TacA (STM3652) acetylates the aminoacyl moiety of several charged tRNAs arresting translation and triggering a persister state [31]. The authors suggested that exiting TacT-induced persistence was due to replenishment of tRNA pools through the hydrolysis of the acetylated-amino group off of the aminoacyl-tRNA by the peptidyl-tRNA hydrolase Pth [31].

Here we show that, in addition to acetylating aminoacyl-tRNAs, TacT acetylates its cognate TacA antitoxin with a concomitant increase in TacT activity without complex dissociation, a unique attribute not seen with other TA modules. We suggest that the acetylation state of the TacA antitoxin, not its degradation, rapidly modulates TacT activity. We also show that the NAD^+-dependent CobB sirtuin deacetylase reverses the effect that acetylated TacA has on TacT activity, and suggest that reversible TacA acetylation plays a key role in exiting the persister state. In addition, we present evidence that the TacA antitoxin binds to the *tacA-tacT* promoter when when in complex with TacT. The involvement of

sirtuin-dependent reversible lysine acetylation (sRLA) frames the persister state of *S. enterica* within the carbon and energy statuses of the cell.

3.4 RESULTS

The absence of TacA antitoxin extends the lag time before the onset of exponential growth in cells with higher levels of *tacT*$^{+}$ expression. Consistent with published data [31], ectopic expression of *tacT*$^{+}$ in *S. enterica tacA*::*cat*$^{+}$ or *tacAT*::*cat*$^{+}$ mutant strains delayed the onset of exponential growth when cells were grown on minimal medium (Fig. 3.1) due to increased TacT-mediated aminoacyl-tRNA acetylation. This delay was shortened when cells were grown in rich medium, suggesting that nutrient availability played a role in the observed phenotype. Notably, the extended lag phase did not have an effect on the final cell density or growth rate of the cultures (Fig. 3.1). The phenotype of *tacA*::*cat*$^{+}$ strains in which *tacT*$^{+}$ was overexpressed (*i.e., tacA*::*cat*$^{+}$ / pTacT^{WT} or *tacAT*::*cat*$^{+}$ / pTacT^{WT}) was corrected by *in trans* expression of *tacA*$^{+}$ (Fig. 3.1), indicating that the observed phenotype was dependent on the absence of the TacA antitoxin.

TacT acetylates residue K44 of TacA. As shown by others, TacA co-purified with TacT, forming a complex. TacA and TacT proteins were synthesized from a vector containing the coding sequence for both proteins. Overexpression of both genes allowed for simultaneous synthesis and folding of each protein, leading to a stable TacAT complex. All toxin-antitoxin complexes (wild-type and variant proteins) reported herein, were co-purified as described under Materials and Methods. Our

assessment of the purity of the proteins used in this study can be found in supplemental Fig. 3.7 (lanes 7, 8). Results of gel permeation chromatography experiments revealed the TacAT complex was a dimer of dimers (Fig. 3.8). When purified TacA and TacT proteins were incubated with [1-^{14}C]AcCoA, TacA was acetylated (Fig. 3.2A, lane 2), a modification that required TacT (Fig. 3.2A, lane 2 *vs* 3). LC/MS/MS peptide fingerprinting analysis of acetylated TacA (TacA^{Ac}) showed that residues K12, K44, and K83 were acetylated to varying degrees (Fig. 3.10). Figure S3 shows peptide masses with different *m/z* ratios used search the database using MASCOT software. Analyses of areas under the peaks for peptides corresponding to acetylated K12, K44, and K83 showed a 2.64-fold increase for K12, 18.86-fold increase for K44, and a 3.12-fold increase for K83.

Correction of the translation initiation codon of TacA. In the *Salmonella* genome, *tacAT* comprise an operon with *tacA* being promoter proximal, with the last 13 bp of *tacA* overlapping with the *tacT* coding sequence. The 5' region of *tacA* contains the annotated start methionine and an additional methionine at position 8. This information was relevant, because the above-mentioned peptide finger printing analysis did not detect the first seven amino acids (M1-L7, grey residues, Fig. 3.9), of the annotated primary sequence of TacA, therefore, we considered the possibility that the true start methionine of TacA was residue M8. This misannotation became relevant during the purification of TacAT complexes, because the nucleotides encoding the first seven amino acids and the *S*-tag fused to TacA were not translated, resulting in co-elution of tag-less, native TacA with

H_6-TacT. The availability of tagless TacA became useful for the performance of DNA binding experiments, because the DNA-binding domain of TacA was on its *N*-terminus. For the experiments mentioned above and hereto forth, the translation start codon of TacA was reassigned to the ATG encoding M8. That is, residue M8 became M1, and the numbering for the rest of the residues was modified accordingly. Consequently, residues K19, K51 and K90 became K12, K44, K83, respectively. The adjusted numbering was used throughout these studies.

Residue K44 is a key acetylation site in TacA, a modification reversed by the NAD^+-dependent CobB sirtuin deacetylase. To validate the putative acetylation sites, TacA variants with substitutions at positions K12, K44 and K83 were isolated. As shown in figure 3.2A, transfer of the acetyl moiety of [^{14}C]-AcCoA was not observed when TacT was in complex with $TacA^{K44A}$, indicating that either K44 was the only acetylation site in TacA that was modified by TacT, or that K44 acetylation triggered K12 and K83 acetylation by TacT. The latter scenario was not pursued. Instead we focused on the question of whether or not K44 acetylation was reversible. In *Salmonella,* the only known protein lysine deacetylase is the NAD^+-dependent CobB sirtuin. This result raised the question of whether CobB could deacetylate $TacA^{Ac}$. To test this idea $TacA^{[14C]Ac}$ was incubated with CobB, NAD^+, CobB + NAD^+, or CobB + NAD^+ + nicotinamide (Nm). Under the conditions tested, CobB and NAD^+ deacetylated ~70% of $TacA^{Ac}$ (Fig. 3.2B, lane 6 *vs* lanes 2, 3, 4, 5). As expected, CobB activity was inhibited by Nm (Fig. 3.2B, lane 5 *vs* 6), a result consistent with CobB dependent deacetylation.

TacA acetylation enhances TacT activity *in vitro*. TacA variants were constructed to determine whether or not acetylation of TacA residue K44 had an effect on TacT-dependent arrest of mRNA translation. TacA ariants with substitutions at position K44 that mimicked acetylation (*i.e.,* K44Q) or deacetylation (*i.e.,* K44R) [2] were isolated in complex with $TacT^{WT}$. No effect on the formation or stability of complexes between $TacT^{WT}$ and TacA variants was detected using gel filtration chromatography (Fig. 3.8). TacT was purified from homogeneous TacAT complex by denaturation and re-folding, as described in the Materials and Methods.

In vitro protein synthesis experiments were performed with $TacAT^{WT}$, $TacA^{K44Q}TacT^{WT}$ and $TacA^{K44R}TacT^{WT}$ complexes or $TacT^{WT}$ alone. Proteins were incubated with or without AcCoA in a cell-free protein synthesis system that contained all machinery necessary for transcription and translation from a synthesized DNA product. DNA that is added to the reactions codes for dihydrofolate reductase (DHFR) and its synthesis is used as a reporter of mRNA translation; DHFR synthesis was monitored using SDS-PAGE [32] (~18 kDa, arrow, Fig. 3.3). $TacT^{WT}$ toxin was expected to acetylate aminoacyl-tRNAs resulting in DHFR protein synthesis inhibition [31]. To quantify differences in DHFR synthesis in reaction mixtures containing or lacking AcCoA, intensities of three bands in each lane (asterisks in Fig. 3.3), were expressed as percentages of the band intensity of DHFR. For example, the level of DHFR synthesis was compared in reaction mixtures containing $TacAT^{WT}$ with or without AcCoA (Fig. 3.3, lanes 4,

5). Control experiments for these studies included conditions for maximal DHFR synthesis (Fig. 3.3, lane 2), with the baseline for the production of DHFR being established by a reaction mixture devoid of DNA. As expected, no DHFR was synthesized (<1%) in the absence of added DHFR encoding DNA (Fig. 3.3, lanes 2 *vs* 3). Two reaction mixtures lacking $TacA^{WT}$, with or without AcCoA added, showed $TacT^{WT}$-, AcCoA-dependent arrest of DHFR synthesis (Fig. 3.3, lanes 10 *vs* 11), indicating that the $TacT^{WT}$ protein was active.

DHFR synthesis was reduced by ~30% in reaction mixtures containing $TacAT^{WT}$ in the presence of AcCoA (Fig. 3.3, lanes 4 *vs* 5). This result was unexpected because, to our knowledge, this is the first report of a toxin suggested to be active in complex with its cognate antitoxin. Reaction mixtures containing $TacA^{K44R}T^{WT}$ (deacetylation mimic variant) and acetyl-CoA showed a 40% reduction in DHFR synthesis (Fig. 3.3, lanes 6 *vs* 7). Although it appears there are differences in aminoacyl-tRNA acetylation by the $TacAT^{WT}$ and $TacA^{K44R}T^{WT}$ proteins, these differences are not statistically significant (see percentages as standard deviations, Fig. 3.3). In sharp contrast, DHFR synthesis was reduced by 92% in reaction mixtures containing the acetylation mimic variant, $TacA^{K44Q}TacT^{WT}$ complex, and AcCoA (Fig. 3.3, lanes 8 *vs* 9), suggesting that the $TacA^{K44Q}$ variant enhanced the acetylation of aminoacyl-tRNA by $TacT^{WT}$ (Fig. 3.3, lanes 8 *vs* 9). These results strongly suggested that $TacA^{WT}$ acetylation by $TacT^{WT}$ up-regulated $TacT^{WT}$-dependent acetylation of aminoacyl-tRNAs. A summary of the above results is presented as percentages underneath SDS-PAGE gel (Fig. 3.3).

TacA acetylation enhances TacT activity *in vivo*. To validate *in vitro* results *in vivo*, *tacA* alleles encoding TacA variants with single-amino acid substitutions at position K44 were introduced by site-directed mutagenesis into a plasmid carrying *tacAT*$^{+}$ (pTacAT-1). One TacA variant within the operon had a K44Q substitution (TacA^{K44Q}) for the purpose of mimicking acetylation (pTacAT-10), another variant had a K44R (TacA^{K44R}) to mimic deacetylation (pTacAT-9). The resulting plasmids, which also carried the *tacT*$^{+}$ allele, were individually introduced into *tacAT*::*cat*$^{+}$ strains. As shown in figure 4, *tacAT*::*cat*$^{+}$ strains that synthesized TacA^{K44Q}TacT^{WT} complex had a striking growth phenotype compared to strains making TacA^{K44R}TacT^{WT} (Fig. 3.4, grey circles, diamonds). This result was consistent with an increase in aminoacyl-tRNA acetylation by TacT^{WT} that slowed down mRNA translation causing the observed phenotype. We note that the growth arrest of strains making TacA^{K44Q}TacT^{WT} was more severe than the growth delay of strains expressing *tacT*$^{+}$ alone and is most likely due to lower inductions of plasmids coding for TacA^{K44Q}TacT^{WT} causing this phenotype (*i.e.* overexpression of *tacT* alone at 25 μM arabinose is not a high enough induction to cause a growth delay). This is most likely due to stability of the toxin when it is in or not in complex with its antitoxin and these results are further discussed below.

Acetylation of residue K44 of TacA alters the DNA-binding of the protein. Electrophoretic mobility shift assays (EMSAs) were performed to determine whether modifications of K44 would affect the DNA-binding activity of TacA. A 157-bp 6-FAM 5'-labeled probe (-163 to -6, probe 1) upstream of the corrected ATG

transcription initiation codon for *tacAT* was chosen for probe design. TacAT^{WT} complex bound to this probe and this binding was specific because TacAT did not bind to the promoter for *argS*, a negative control (Fig. 3.10). To narrow down the minimal sequence necessary for TacAT.binding, two probes within the 157-bp region upstream the putative ATG codon for *tacAT* were designed. One probe was 71-bp long (-163 to -92, probe 2), and the other was 75-bp (-81 to -6, probe 3) long (Fig. 3.12, probe 2 and 3). There was no shift seen with probe 2, but the electrophoretic mobility of the 75-bp probe (probe 3) changed in the presence of the TacAT^{WT} complex (Fig. 3.11).

TacA^{WT} or TacT^{WT} alone did not bind to probe 3 (Fig. 3.12A). This result suggested that TacA needed TacT to bind to DNA. An EMSA was conducted in which TacA and TacT protein were present to test whether or not they could reform a complex capable of binding to probe 1. No binding was detected at 5-fold excess TacA and 50-fold excess TacT, nor was DNA binding observed when TacA protein was incubated with probe 1 at 50-fold excess (Fig. 3.12B). We concluded that refolded TacT was active based on the results presented in figure 3.2. However, we cannot rule out the possibility that TacA did not refold correctly, thus preventing DNA binding. It is also possible that the complex could not reform after TacA and TacT were separated.

We observed two different probe shifts (Fig. 3.5A, labeled with one- and two-site binding), which were not due to a mixed population of complex (*i.e.* acetylated *vs* non-acetylated) since a TacA^{K44R}TacT^{WT} complex (*i.e.,* not acetylatable variant, but positively charged substitution of K44) produced the same

two shifts (Fig. 3.5B). To ascertain whether the observed two bands reflected the presence of two binding sites within probe 3 the probe was split into two pieces, probe 4 and 5 (Fig. 3.12, sequence representation). Probe 4 (-81 to -46) or probe 5 (-46 to -6) were mixed with TacAT^{WT} complex, and interactions assessed by EMSAs (Fig. 3.12). The two bands observed when probe 1 and probe 3 were used (Fig. 3.5, Fig. 3.12) were not observed when probes 4 and 5 were used (Fig. 3.13), indicating that only one binding site was present in each probe. Work on the MsqRA TA system in *E. coli* showed that two band shifts occur when two binding sites were present (Brown 2013). Such a scenario is consistent with our observations (Fig. 3.5A) in that the lower band assigned to TA complex bound to either site, and the upper band reflecting interactions of complexes bound to two sites (Fig. 3.5A, cartoon representation).

In contrast to the TacA^{K44R}TacT^{WT} complex, the TacA^{K44Q}TacT^{WT} complex (*i.e.,* mimicking acetylated K44) did not display the second band, in support of the presence of two sites (Fig. 3.5). These data suggested that residue K44 did not directly affect the DNA-binding activity of TacA, but that acetylation altered TacA recognition of or binding to the *tacAT* promoter. It was further validated that K44 was not directly involved in DNA binding because the TacA^{K12A}TacT^{WT} complex had DNA binding activity that was abrogated (Fig. 3.13), suggesting the *N*-terminus included the DNA-binding domain of TacA. In contrast to this finding, binding of TacA^{K83A}TacT^{WT} and TacAT^{WT} complexes to probe 3 was very similar, suggesting that residue K83 and therefore the *C*-terminus was not involved in TacA

binding to DNA. A predicted structural representation of these residues can be seen in figure 3.14.

To determine whether the altered binding to DNA by TacA variants in complex with $TacT^{WT}$ had an effect on repression of *tacAT* transcription, chromosomal mutations coding for K44R or K44Q TacA variants were constructed as described in materials and methods. Cells coding for $TacA^{WT}TacT^{WT}$, $TacA^{K44R}TacT^{WT}$, $TacA^{K44Q}TacT^{WT}$, or cells lacking TacT were grown to mid-log on minimal medium supplemented with 22mM glycerol and total RNA was extracted as described (Stead 2012). We performed RT-qPCR with total RNA to measure the differences in level of mRNA transcript of the *tacA* gene in the mutants as compared to the $tacAT^{+}$ strain. We observed a large increase in *tacA* transcript (28.6-fold) in the strain in which the *tacT* gene was deleted and only *tacA* remained (Fig. 3.6). This suggested that the native TacAT complex repressed the operon under this condition and that the presence of TacT was needed for TacA-mediated repression. When measuring *tacA* transcript in the chromosomal variants, we did not detect significant transcript differences between the $TacA^{K44R}TacT^{WT}$ or $TacA^{K44Q}TacT^{WT}$ strains as compared to the parent strain. These results suggested that the altered DNA binding activity of $TacA^{K44Q}T$ variants seen *in vitro* might either be an artifact of the assay or that an additional factor for *tacAT* expression might be needed, and may only be induced under different conditions.

3.5 DISCUSSION

Here we report insights into the functionality of one of the three type II toxin-antitoxin systems encoded by the genome of *Salmonella enterica*, a human pathogen. The *S. enterica* TacAT system (encoded by *tacA* and *tacT*, formerly STM3651 and STM3652, respectively) is the focus of the studies reported herein. The TacAT system is the first example of a toxin-antitoxin system whose function is not regulated by dynamic association/dissociation of its components and the first example of a system where the toxin is a GNAT that recognizes a protein and non-protein substrate. In the case of TacAT, TacT acetylates the □-amino group of aminoacyl-tRNAs [29], and also acetylates a lysine residue of TacA (this work). Furthermore, we have shown that sirtuin-dependent reversible lysine acetylation (sRLA) deacetylates $TacA^{Ac}$ and may play a role in controlling the activities of TacA and TacT *in vivo*.

A new role for sRLA in *Salmonella* pathogenesis. We hypothesize that the *S. enterica* TacAT system is different than other type II toxin-antitoxin systems in that the activity of TacT is not up regulated as a result of the cleavage or cellular degradation of the cognate antitoxin (*i.e.,* TacA) (Aakre 2013; Maisonneuve 2013; Germain 2015). Our data support the conclusion that TacAT activities are modulated post-translationally through sRLA (Figs. 3, 4). The involvement of sRLA in the modulation of the entrance or egress into or out of a persister state suggests a metabolic link between the latter and the carbon (AcCoA) and energy (NAD^+) statuses of the cell (see below).

Although it is clear that acetylation of TacA alters binding of TacA to the two sites present inside the *tacAT* promoter *in vitro* (Fig. 3.5), more work is needed to understand why and how $TacA^{Ac}$ recognition of two sites is favored over recognition of either one of two sites, and whether or not there are additional factors involved in *tacAT* expression.

Other unique features of the *S. enterica* TacAT system. As stated above, a unique feature of TacT is that it can acetylate non-protein and protein substrates. Also unique is the fact that TacT activity is enhanced as a result of the acetylation of TacA by TacT while the proteins are in complex. At present, there is no indication that TacA and TacT ever dissociate from each other. In fact, the aminoacyl-tRNA acetylating activity of TacT is enhanced when TacT is in a complex with $TacA^{K44Q}$ (an acetylation mimic) (Fig. 3.3). In contrast, TacT activity decreases in complexes containing TacA or $TacA^{K44R}$ (a deacetylated mimic).

***In vivo* evidence supports the idea that TacA acetylation enhances TacT activity.** Results from *in vivo* experiments are consistent with *in vitro* results. For example, the phenotype associated with the synthesis of $TacA^{K44Q}TacT^{WT}$ occurred even at low levels of induction (25 μM arabinose), whereas the phenotype generated by high level of expression of TacT alone was erratic, and when it was observed, it occurred only at high levels of induction (150-1000 μM arabinose). These differences could be due to increased turnover rate or to poor solubility of TacT in the absence of TacA. Therefore, when TacT is in complex with $TacA^{K44Q}$,

TacT is not only stabilized, but its activity is increased leading to growth arrest (Fig. 3.4). This result suggests that $TacA^{Ac}T^{WT}$ blocks protein synthesis very efficiently. We posit that interaction of *Salmonella* with the macrophage leads to increased AcCoA levels as a result of the inhibition of AcCoA-consuming processes. Such an increase in AcCoA could trigger TacA acetylation by TacT, increased stability of the complex and increased TacT activity, ultimately resulting in the arrest of protein synthesis. These ideas await further experimentation.

Additional perspective. One central, unanswered question is how *S. enterica* monitors and modulates the levels of $TacA^{WT}$:$TacT^{WT}$ and $TacA^{Ac}$:$TacT^{WT}$ complexes. Our data suggest that the NAD^+-dependent CobB protein deacetylase prevents the accumulation of $TacA^{Ac}$:$TacT^{WT}$ complex that would drive the cell into a prolonged persister state. Given that CobB can deactylate $TacA^{Ac}$, and that NAD^+ is needed for CobB activity, a plausible answer to the above question may lie on the intracellular NAD^+ level. A decrease in the AcCoA level could be due to its consumption by anabolic processes, which would require a robust energy charge and would correlate with increased NAD^+ levels in the cell. NAD^+ would then be used by CobB to reduce TacT acitivity as a result of $TacA^{Ac}$ deacetylation.

Importance. The TacAT toxin-antitoxin system of *S. enterica* appears to have evolved to use sirtuin-dependent reversible lysine acetylation (sRLA) as a mechanism to rapidly enter and leave a persister state. There are several methods of other type II TA systems that use transcriptional regulation to either replenish

pools of antitoxin to neutralize toxin activity as with HicAB [33] or more commonly use the toxin as a co-repressor until a stoichiometric imbalance occurs to de-repress the operon (known as conditional cooperativity) [34, 35]. With the TacAT system, the use of sRLA could quickly and efficiently modulate toxin and antitoxin activities without the need for threshold de-repression of transcription. By maintaining a stable TacAT complex the cell may enter or egress the persister state as a function of AcCoA, NAD^+, and probably other as-yet unknown signals. Additional evidence for the involvement of sirtuins in the control of TacAT activities inside the macrophage is needed to advance our understanding of the persister state and how *S. enterica* gets in and out of it. Regardless, the new knowledge reported here suggests a new use of sirtuin activators to help reduce the probability of *S. enterica* maintaining a persister state inside a host.

3.6 AUTHOR CONTRIBUTIONS.

CMV, ARP designed and performed the reported experiments, analyzed the data and wrote the paper; they contributed equally to the performance of the work. JCES conceived the project, helped design experiments, analyzed data and wrote the paper.

3.7 FUNDING INFORMATION.

This work was supported by USPHS grant from the National Institutes of Health R01 GM062203 to J.C.E.-S. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

3.8 ACKNOWLEDGMENTS.

The authors have no conflict of interest to declare. We gratefully acknowledge valuable discussions with Diana Downs, Janet Westpheling, Joseph Groom, and Andrew Borchert.

3.9 MATERIALS AND METHODS

Bacterial strains. All strains constructed were derivatives of *Salmonella enterica* subsp. *Enterica* sv Typhimurium LT2 (hereafter *S. enterica*) using the Wanner in-frame gene deletion method [36]. Bacteria were grown shaking at 37 ºC and medium used for growth are described under *culture media and chemicals*.

Culture media and chemicals. All of the bacterial strains and plasmids used are listed in table S1. Strains were grown in lysogenic broth (LB, Difco), nutrient broth (NB, Difco), or no-carbon essential (NCE) minimal medium [37]. Growth studies where glycerol was the sole carbon and energy source, were performed in NCE medium supplemented with $MgSO_4$ (1 mM), Wolfe's trace minerals (1x) [38], and glycerol (22 mM). When used, antibiotics were added at the following concentrations: Ampicillin, 100 μg/ml; chloramphenicol, 20 μg/ml; and kanamycin, 50 μg/ml. Cultures used as inoculum were grown overnight at 37 ºC in NB, and small samples (1%, v/v) were used to inoculate 198 μL of fresh medium placed in each well of a 96-well microtiter plate. L-(+)-arabinose was used as inducer wherever indicated. Microtiter plates were incubated at 37 ºC inside the

temperature-controlled chamber of a Powerwave microtiter plate reader (Bio-Tek Instruments), and plates were continuously shaken using the medium setting of the instrument. Cell density was monitored at 630 nm, and data were analyzed using Prism 6 software package (GraphPad).

Strain construction. All strains constructed were *ara-9* derivatives of *S. enterica* using the Wanner in-frame gene deletion method [36]. Primers used in this study were synthesized by Integrated DNA Technologies (IDT, Coralville, IA), and are listed in table S2. *tacA*::*cat*$^{+}$ and Δ*tacA* strains were engineered as follows. Using Pfu Ultra II Fusion DNA polymerase (Stratagene), flanking regions of the plasmid pKD3 [36] were amplified with primers designed with 36-39bp of overlapping region at the beginning of either the *tacA* or *tacT* genes, and with 50 bp overlapping region at the end of the *tacA* or *tacT* genes. PCR fragments were analyzed by agarose gel electrophoresis on 1% (w/v) agarose gels post-stained with 0.5 μg/ml ethidium bromide for 15 min. PCR fragments were PCR cleaned up using Wizard SV Gel and PCR Clean-Up System (Promega), and 5-10 μl of product were electroporated into *S. enterica* strain JE10813 (Δ*ara-9*) harboring plasmid pKD46 [36]. After electroporation, cells were grown up to an optical density (630 nm) of ~0.6 at 30ºC, followed by three washes with glycerol (10%, v/v). Electroporation was performed in 0.2 cm electroporation cuvettes (MidSci) in a BioRad MicroPulser Electroporator on the Ec2 setting. Cells were then recovered in 0.5 ml of LB for 1 h at 37 ºC, plated on LB + agar + antibiotic, and incubated overnight at 37 ºC. Drug resistant transformants were streaked repeatedly on antibiotic plates

at 42ºC to cure the strains of plasmid pKD46. Strains were then reconstructed by P22-mediated transduction of the drug marker into strain JE10079. Strains containing chloramphenicol insertions were transformed with the plasmid pCP20 to resolve out the chloramphenicol insertion and make a scarred deletion.

Plasmid construction for complementation and overexpression. All plasmids used in this work are listed in table S3.3. Primers used in this study were synthesized by Integrated DNA Technologies (IDT, Coralville, IA), and are listed in table S3.2. We used the high-efficiency cloning method described elsewhere [39] to clone *tacA*, *tacT*, *tacAT* and *cobB* genes into pCV1 and pCV3 vectors. Plasmid pCV1 is a modified plasmid of pBAD24 [40] with BspQI sites added, and confers ampicillin resistance, and expression of genes cloned into it can be induced with L(+)-arabinose. Plasmid pCV3 is a modified plasmid of pBAD33 [40] with added BspQI sites and confers chloramphenicol resistance and arabinose induction. Genes *tacA, tacT*, *tacAT*, and *cobB* were amplified from the *S. enterica* chromosome using PfuUltra II Fusion DNA polymerase (Stratagene). PCR fragments were analyzed by agarose gel electrophoresis on 1% (w/v) agarose gels stained with ethidium bromide. PCR fragments were PCR cleaned up using Wizard SV Gel and PCR Clean-Up System (Promega) and digested with the restriction enzyme BspQI (NEB) at 50 ºC for 1 h; products were ligated with T4 DNA ligase (Fisher).

The overexpression vector pACYCDuet (EMD Millipore Biosciences) possesses two multiple cloning sites (MCS) for which two separate genes may be

cloned into and overexpressed simultaneously from the same vector. This vector was used for the overexpression of TacT and TacA from the same plasmid. pACYCDuet was first digested with FastDigest (ThermoFisher Scientific) BamHI and EcoRI for 1 h at 37 ºC. *tacA* and *tacT* were amplified from the *S. enterica* chromosome using PfuUltra II Fusion DNA polymerase (Stratagene). PCR fragments were analyzed by agarose gel electrophoresis on 1% (w/v) agarose gels stained with ethidium bromide. PCR fragments were PCR cleaned up using Wizard SV Gel and PCR Clean-Up System (Promega). The *tacT* PCR product was digested with FastDigest (ThermoFisher Scientific) BamHI and EcoRI for 1 h at 37 ºC. The digested PCR product and pACYCduet plasmid were PCR cleaned up with the same clean-up kit stated above, and the digested *tacT* product was ligated into MCS1 of digested pACYCDuet with T4 DNA ligase (Fisher) at room temperature for 30 min. Once transformants were verified, the *tacA* PCR product was digested with EcoRI and XhoI for 1 h at 37 ºC, PCR cleaned, and ligated as described above into the second MCS of pACYCDuet that had the *tacT* gene cloned into MSC1. Plasmids were isolated using Wizard Plus SV Miniprep Kit (Promega). To confirm the correct sequences were cloned without mutations, DNA sequencing reactions were analyzed at the Georgia Genomics Facility, UGA. Site-directed mutagenesis (Stratagene) was performed on pTacAT-2 or pTacAT-1 to change mentioned residues K12, K44, K83 to A, Q, or R. Polymerase chain reaction was performed using PfuUltra II DNA polymerase using primers listed in (Table S3.2). Modifications included an anneal time of 60 s, an extension temperature of 68 °C,

and an extension time of 2.5 min kb^{-1}. DNA changes were confirmed by sequencing.

Protein purification. Plasmids encoding H_6-TacT^{WT} and TacA^{WT} (pTacAT-2), H_6-TacT^{WT} and TacA^{K12A} (pTacAT-12), H_6-TacT^{WT} and TacA^{K44A} (pTacAT-13), H_6-TacT^{WT} and TacA^{K44Q} (pTacAT-16), H_6-TacT^{WT} and TacA^{K44R} (pTacAT-15), and His_6-TacT^{WT} and TacA^{K83A} (pTacAT-14) were electroporated into *Escherichia coli* strain C41 (λDE3) [41] Δ*pat* (strain JE9314). Cultures of cells containing plasmids were grown to stationary phase (OD_{650} ~1.3) and sub-cultured (1:100 v/v) into 6 L of LB + chloramphenicol. Cultures were grown shaking at 25 °C to an OD_{650} of 0.5, after which ectopic gene expression was induced with IPTG (0.5 mM). Cultures were grown overnight at 25°C, cells were harvested by centrifugation at 6,000 x *g* for 15 min at 4°C, and cell pellets were stored at -80°C until used.

Cell pellets were re-suspended in 50 mL of buffer A containing 4-(2-hydroxymethyl)-1-piperazineethanesulfonic acid (50 mM HEPES, pH 7.0 @ 4°C), NaCl (500 mM), imidazole (20 mM), glycerol (20 %, v/v), lysozyme (1 mg/mL), DNase (1 μg/ml), and protease inhibitor phenylmethane sulfonyl fluoride (PMSF, 1 mM). Cells were sonicated for 60 s using a Qsonica sonicator at 60% duty with 2 s pulses. Lysates were centrifuged using a Beckman Coulter Avanti J-251 centrifuge at 40,000 x g for 30 min equipped with JA-25.50 rotor. Clarified lysates were filtered through a 0.45 μm and applied to a 2-mL HisTrap FF (GE Healthcare Sciences) column using an ÄKTA FPLC system (GE Healthcare Sciences). Column was washed with 10 column volumes of bind buffer, 7 CV of 8% elution

buffer (50 mM HEPES, pH 7.0 @ 4 °C, 500 mM NaCl, 500 mM imidazole, and 20% glycerol v/v), and a 20-column volumes gradient to 100% elution buffer. When separation of TacT and TacA complexes was necessary, proteins were denatured on the column as described [31]. Fractions were run on an SDS-PAGE gel and fractions containing desired protein were combined and dialyzed for 3 h each in storage buffer 1 (50 mM HEPES, pH 7.0 @ 4 °C, 400 mM NaCl, and 20% glycerol v/v), storage buffer 2 (50 mM HEPES, pH 7.0 @ 4 °C, 200 mM NaCl, and 20% glycerol v/v), and storage buffer 3 (50 mM HEPES, pH 7.0 @ 4 °C, 150 mM NaCl, and 20% glycerol v/v). Proteins were flash frozen in liquid N_2 and stored at -80 °C. Proteins were quantified using a NanoDrop™ 1000 Spectrophotometer (Thermo Scientific) using the molecular mass (28.67 kDa) and extinction coefficient (18,700 M^{-1} cm^{-1}; ExPASy ProtParam) of the complex (assuming a 1:1 ratio of TacT to TacA, as determined via size exclusion chromatography, see below). Percent purity was calculated using ImageQuant™ v5.2 software. CobB protein was purified as described elsewhere [42]. TacT was isolated from TacAT complex using a purification protocol reported elsewhere [31]. Briefly, $TacAT^{WT}$ complex was purified as above and was separated into its components by denaturation with guanidine-HCl (5M) followed by overnight dialysis. Denatured proteins were resolved by Ni-affinity chromatography and refolded by dialysis of the denaturant. As refolding occurred, the bulk of TacT became insoluble while TacA remained stable in solution. This procedure yielded 6-fold larger amounts of TacA than TacT (*e.g.,* 0.5 mg of TacA *vs* 0.08 mg TacT per liter of culture).

Size exclusion chromatography. A Superose 12 10/300 GL gel filtration column (GE Healthcare Life Sciences) was equilibrated as per manufacturer's protocol using water and elution buffer (50 mM HEPES, pH 7.0 @ 4°C, 150 mM NaCl, 20% glycerol v/v). Samples were applied to column using a 100-μL superloop. Gel filtration standards (Bio-Rad) were applied to column first and a standard curve was calculated using the MW_{log} of each standard against retention time of each protein. Purified TacAT complexes were eluted from the column in the same manner and retention times were recorded and molecular weights were determined using equation calculated from the standard curve.

***In vitro* acetylation assays.** Homogeneous TacAT complex (3 μM) was incubated with or without [1-^{14}C]-acetyl-CoA (20 □M) in HEPES (50 mM, pH 7.5) and *tris*(2-carboxyethyl)phosphine (TCEP, 1 mM), for 1 h at 37 ºC in a total volume of 25 μl. Reactions were quenched by the addition of SDS-loading buffer (60% (v/v) glycerol, Tris-HCl pH 6.8 (0.3 M), EDTA (12 mM), 12% SDS, 2-mercaptoethanol (0.87 mM), bromophenol blue (0.05%, w/v) and reaction mixtures were resolved by SDS-PAGE on a 15% (w/v) polyacrylamide gel with Tris-HCl buffer pH 8.8 (resolving gel)/Tris-HCl pH 6.8 (stacking gel). Samples were run at 200V for 45 min. Transfer of the radiolabel onto TacA was visualized using a Typhoon Trio+ variable mode imager (GE Healthcare).

***In vitro* deacetylation assays.** To determine whether or not TacA^{Ac} was a substrate for CobB sirtuin, 200 μL reaction mixtures containing radiolabeled

$TacA^{Ac}$ synthesized as described above was treated NAD^+-dependent CobB sirtuin deacetylase [1-^{14}C]-acetyl-CoA. Briefly, excess [1-^{14}C]-acetyl-CoA was removed by buffer exchange using Amicon®Ultra – 0.5 ml centrifugal filters (Ultracell®, 10K MWCO) and HEPES buffer (50 mM, pH 7.5). Reaction mixtures were concentrated to 100 μL and served as a 2X stock of $TacA^{Ac}$. After removal of [1-^{14}C]-acetyl-CoA, $TacA^{Ac}$ was added to reaction mixtures (1X) that contained; i) NAD^+ (1 mM) + CobB (75 pmol; 3 μM final concentration), ii) NAD^+ (1 mM) + CobB (3 μM), or iii) NAD^+ (1 mM) + CobB (3 μM) + nicotinamide (5 mM). Samples were incubated at 37 ºC for 1 h, then resolved by SDS-PAGE. Deacetylation of $TacA^{Ac}$ was monitored with a phosphor imager as described above.

***In vitro* DHFR protein synthesis assay.** The PureExpress® *In Vitro* Protein Synthesis Kit (New England BioLabs) reactions were set up per manufactures protocol in RNase free tubes with the following modifications. Reactions were adjusted to 15 μL total volume (5 μL of Solution A, 3.75 μL Solution B, 100 ng DHFR DNA) and when noted, supplemented with acetyl-CoA (2 mM), TacAT complex (2 μM), or TacT (2 μM). All samples contained 0.5 μL of SUPERase In™ RNase inhibitor (ThermoFisher Scientific). Reactions were brought up to equal volumes with the addition of RNase free water, per manufactures protocol. Reaction mixtures were incubated in a 37°C sand bath for 2 h, after which tubes were placed on ice and 2.5 μL of each reaction mixture was added to 12 μL of 1x loading dye (60% (v/v) glycerol, Tris-HCl pH 6.8 (0.3 M), EDTA (12 mM), 12% SDS, 2-mercaptoethanol (0.87 mM), bromophenol blue (0.05%, w/v)) and heated

at 100 °C for 10 min. 5 μL of each denatured sample was resolved by SDS-PAGE on a 15-comb, 15% (w/v) polyacrylamide gel with Tris-HCl buffer pH 8.8 (resolving gel)/Tris-HCl pH 6.8 (stacking gel). Gels were run for 45 min at 220 V and visualized via Coomassie Blue staining and acetic acid destaining. Gels were imaged and analyzed for DHFR protein production using ImageQuant™ v5.2 software. The intensity of DHFR from each lane on the polyacrylamide gel was normalized to the asterisk-indicated bands in figure 3 using the ImageQuant™ v5.2 software. These values were used to calculate the percent decrease in DHFR per reaction. The mean percentages were plotted using Prism6 software to obtain the standard deviation of each percent decrease as shown in Table 3.1.

DNA-binding assays. Electrophoretic mobility shift DNA-binding assays were performed using DNA probes with 6-carboxyfluorescein (6-FAM) covalently attached to the 5' end of the probe. Primers used in these experiments were manufactured by Integrated DNA Technologies (IDT, Coralville, IA). Probes were generated from PCR amplification of strain JE10079 (*tacAT*$^{+}$) chromosomal DNA. PCR products were sized on a 1% (w/v) agarose gel and were purified using a Wizard SV gel and the PCR cleanup system (Promega). Binding reaction mixtures contained 6-FAM dsDNA probe (50 ng), HEPES-NaOH buffer (50mM, pH 7) containing KCl (50 mM), $MgCl_2$ (10 mM), disodium ethylenetetraacetic acid (Na_2EDTA, 0.5 mM), glycerol (10%, v/v), 25 μg / μl poly dI-Dc (deoxyinosinic-deoxycytidylic acid [(Poly(dI-dC) Sigma-Aldrich]), and when added, TacT-TacA complex protein in molar excess of probe as indicated in figure legends. Reaction

mixtures (25 μl) were incubated at 25 ºC for 45 min. Glycerol (27 μmol; 5 μl of a 50% v/v solution) was added reaction mixtures, which were resolved using a non-denaturing Criterion Tris-HCl buffer (375 mM, pH 8.6) 7.5% (w/v) polyacrylamide gel (BioRad) at 120 V. Gels were imaged using a Typhoon Trio+ variable mode imager (GE Healthcare) wavelength 488nm (Blue) and analyzed with ImageQuant v5.2 software.

Construction of a strain carrying a chromosomal *tacA* allele encoding TacA^{K44Q} or TacA^{K44R}. A region 792-bp upstream of *tacAT* in frame with the coding region of *tacAT* was cloned into pCV1 (pTacAT-17). Site-directed mutagenesis was performed to mutate *tacA* to encode either TacA^{K44R} to TacA^{K44Q}, as described above (pTacAT-22, pTacAT-23). This DNA fragment (792 bp upstream *tacAT* in frame with *tacAT* coding for variants) was designated as fragment 1 (Fig. 3.16). Gene splicing by overlap extension quantitative polymerase chain reaction (SOE-qPCR) was utilized to fuse this fragment to a *cat*$^{+}$ gene (fragment 2) and the downstream region of *tacAT* (fragment 3) The *cat*$^{+}$ gene from pKD3 [36] was used as a template for fragment 2. Fragments 1, 2, and 3 were amplified with Pfu Ultra II Fusion DNA polymerase (Stratagene) with primers listed in Table S3.2 and an annealing temperature of 61.8°C. PCR products were sized on a 1% (w/v) agarose gel and were purified using a Wizard SV PCR cleanup system (Promega). Fragments 1 and 2 (50 ng each per 50 μL PCR reaction) were annealed using SOE PCR primers (5' SOE PCR fragment 1 and 3' SOE PCR fragment 2, Table S2) with an annealing temperature of 61.8 °C and an extension time of 30 sec/kb

PCR products were sized on a 1% (w/v) agarose gel and were purified using a Wizard SV PCR cleanup system (Promega). Fragment 1-2 was annealed to fragment 3 using the same protocol. The linear PCR fragment 1-3 was transformed into a Δ*tacAT* strain harboring the helper plasmid pKD46 using protocol described above under *Strain construction* [36]. The linear PCR fragment recombined with the upstream and downstream region of *tacAT,* which inserted the *tacAT* operon coding for K44Q or K44R in place of its absence. Cells were plated on LB agar + chloramphenicol (10 μg/mL) and individual colonies were screened for acquisition of *tacAT* compared to Δ*tacAT*. Chromosomal mutations were confirmed by sequencing.

RNA isolation. Strains JE10079 (*tacAT*$^{+}$), JE23754 (*tacA1* encoding TacA^{K44Q} *tacT*$^{+}$), JE23755 (*tacA2* encoding TacA^{K44R} *tacT*$^{+}$), JE23438 (□*tacT*) were grown overnight in triplicate in nutrient broth (2 ml; NB, Difco) with shaking at 37ºC. After incubation, strains were diluted 1:100 into 5 ml of fresh NCE minimal medium supplemented with $MgSO_4$ (1 mM), Wolfe's trace minerals (1x), and glycerol (22 mM). Cultures were grown shaking at 37 ºC to an optical density (600 nm) of 0.5, then 5 ml of each sample were quickly centrifuged in 1.5 mL Eppendorf tubes at 6000 x g, supernatant was removed, and pellets were flash-frozen in liquid nitrogen and kept on dry ice. RNA was isolated following the RNAsnap^{TM} protocol [43]. Pellets were re-suspended in 150 μl of boil solution (ethylenediaminetretraacetic acid (EDTA, 18 mM), SDS (0.025%, w/v) formamide (95%, v/v; RNA grade), 2-mercaptoethanol (1% v/v) in RNase-free water) and were mixed vigorously to

break up the cell pellet. Pellets were incubated at 95 ºC for 7 min and centrifuged at 16,000 x g for 5 min at room temperature; 100 μl of supernatant was transferred to a fresh tube. A sodium acetate/ethanol RNA precipitation was then conducted by the addition of 300 μl of RNase-free water, 40 μl of sodium acetate (3 M, pH 5.2; final concentration of 0.3 M), and finally 400 μl of ice-cold absolute ethanol (100%), with mixing briefly before the addition of the next reagent. The mixture was incubated on ice for 15 min, centrifuged at 16,000 x g for 15 min at 4ºC, and ethanol was decanted off. Ethanol (400 μl of cold 70% v/v) was added and pellets were centrifuged at 12,000 x g for 10 min at 4ºC in an Eppendorf 5415D centrifuge. Ethanol was removed and pellets were allowed to dry. RNA pellets were re-suspended in RNase-free water at 4ºC on ice overnight. Subsequent RNase-free DNase I treatment was conducted using the Ambion Turbo DNA-free kit according to manufacturer's instructions (ThermoFisher Scientific). After DNA cleavage, a final sodium acetate/ethanol precipitation was performed as described above. RNA was allowed to re-suspend at 4ºC on ice for 4 h, then were flash frozen in liquid nitrogen and stored at -80ºC until used. A small aliquot of each sample was sent for quality control analysis using the RNA 600 nano kit of the Agilent 2100 bioanalyzer through the Georgia Genomics Facility. Primers for qPCR were designed using Primer 3 software and were evaluated for specificity and melting curve prior to running the qPCR.

cDNA synthesis and quantitative real time polymerase chain reaction. Total RNA (620 ng) from each sample was used for the synthesis of cDNA using the

iScript™ cDNA synthesis Kit from BioRad Laboratories according to manufacturer's protocol. Each cDNA reaction was then dilute to 7.5 ng/μl and used as template for PCR. For real-time PCR, 20 μl reactions were prepared with 10 μl of 2X FastSYBR Green master mix (Applied Biosystems), 500 nM of each gene-specific primer (1μl of 10 μM primer stock), and 15 ng of cDNA (2 μl of 7.5 ng/μl cDNA). The real-time PCR reaction was performed using a 7500 Fast real-time PCR system (Applied Biosystems). The threshold cycle values of *rpoB* and *gyrB* were checked first to ensure that both genes were optimal for use as reference genes for these strains under the conditions chosen for RT-qPCR. Cycle threshold (C_T) data were normalized to the *rpoB* gene [44]. These normalized values (ΔC_T) were transformed using $2(e\text{-}\Delta Ct)/10^{-6}$ [45], and were reported as arbitrary gene expression units (EU), or the gene expression ratio of the mutant strains/the parent strain (JE10079 *tacAT*$^+$). Mean EU values were used to calculate the standard error of the mean (SEM) using Prism6 from three biological replicates that were each tested in technical triplicate. Differences in EU between mutant strains and JE10079, and between JE23754 (*tacA1* encoding TacA^{K44Q} *tacT*$^+$), JE23755 (*tacA2* encoding TacA^{K44R} *tacT*$^+$), were compared using Welch's t-test with the Graphpad Prism6 software as shown in figure 3.6.

3.10 REFERENCES

1. Hentchel, K.L. and J.C. Escalante-Semerena, *Acylation of biomolecules in prokaryotes: a widespread strategy for the control of biological function and metabolic stress.* Microbiol. Mol. Biol. Rev., 2015. **79**: p. 321-346.

2. Bheda, P., et al., *The Substrate Specificity of Sirtuins.* Annu. Rev. Biochem., 2016. **85**: p. 405-429.

3. Rocker, A. and A. Meinhart, *Type II toxin: antitoxin systems. More than small selfish entities?* Curr. Genet., 2016. **62**: p. 287-290.

4. Chan, W.T., M. Espinosa, and C.C. Yeo, *Keeping the wolves at bay: Antitoxins of prokaryotic type II toxin-antitoxin systems.* Front. Mol. Biosci., 2016. **3**: p. 9.

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

6. Roberts, R.C., A.R. Strom, and D.R. Helinski, *The parDE operon of the broad-host-range plasmid RK2 specifies growth inhibition associated with plasmid loss.* J. Mol. Biol., 1994. **237**: p. 35-51.

7. Ogura, T. and S. Hiraga, *Partition mechanism of F plasmid: two plasmid gene-encoded products and a cis-acting region are involved in partition.* Cell, 1983. **32**: p. 351-360.

8. Bernard, P., et al., *The F plasmid CcdB protein induces efficient ATP-dependent DNA cleavage by gyrase.* J. Mol. Biol., 1993. **234**: p. 534-541.

9. Lehnherr, H., et al., *Plasmid addiction genes of bacteriophage P1: doc, which causes cell death on curing of prophage, and phd, which prevents host death when prophage is retained.* J. Mol. Biol., 1993. **233**: p. 414-428.

10. Buts, L., et al., *Toxin-antitoxin modules as bacterial metabolic stress managers.* Trends Biochem Sci, 2005. **30**: p. 672-679.

11. Page, R. and W. Peti, *Toxin-antitoxin systems in bacterial growth arrest and persistence.* Nat. Chem. Biol., 2016. **12**: p. 208-214.

12. Christensen, S.K., et al., *RelE, a global inhibitor of translation, is activated during nutritional stress.* Proc. Natl. Acad. Sci. U S A, 2001. **98**: p. 14328-14333.

13. Moyed, H.S. and K.P. Bertrand, *hipA, a newly recognized gene of Escherichia coli K-12 that affects frequency of persistence after inhibition of murein synthesis.* J. Bacteriol., 1983. **155**: p. 768-775.

14. Germain, E., et al., *Molecular mechanism of bacterial persistence by HipA.* Mol. Cell, 2013. **52**: p. 248-254.

15. Christensen, S.K., et al., *Toxin-antitoxin loci as stress-response-elements: ChpAK/MazF and ChpBK cleave translated RNAs and are counteracted by tmRNA.* J. Mol. Biol., 2003. **332**: p. 809-819.

16. Maisonneuve, E., et al., *Bacterial persistence by RNA endonucleases.* Proc. Natl. Acad. Sci. U S A, 2011. **108**: p. 13206-13211.

17. Balaban, N.Q., et al., *Bacterial persistence as a phenotypic switch.* Science, 2004. **305**: p. 1622-1625.

18. Keren, I., et al., *Specialized persister cells and the mechanism of multidrug tolerance in Escherichia coli.* J. Bacteriol., 2004. **186**: p. 8172-8180.

19. Lewis, K., *Persister cells, dormancy and infectious disease.* Nat. Rev. Microbiol., 2007. **5**: p. 48-56.

20. Hayes, F., *Toxins-antitoxins: plasmid maintenance, programmed cell death, and cell cycle arrest.* Science, 2003. **301**: p. 1496-1499.

21. Van Melderen, L., P. Bernard, and M. Couturier, *Lon-dependent proteolysis of CcdA is the key control for activation of CcdB in plasmid-free segregant bacteria.* Mol. Microbiol., 1994. **11**: p. 1151-1157.

22. Maisonneuve, E., M. Castro-Camargo, and K. Gerdes, *(p)ppGpp controls bacterial persistence by stochastic induction of toxin-antitoxin activity.* Cell, 2013. **154**: p. 1140-1150.

23. Aakre, C.D., et al., *A bacterial toxin inhibits DNA replication elongation through a direct interaction with the beta sliding clamp.* Mol. Cell, 2013. **52**: p. 617-628.

24. Germain, E., et al., *Stochastic induction of persister cells by HipA through (p)ppGpp-mediated activation of mRNA endonucleases.* Proc Natl Acad Sci U S A, 2015. **112**: p. 5171-5176.

25. Cataudella, I., et al., *Conditional cooperativity in toxin-antitoxin regulation prevents random toxin activation and promotes fast translational recovery.* Nucleic Acids Res., 2012. **40**: p. 6424-6434.

26. Gelens, L., et al., *A general model for toxin-antitoxin module dynamics can explain persister cell formation in E. coli.* PLoS. Comput. Biol., 2013. **9**: p. e1003190.

27. Bertram, R. and C.F. Schuster, *Post-transcriptional regulation of gene expression in bacterial pathogens by toxin-antitoxin systems.* Front. Cell. Infect .Microbiol., 2014. **4**: p. 6.

28. Hayes, F. and B. Kedzierska, *Regulating toxin-antitoxin expression: controlled detonation of intracellular molecular timebombs.* Toxins (Basel), 2014. **6**: p. 337-358.

29. Helaine, S., et al., *Internalization of Salmonella by macrophages induces formation of nonreplicating persisters.* Science, 2014. **343**: p. 204-208.

30. Lobato-Marquez, D., et al., *Distinct type I and type II toxin-antitoxin modules control Salmonella lifestyle inside eukaryotic cells.* Sci Rep, 2015. **5**: p. 9374.

31. Cheverton, A.M., et al., *A Salmonella Toxin Promotes Persister Formation through Acetylation of tRNA.* Mol. Cell., 2016. **63**: p. 86-96.

32. Laemmli, U.K., *Cleavage of structural proteins during the assembly of the head of bacteriophage T4.* Nature, 1970. **227**: p. 680-685.

33. Turnbull, K.J. and K. Gerdes, *HicA Toxin of Escherichia coli derepresses hicAB transcription to selectively produce HicB antitoxin.* Mol. Microbiol., 2017.

34. Turnbull, A.L. and M.G. Surette, *Cysteine biosynthesis, oxidative stress and antibiotic resistance in Salmonella typhimurium.* Res. Microbiol., 2010. **161**: p. 643-650.

35. Winther, K.S. and K. Gerdes, *Regulation of enteric vapBC transcription: induction by VapC toxin dimer-breaking.* Nucleic Acids Res., 2012. **40**: p. 4347-4357.

36. Datsenko, K.A. and B.L. Wanner, *One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products.* Proc. Natl. Acad. Sci. USA, 2000. **97**: p. 6640-6645.

37. Berkowitz, D., et al., *Procedure for identifying nonsense mutations.* J. Bacteriol., 1968. **96**: p. 215-220.

38. Balch, W.E. and R.S. Wolfe, *New approach to the cultivation of methanogenic bacteria: 2-mercaptoethanesulfonic acid (HS-CoM)-dependent growth of Methanobacterium ruminantium in a pressurized atmosphere.* Appl. Environ. Microbiol., 1976. **32**: p. 781-791.

39. VanDrisse, C.M. and J.C. Escalante-Semerena, *New high-cloning-efficiency vectors for complementation studies and recombinant protein overproduction in Escherichia coli and Salmonella enterica.* Plasmid, 2016. **86**: p. 1-6.

40. Guzman, L.M., et al., *Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter.* J Bacteriol, 1995. **177**(14): p. 4121-30.

41. Miroux, B. and J.E. Walker, *Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels.* J. Mol. Biol., 1996. **260**: p. 289-298.

42. Tucker, A.C. and J.C. Escalante-Semerena, *Biologically active isoforms of CobB sirtuin deacetylase in Salmonella enterica and Erwinia amylovora.* J. Bacteriol., 2010. **192**: p. 6200-6208.

43. Stead, M.B., et al., *RNAsnap: a rapid, quantitative and inexpensive, method for isolating total RNA from bacteria.* Nucleic Acids Res., 2012. **40**: p. e156.

44. Rocha, D.J., C.S. Santos, and L.G. Pacheco, *Bacterial reference genes for gene expression studies by RT-qPCR: survey and analysis.* Antonie Van Leeuwenhoek, 2015. **108**: p. 685-693.

45. Livak, K.J. and T.D. Schmittgen, *Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method.* Methods, 2001. **25**: p. 402-408.

Table 3.1. Strains and plasmids used in this study.		
Strain	**Relevant genotype**	**Reference/source[a]**
JE10079	*ara-9*	Laboratory strain
Derivatives of JE10079		
JE21469	*tacAT*$^{+}$ / pCV1	This work
JE21486	*tacAT*$^{+}$ / pTacT-4	This work
JE21493	*tacA::cat*$^{+}$ / pCV1	This work
JE21494	*tacA::cat*$^{+}$ / pTacT-4	This work
JE21489	*tacT::cat*$^{+}$/ pCV1	This work
JE21490	*tacT::cat*$^{+}$/ pTacT-4	This work
JE21500	*tacAtacT::cat*$^{+}$/ pCV1	This work
JE21501	*tacAtacT::cat*$^{+}$/ pTacT-4	This work
JE21503	*tacAtacT::cat*$^{+}$/ pTacAT-1	This work
JE23697	*tacAtacT::cat*$^{+}$/ pTacA^{K44R}T	This work
JE23698	*tacAtacT::cat*$^{+}$/ pTacA^{K44Q}T	This work
JE23754	*tacA1 tacT*$^{+}$	This work
JE23755	*tacA2 tacT*$^{+}$	This work
JE23438	Δ*tacT*	
JE10813	Δ*ara-9* / pKD46	
***E. coli* strains**		
E. coli C41 (λDE3)	*pka12*::*kan*$^{+}$ *ompT hsdS* (r_Bm_B) *gal λ (DE3)*	Laboratory collection
E. coli DH5α	Φ80d*lacZ* ΔM15 *recA*1 *endA*1 *gyrA*96 *thi*-1 *hsdR*17 (r_k^-, m_k^+) *supE*44 *relA*1 *deoR* Δ(*lacZYA-argF)* U169 *phoA*	NEB

Table 3.2. Primers and DNA probes used in this study.	
Primer Name	**Primer Sequence 5' → 3'**
Strain construction	
5' *tacT* Wanner primer	CACCAGAACCTTTGTCCGCTTTTCATCAGGTAGCTGGTGTAGGCTGGAGCTGCTTC
3' *tacT* Wanner primer	CGCTTATAGCGATTTGAACATAACAGTCTTTGCATTCTATTGAGGGAGCCCATATGAATATCCTCCTTAG
5' *tacAT* Wanner primer	GCTATACATGGTGGTTGTGCTATTCTTGTAAAGCAAATGGTGTAGGCTGGAGCTGCTTC
3' *tacAT* Wanner primer	CGCTTATAGCGATTTGAACATAACAGTCTTTGCATTCTATTGAGGGAGCCCATATGAATATCCTCCTTAG
5' *tacA* Wanner primer	GCTATACATGGTGGTTGTGCTATTCTTGTAAAGCAAATGGTGTAGGCTGGAGCTGCTTC
3' *tacA* Wanner primer	GGTTTCCTTGCCAGCAGTTTTTCGATAACGGGATCATCTGCGACCGGTGCCATATGAATATCCTCCTTAG
5' SOE PCR fragment 1	AATTAACAAGGTCTGTACAGGGATGCTATCAG
3' SOE PCR fragment 1	CTCCAGCCCTCCAGCCTACACCTATTGAGGGAGCCTAAGGAACAATGT
5' SOE PCR fragment 2	CTCCCTCAATAGGTGTAGGCTGGAGCTGCT
3' SOE PCR fragment 2	CCGGCTGTCAGGGCATATGAATATCCTCCTTAG
5' SOE PCR fragment 3	GATATTCATATGCCCTGACAGCCGGAAAGG
3' SOE PCR fragment 3	CGAAACGCGAAAAACCCACG
Site-Directed Mutagenesis	
5' $TacA^{K12A}$	GCCCGCTGCGATTCCGCAGCTCTAAGGTTAAGTTGAAC
3' $TacA^{K12A}$	GTTCAACTTAACCTTAGAGCTGCGGAATCGCAGCGGGC
5' $TacA^{K44A}$	TGCCAGGCTGCCGAGGCAGTGATCCTTGACCG
3' $TacA^{K44A}$	CGGTCAAGGATCACTGCCTCGGCAGCCTGGCA
5' $TacA^{K83A}$	CTGCTGGCAAGGGCGCCTCAGTGGGACGTG
3' $TacA^{K83A}$	CACGTCCCACTGAGGCGCCCTTGCCAGCAG
5' $TacA^{K44R}$	GTCAAGGATCACTCTCTCGGCAGCCTGGCAG
3' $TacA^{K44R}$	CTGCCAGGCTGCCGAGAGAGTGATCCTTGAC
5' $TacA^{K44Q}$	CGGTCAAGGATCACCTGCTCGGCAGCCTGGC
3' $TacA^{K44Q}$	GCCAGGCTGCCGAGCAGGTGATCCTTGACCG
Cloning [b]	
5' *tacT* cloned into pCV1	NNGCTCTTCNTTCGTGGGACGTGTAACAGCACC
3' *tacT* cloned into pCV1	NNGCTCTTCNTTACTATTGAGGGAGCCTAAGGA

5' *tacA* cloned into pCV1	NNGCTCTTCNTTCATGCTATACAAGGGGTGTCT
3' *tacA* cloned into pCV1	NNGCTCTTCNTTATTACACGTCCCACTGAGGTT
5' *tacT* cloned into MCS1 pACYCDuet BamHI	NNNNNNGGATCCGTGGGACGTGTAACA
3' *tacT* cloned into MCS1 pACYCDuet EcoRI	NNNNNNGAATTCCTATTGAGGGAGCCTAAGGAACAATGTTC
5' *tacA* cloned into MCS2 pACYCDuet EcoRV	NNNNNNGATATCATGCTATACAAGGGGTGTCT
3' *tacA* cloned into MCS2 pACYCDuet XhoI	NNNNNNCTCGAGTTACACGTCCCACTGAGGTT
5' Upstream *tacA*	NNGCTCTTCNTTCAATTAACAAGGTCTGTACAGGGATGCTATCAG
Sequence verification primers	
5' T7 Foward	TAATACGACTCACTATAGGG
3' T7 Reverse	GCTAGTTATTGCTCAGCGG
5' pBAD sequencing primer	CTGTTTCTCCATACCCGTT
3' pBAD sequencing primer	GGCTGAAAATCTTCTCT
5' *tacT* sequencing primer	TCAACTTAACCTTAGAGCTAAGGA
3' *tacT* sequencing primer	GCTCACTTTGTACTGACCCC
5' *tacA* sequencing primer	CACGGCTTTCATATAGTGGAGCG
3' *tacA* sequencing primer	GCAAGACGGGCAAGTATAATGACAGGG
EMSA primers	
5' Probe 1, Probe 2 EMSA primer	CCTTGTTGAACTCAGTAAACCG
3' Probe 1, Probe 3, Probe 5 EMSA primer	ACCCCTTGTATAGCATTTGC
3' Probe 2 EMSA primer	TGAATACTGATGTTCGCTT
5' Probe 3, Probe 4 EMSA primer	ACGACGACTTGATGTATAC
3' Probe 4 EMSA primer	ACCATGTATAGCAAACGG
5' Probe 5 EMSA primer	TGGTTGTGCTATTCTTGTA
RT-qPCR primers	
5' *tacA* RTqPCR	CCTTGACCGCCGTGTATTT
3' *tacA* RTqPCR	GGTTTCCTTGCCAGCAGTT
5' *rpoB* RTqPCR	AGTCGACCTGAGCACTTCA
3' *rpoB* RTqPCR	CAAACACTGGTGTGGCAATC
5' *gyrB* RTqPCR	CCGTTGGATCACGAGTTTG
3' *gyrB* RTqPCR	AACGCGTCCTCTTCAATCAG

Table 3.3. Plasmids used in this study[a]			
Plasmid	**Genotype**	**Description**	**Source**
pCV1	*araC*$^+$ *bla*$^+$	pBAD24 with BspQI MCS	[39]
pCV3	*araC*$^+$ *cat*$^+$	pBAD33-SD1 with BspQI MCS	[39]
pACYCDuet-1	*cat*$^+$	Vector for coexpression of two genes	Novagen
pTacT-4	*tacT*$^+$ *bla*$^+$	*tacT*$^+$ cloned into pCV1	This study
pTacAT-1	*tacAT*$^+$ *bla*$^+$	*tacAT*$^+$ cloned into pCV1	This study
pTacAT-2	*tacAT*$^+$ *cat*$^+$	pACYCDuet plasmid with *tacT* cloned into MCS1, *tacA* cloned into MCS2	This study
pTacAT-12	*tacAT*$^+$ *cat*$^+$	pACYCDuet plasmid with *tacT* cloned into MCS1, *tacA* coding for TacA^{K12A} cloned into MCS2	This study
pTacAT-13	*tacAT*$^+$ *cat*$^+$	pACYCDuet plasmid with *tacT* cloned into MCS1, *tacA* coding for TacA^{K44A} cloned into MCS2	This study
pTacAT-14	*tacAT*$^+$ *cat*$^+$	pACYCDuet plasmid with *tacT* cloned into MCS1, *tacA* coding for TacA^{K83A} cloned into MCS2	This study
pTacAT-15	*tacAT*$^+$ *cat*$^+$	pACYCDuet plasmid with *tacT* cloned into MCS1, *tacA* coding for TacA^{K44R} cloned into MCS2	This study
pTacAT-16	*tacAT*$^+$ *cat*$^+$	pACYCDuet plasmid with *tacT* cloned into MCS1, *tacA* coding for TacA^{K44Q} cloned into MCS2	This study
pTacAT-9	*tacAT*$^+$ *bla*$^+$	*tacAT*$^+$ cloned into pCV1, *tacA* coding for TacA^{K44R}	This study
pTacAT-10	*tacAT*$^+$ *bla*$^+$	*tacAT*$^+$ cloned into pCV1, *tacA* coding for TacA^{K44Q}	This study
pTacAT-17	P_{tacAT} *bla*$^+$	Upstream region (-792 bp) of *tacAT*$^+$ with *tacAT*$^+$ cloned into pCV1	This study
pTacAT-22	*tacAT*$^+$ *bla*$^+$	Upstream region (-792 bp) of *tacAT*$^+$ with *tacAT* (coding for TacA^{K44R}) cloned into pCV1	This study
pTacAT-23	P_{tacAT} *bla*$^+$	Upstream region (-792 bp) of *tacAT*$^+$ with *tacAT* (coding for TacA^{K44Q}) cloned into pCV1	This study
pCobB71	*cobB*$^+$ *bla*$^+$	*cobB* ORF in pTEV6	[42]
pKD46	*bla*$^+$	Express the Red system to avoid unwanted recombination	[36].
pKD3	*cat*$^+$	Template plasmid carrying *cat*$^+$ gene	[36].

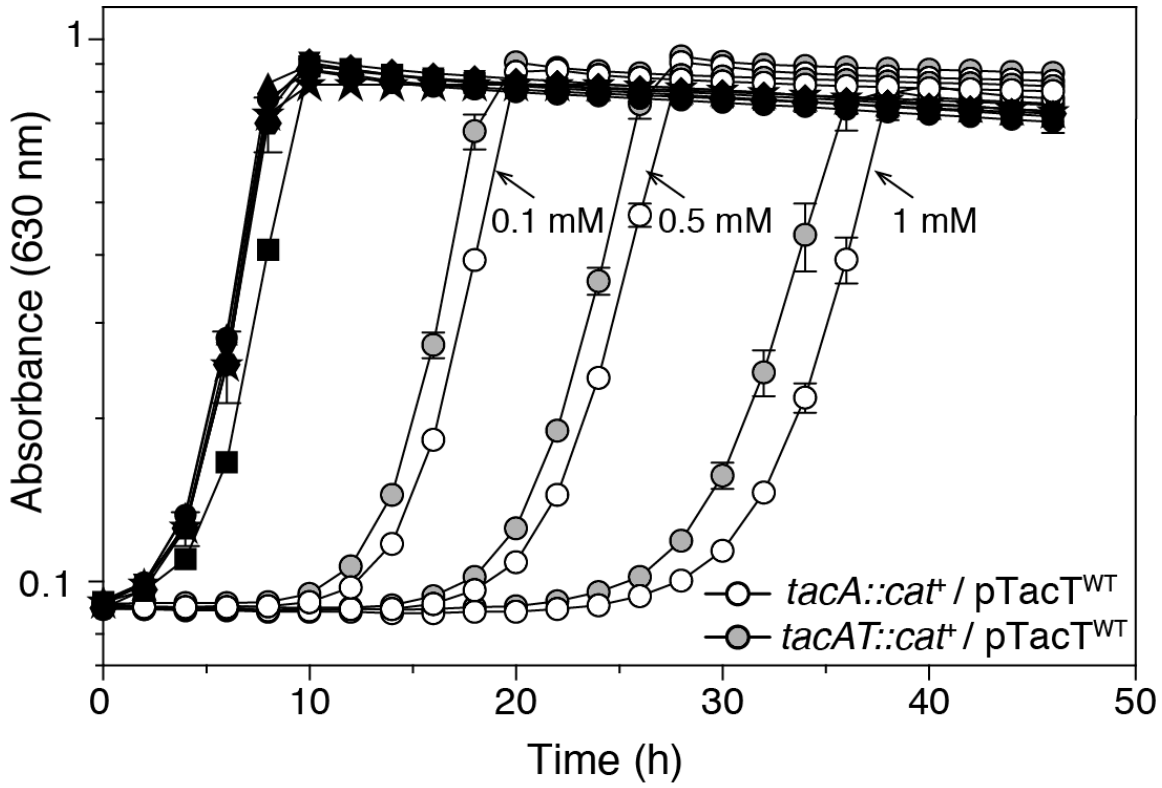

Figure 3.1. Growth of *S. enterica* cells lacking tacA and harboring tacT$^+$ plasmids have growth delays associated with increased expression of tacT. Genes coding for TacT (pTacT) or TacAT (pTacAT) were cloned into a complementation vector and introduced into strains with genetic backgrounds identified by the symbols in the graph. Overnight cultures were grown in NB rich medium. Cells were sub-cultured (1% v/v) and grown in NCE minimal medium supplemented with glycerol (22 mM), ampicillin, and transcription of plasmids containing *tacT*$^+$ were induced using levels of arabinose designated by arrows (grey and white circles). All strains with solid black symbols contained 0.15 mM arabinose. Growth curves were obtained in technical triplicates of biological triplicates and was repeated three times using a microplate reader (BioTek instruments). Error bars represent standard deviation of technical triplicates.

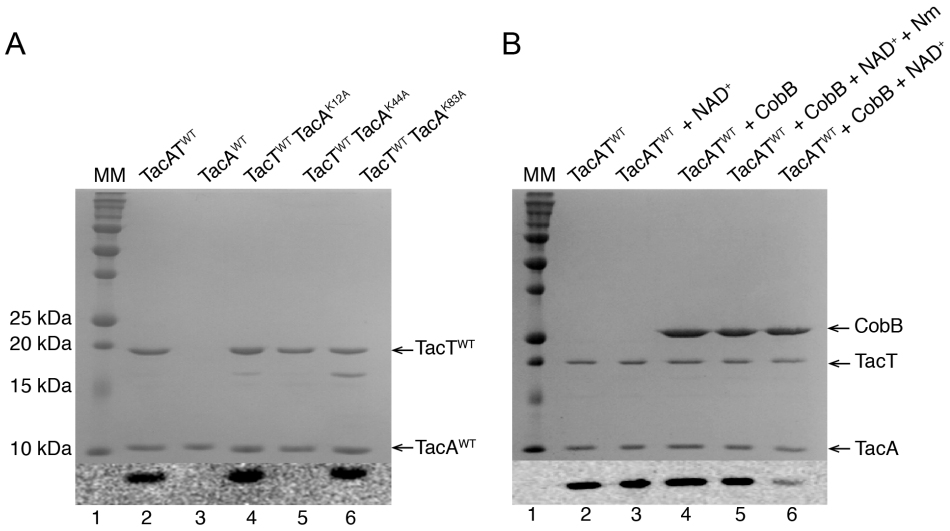

Figure 3.2. A. **TacA residue K44 is acetylated by TacT.** Wild-type TacA-TacT complex (lane 2) or variant (lanes 4, 5, 6) complexes and TacA (lane 3) were incubated with [1-^{14}C]-acetyl-CoA. Proteins were separated by SDS-PAGE and stained with Coomasie Blue to visualize proteins. Precision Plus Protein™ (Bio-Rad) standard was used as a molecular marker (MM). Acetylation was visualized by phosphor imaging. Extra bands in lanes 4 and 5 were analyzed via mass spectrometry and were identified as TacT, which we presumed was cleaved or lacked the hexahistidine tag. **B.** *TacA^{Ac} is deacetylated by CobB.* The reaction mixture used as positive control contained wild-type TacAT (TacAT^{WT}) complex and [1-^{14}C]-acetyl-CoA (lane 2); positive control mixture + NAD$^+$ (lane 3); positive control mixture + CobB (lane 4); positive control mixture + CobB + NAD$^+$ + Nm (lane 5); or positive control mixture + CobB + NAD$^+$ (lane 6). Proteins were separated by SDS-PAGE and stained with Coomasie Blue to visualize proteins. Precision Plus Protein™ (Bio-Rad) standard was used as a molecular marker (MM, lane 1). Acetylation was visualized by phosphor imaging. On panels A and B, the thin black line at the bottom of the figures separates the SDS-PAGE gel from the phosphor image. CobB: sirtuin deacetylase, NAD$^+$: nicotinamide adenine dinucleotide, Nm: nicotinamide, TacT: toxin, TacA: antitoxin.

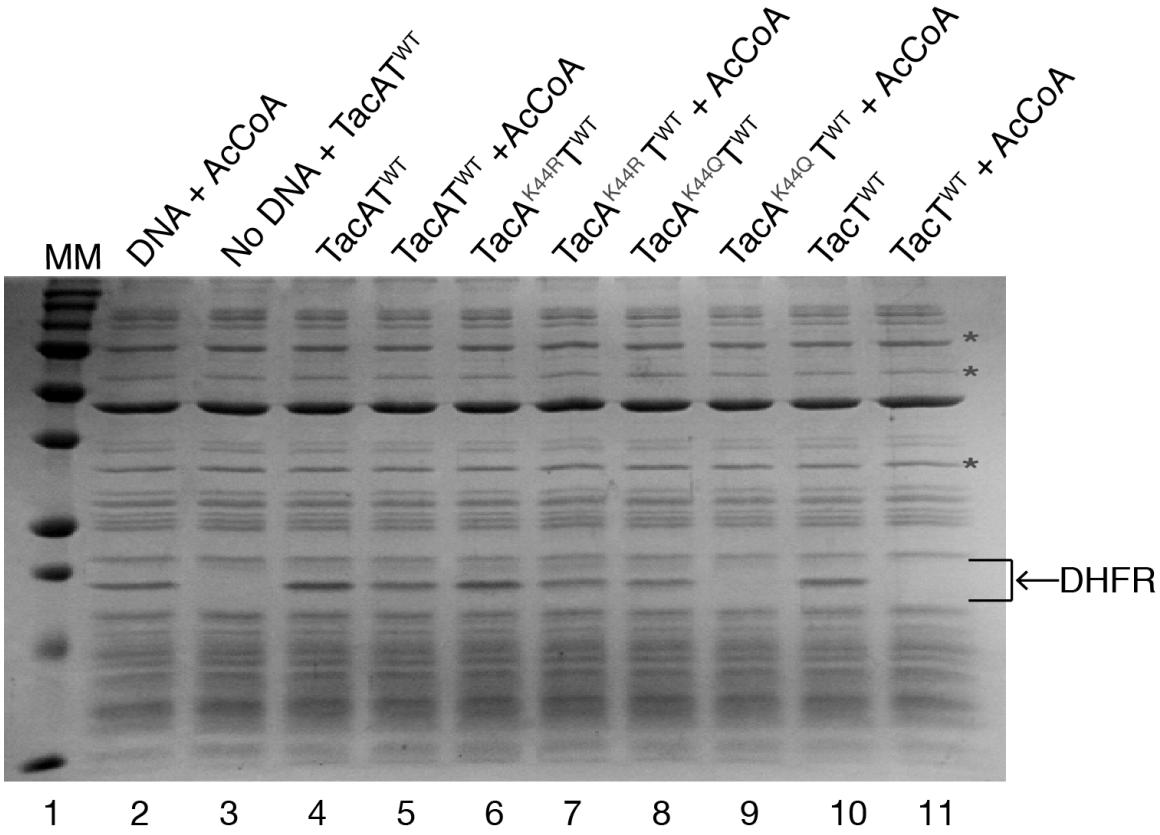

Figure 3.3. *In vitro* DHFR protein synthesis used to quantify aminoacyl-tRNA acetylation by TacT shows TacA acetylation variants increase TacT-mediated aminoacyl-tRNA aceylation. SDS-PAGE gel showing *in vitro* synthesis of dihydrofolate reductase (DHFR, indicated by arrow), in reactions mixtures containing the additions indicated above each lane. Precision Plus Protein™ (Bio-Rad) standard was used as a molecular marker (MM, lane 1). Lane 2 represents a positive control for DHFR production, where the reaction mixture contained DHFR DNA (100 ng) and acetyl-CoA to ensure acetyl-CoA stocks did not interfere with reaction components. The reaction mixture in lane 3 lacked DNA and contained TacAT^{WT} complex to control for any added bands that might interfere with interpretations. Lanes 4-9, reaction mixtures that contained TacAT^{WT} or variant complexes (indicated by labels above each lane) with or without acetyl-CoA and DNA was added to lanes 4-11. Lanes 10 and 11, positive control for aminoacyl-tRNA acetylation. These reaction mixtures contained TacT with or without acetyl-CoA. All reactions were incubated for 2 h at 37°C. Details of the procedure can be found under in materials and methods. Asterisks represent bands used to normalize the intensity of the DHFR protein within each lane (ImageQuant™ v5.2 software). Comparison of DHFR band intensity to bands with asterisks was averaged and the standard deviation of these three numbers was calculated. The percent decrease with SD for each sample is indicated below the corresponding lanes.

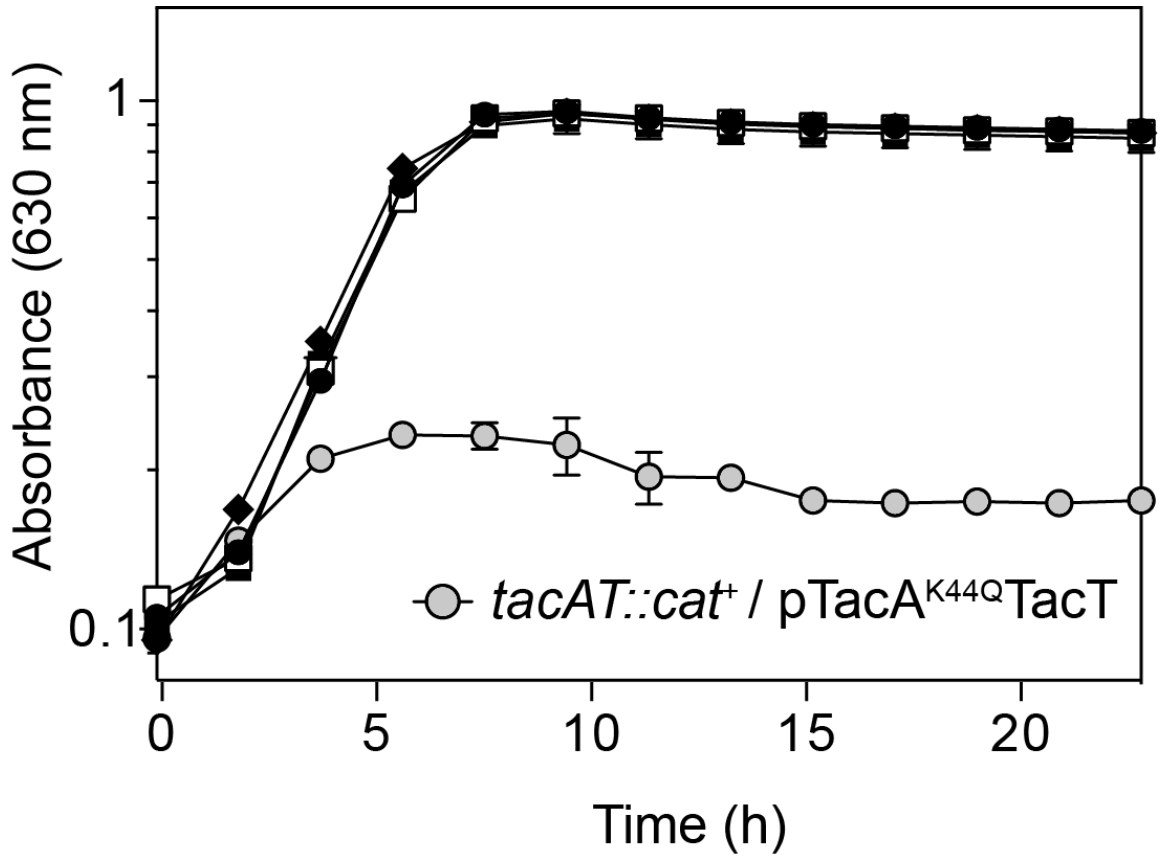

Figure 3.4. TacA variant mimicking acetylation in complex with TacT in *S. enterica tacAT::cat*$^{+}$ strains causes growth defect. Genes coding for TacT^{WT} (pTacT), TacAT^{WT} (pTacAT^{WT}), TacA^{K44R}TacT^{WT} (pTacA^{K44R}TacT^{WT}), TacA^{K544Q}TacT^{WT} (pTacA^{K44Q}TacT^{WT}), were cloned into a complementation vector pCV1) and introduced into strain backgrounds indicated by the following symbols: *tacAT*::*cat*$^{+}$ / pCV1 (empty cloning vector control), solid circles; *tacAT*::*cat*$^{+}$ / pTacT^{WT}, open squares; *tacAT*::*cat*$^{+}$ / pTacAT^{WT}, solid squares; *tacAT*::*cat*$^{+}$ / pTacA^{K44R}T^{WT}, diamonds; *tacAT*::*cat*$^{+}$ / pTacA^{K44Q}T^{WT}, gray circles. Overnight cultures were grown in NB rich medium. Cells were sub-cultured (1% v/v) and grown in NCE minimal medium supplemented with glycerol (22 mM), ampicillin. Expression of plasmid-borne genes was induced with 25 μM L(+)-arabinose. Growth curves were obtained in technical triplicates of biological triplicates and was repeated three times using a microplate reader (BioTek instruments). Error bars represent standard deviation of technical triplicates.

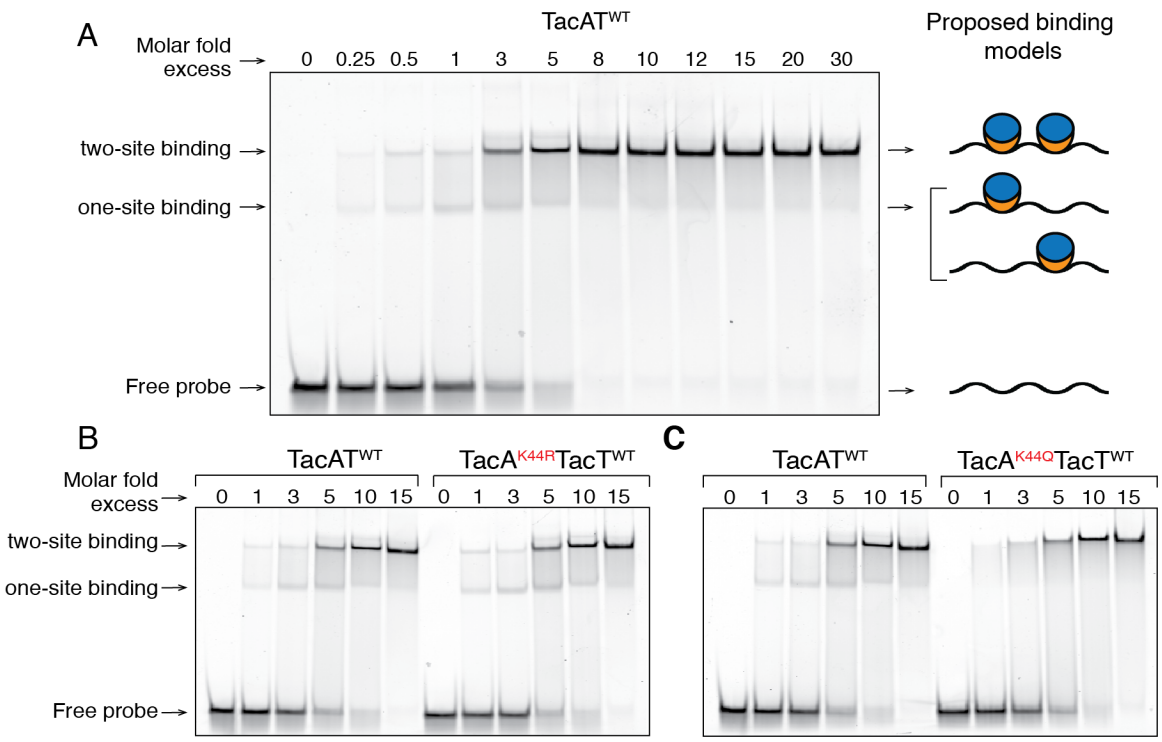

Figure 3.5. Binding of TacA^{WT}TacT^{WT} to the *tacAT* promoter differs from binding of TacA^{K44Q}TacT^{WT}. Binding of increasing concentrations of TacAT complexes to the *tacAT* promoter was assessed by electrophoretic mobility shift assays using 6-FAM 5'-labeled probes. A. Probe 3 (75 bp, 1.01pmol, position -81 to -6) was incubated with increasing concentrations of TacAT^{WT} protein (0.25, 0.5, 1.01, 3.03, 5.05, 8.08, 10.1, 12.12, 15.15, 20.2, and 30.3 pmol). The diagram on the right side of the figure depicts a model that could explain the occurrence of the two shifted bands seen (TacT, blue ovals; TacA, orange crescents). **B**. Probe 3 (1.06 pmol) incubated with increasing concentrations of either TacAT^{WT} or TacA^{K44R}TacT^{WT} (molar excesses of both complexes: 0.53, 1.06, 3.2, 5.3, and 10.6 pmol). **C**. Probe 3 (1.06 pmol) incubated with TacAT^{WT} or TacA^{K44Q}TacT^{WT} at same molar fold-excess as figure 3.5C.

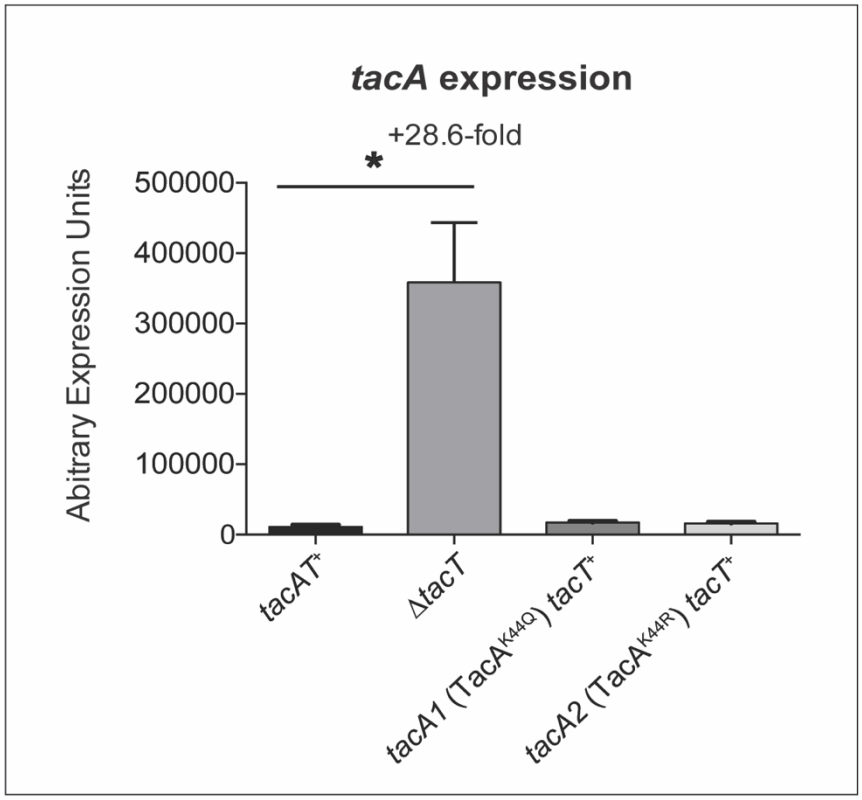

Figure 3.6. Deletion of *tacT* relieves repression of *tacA*. RT-qPCR was used to evaluate the expression of the *tacA* gene in the backgrounds of *tacAT*$^{+}$, Δ*tacT, tacA1 tacT*$^{+}$ (TacA^{K44Q}), and *tacA2 tacT*$^{+}$ (TacA^{K44R}) under the condition of NCE minimal medium supplemented with glycerol (22mM). The gene expression of *tacA* in the toxin deletion strain Δ*tacT* was 28.6-fold higher as compared to the *tacAT*$^{+}$ strain. Interestingly, the two lysine variant strains yielded no significant difference in expression of *tacA* as compared to the *tacAT*$^{+}$ strain. Error bars represent SEM. Asterisk indicates the Welch's *t* test gave a p value of 0.01. Experiment was conducted in technical triplicates of biological triplicates and was repeated three times.

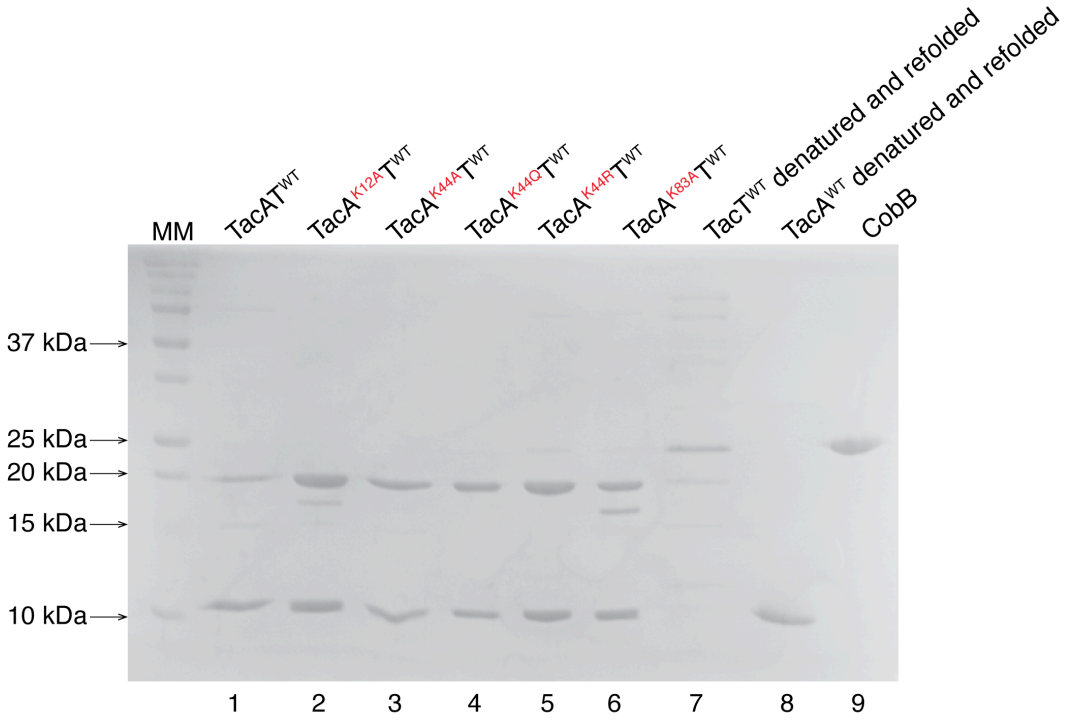

Figure 3.7. Percent purity of proteins used in these studies. A 15% (w/v) SDS-PAGE gel loaded with ~1 mg of each protein that is designated above each well. Percent purities are as follows, TacAT^{WT} : 88%, TacA^{K12A}TacT^{WT} : 97%, TacA^{K44A}TacT^{WT} : 99%, TacA^{K44Q}TacT^{WT} : 99%, TacA^{K44R}TacT^{WT} : 99%, TacA^{K83A}TacT^{WT} : 85%, TacT : 10%, TacA: 99%, CobB: 99%. Percent purity was calculated by diluting proteins (3, 2, 1, 0.5, 0.25 mg) on a separate gel and bands were quantified as percentages using ImageQuant™ TL. Contaminants of TacT denatured and refolded prep include: ClpB (ATP-dependent chaperone), HtpG (molecular chaperone), MdaA (NADPH nitroreductase), elongation factor Tu, peptidyl-prolyl *cis-trans* isomerase, Ferric uptake regulation protein, transposase, and YdaS (uncharacterized protein).

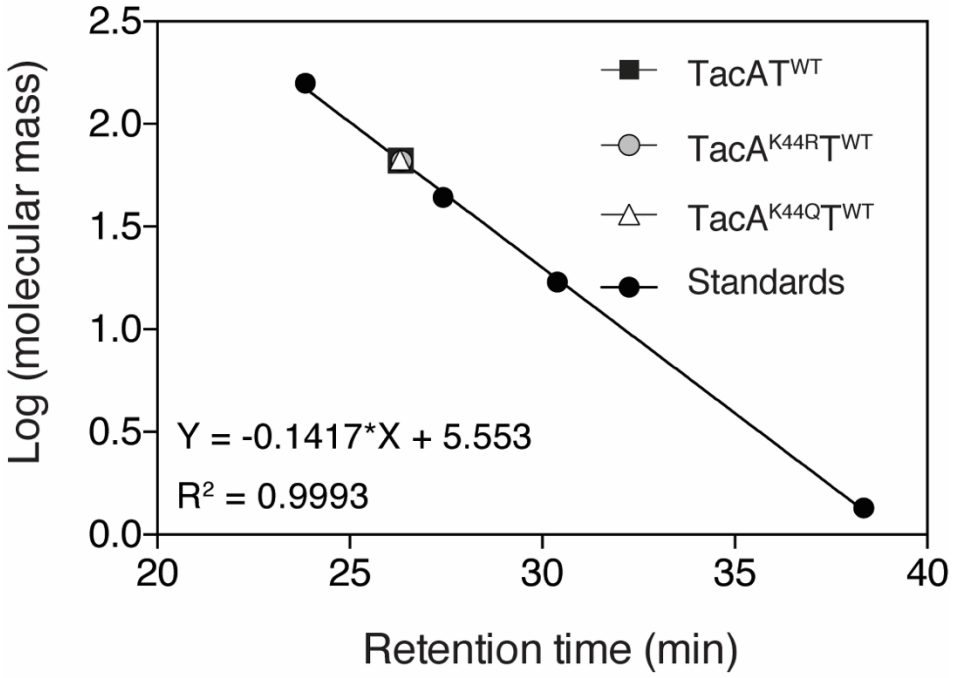

Figure 3.8. Size exclusion chromatography of purified complexes. The molecular masses of TacAT^{WT}, TacA^{K44R}TacT^{WT}, and TacA^{K44Q}TacT^{WT} were determined by size exclusion chromatography using a Superose 12 10/300 GL column as detailed under Materials and Methods. The molecular mass standards (white circles) were thyroglobulin (bovine; 670,000 Da), gamma globulin (bovine; 158,000 Da), ovalbumin (chicken; 44,000 Da), myoglobin (horse; 17,000 Da), and vitamin B_{12} (1,350 Da).

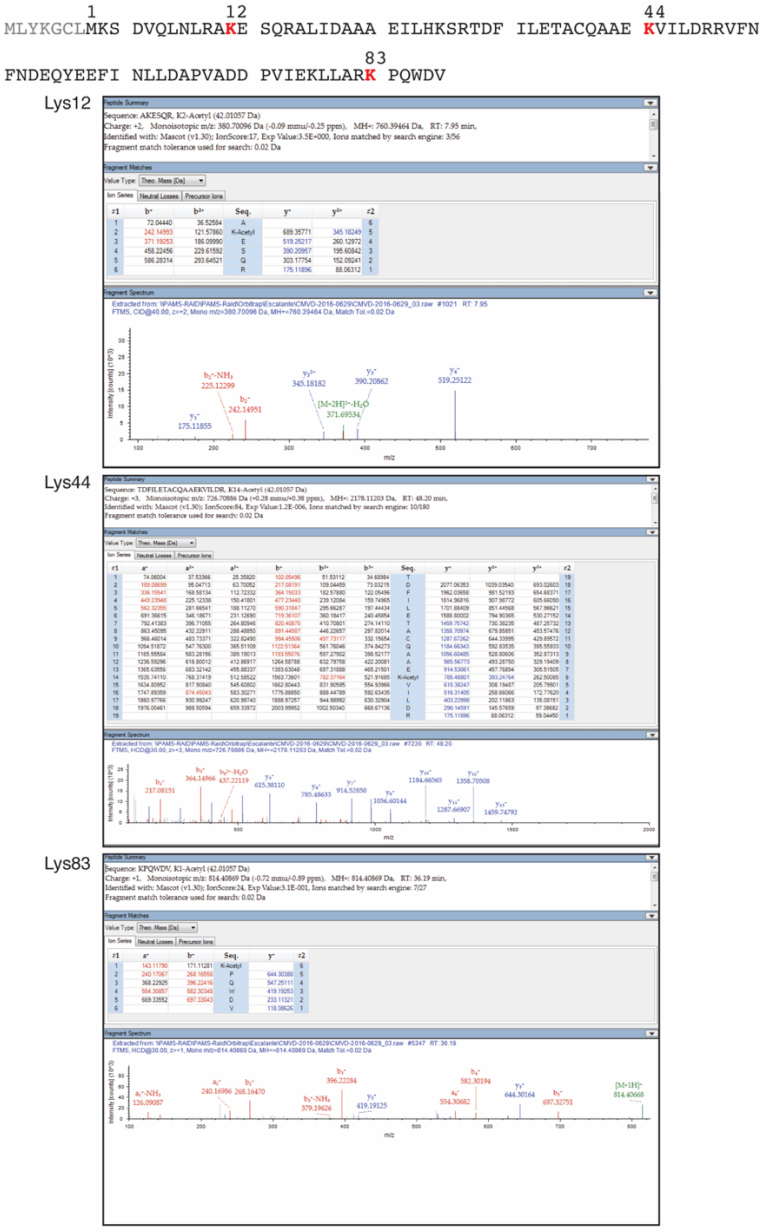

Figure 3.9. Mass spectrometry analysis of acetylated TacA. Sequence above represents annotated TacA protein sequence. Residues in grey represent amino acids not detected by LC/MS MS and therefore the start methionine was repositioned to begin with black sequence. Residues identified to be acetylated in TacAT^{WT} samples containing 1 mM acetyl-CoA are highlighted in red. *b* ions is the series of fragments that extend from the amino terminus; *y* ions are the series of ions that extend from the *C*-terminus. MASCOT software (http://www.matrixscience.com) was the online search engine used to identify peptides on the basis of their masses. Ion series a, b, y are expected ion types. The subscript of a, b or y ions represents the number of residues in the peptide; the superscript represents the peptide charge.

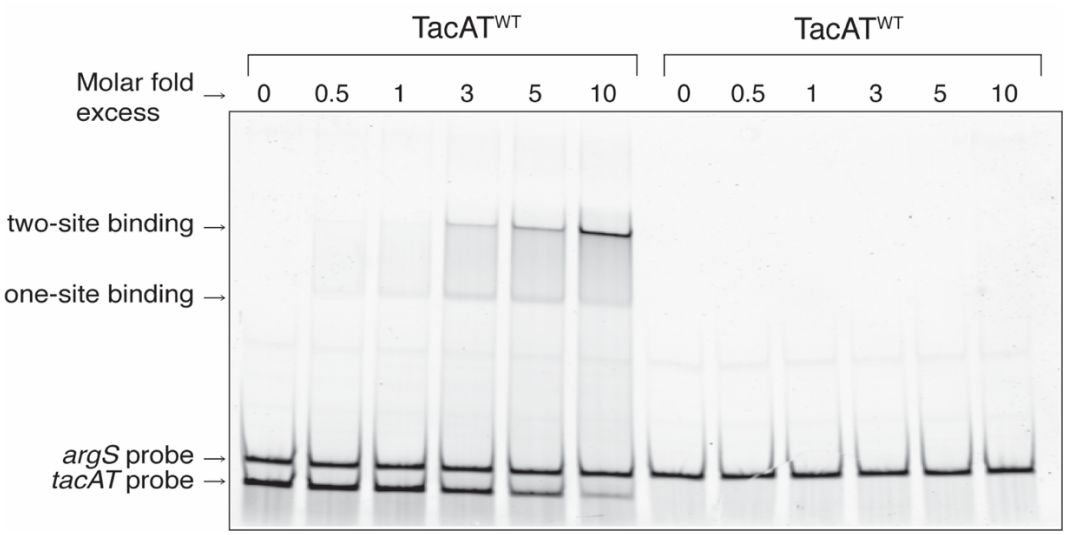

Figure 3.10. TacAT^{WT} complexes bind specifically to the *tacAT* promoter. The specificity of binding of TacAT^{WT} complexes to the *tacAT* promoter was analyzed by electrophoretic mobility shift assays using 6-FAM 5'-labeled probes. Left half: Probe 1 (157 bp, 0.48 pmol, position -163 to -6) and competitor DNA (*argS* promoter DNA, 0.38 pmol, 196 bp) were incubated together with increasing concentrations of TacAT^{WT} complex (0.24, 0.48, 1.45, 2.4, and 4.8 pmol protein to DNA). Right half: Promoter *argS* (0.38pmol) incubated with TacAT^{WT} at increasing molar fold excess (0.54, 1.08, 3.25, 5.41, and 10.8 pmol).

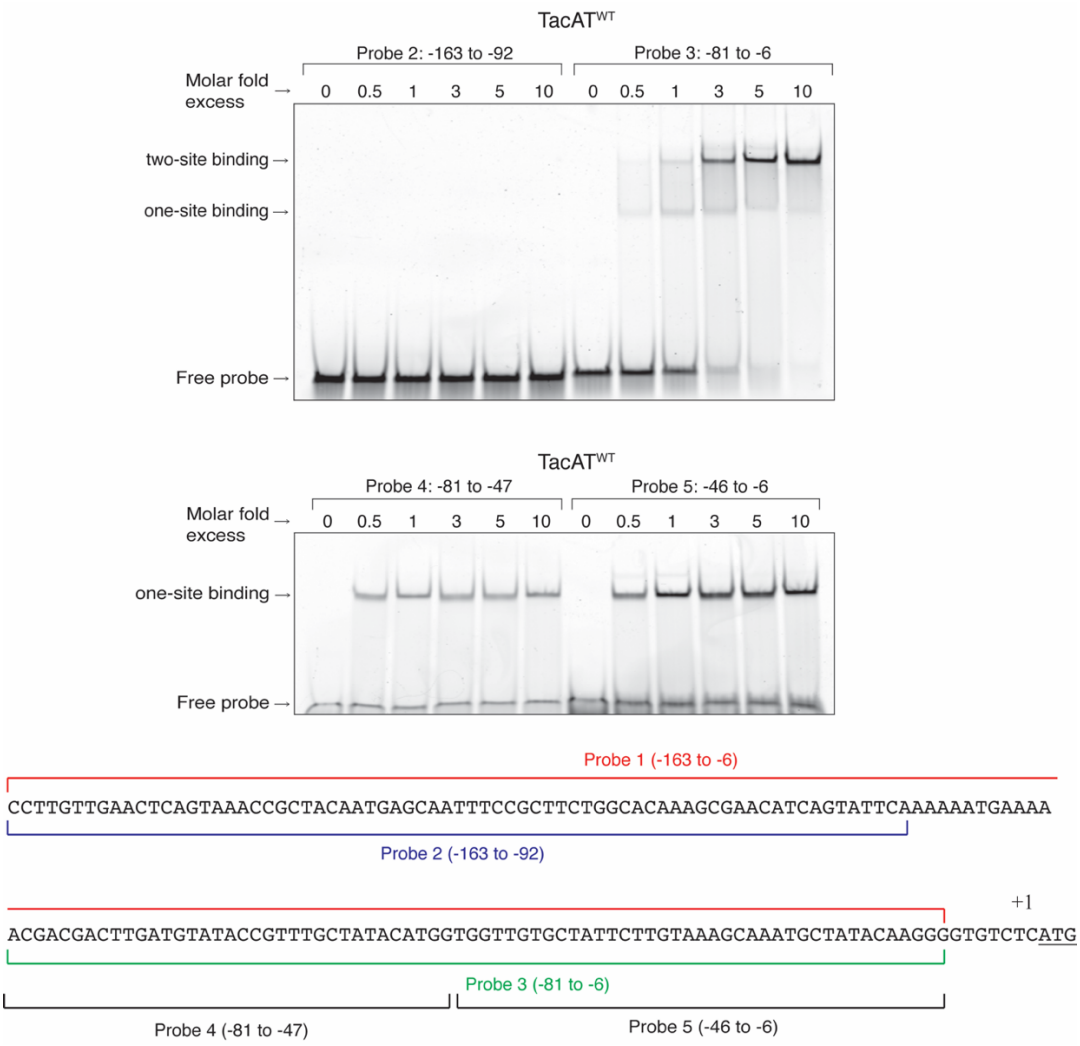

Figure 3.11. TacA-TacT complexes bind to a specific region within the *tacAT* promoter. An attempt to narrow down the binding region of the TacAT^{WT} complex within the *tacAT* promoter was conducted using electrophoretic mobility shift assays with 6-FAM 5'-labeled probes. Top figure, left: A fragment of DNA consisting of the upstream 71 base pairs of Probe 1 called Probe 2 (1.06 pmol, 71 bp) was incubated with increasing concentrations of TacAT^{WT} complex (Protein added to probe 2: 0.53, 1.06, 3.2, 5.3, 10.6 pmol). Top figure, right: TacA^{WT}TacT^{WT} complex (0.5, 1.01, 3.03, 5.05, 10.1 pmol) was incubated with the downstream 75 base pairs of Probe 1 called Probe 3 (1.01pmol). Bottom figure: To determine if two binding sites exist within Probe 3, the sequence of Probe 3 was then divided in half to form Probe 4 (upstream 35 bp region, 2.16 pmol) and Probe 5 (downstream 40 bp region, 1.89 pmol). TacAT^{WT} complex was incubated with either Probe 4 or Probe 5 to show binding to two different sites at the molar fold excess of 1.08, 2.16, 6.49, 10.8, and 21.6 pmol Protein to Probe 4 and 0.95, 1.89, 5.68, 9.5, and 18.9 pmol Protein to Probe 5. DNA sequence represents -163 bp to ATG of *tacA*. ATG start codon matches correct reading frame as mentioned in Results section.

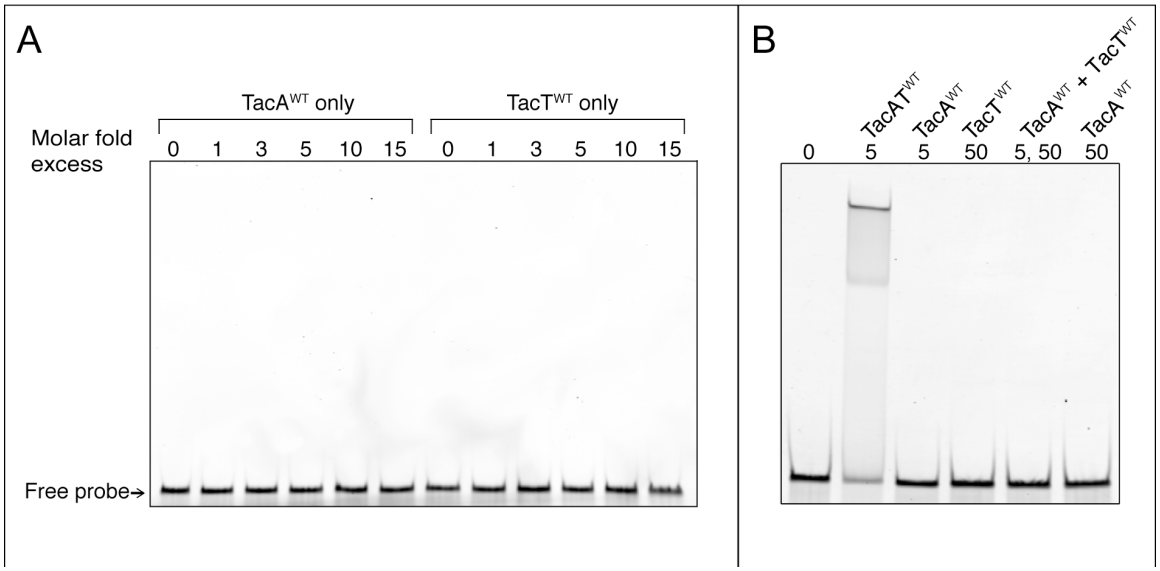

Figure 3.12. TacA and TacT alone do not bind to the *tacAT* promoter. (A) To test whether or not TacA^{WT} could bind to its own promoter in the absence of TacT^{WT}, probe 3 and TacA^{WT} protein were incubated together at 1.01, 3.03, 5.05, 10.1, and 15.15 molar fold excess protein to probe (1.01pmol). As a control to ensure that TacT^{WT} alone had no DNA binding activity, TacT^{WT} protein was incubated with probe 3 at the same molar fold excess as stated above. (B) To test whether TacA and TacT^{WT} protein could re-form a complex and bind to DNA, probe 1 (0.24 pmol) was incubated with either 5 molar-fold excess TacAT^{WT} complex, 5-fold excess TacA^{WT}, 50-fold excess TacT^{WT}, or mixed 5-fold excess TacA^{WT} and 50-fold excess TacT^{WT} together. Fifty-fold excess TacT^{WT} was used because TacT^{WT} was 10% pure (Fig. 3.8).

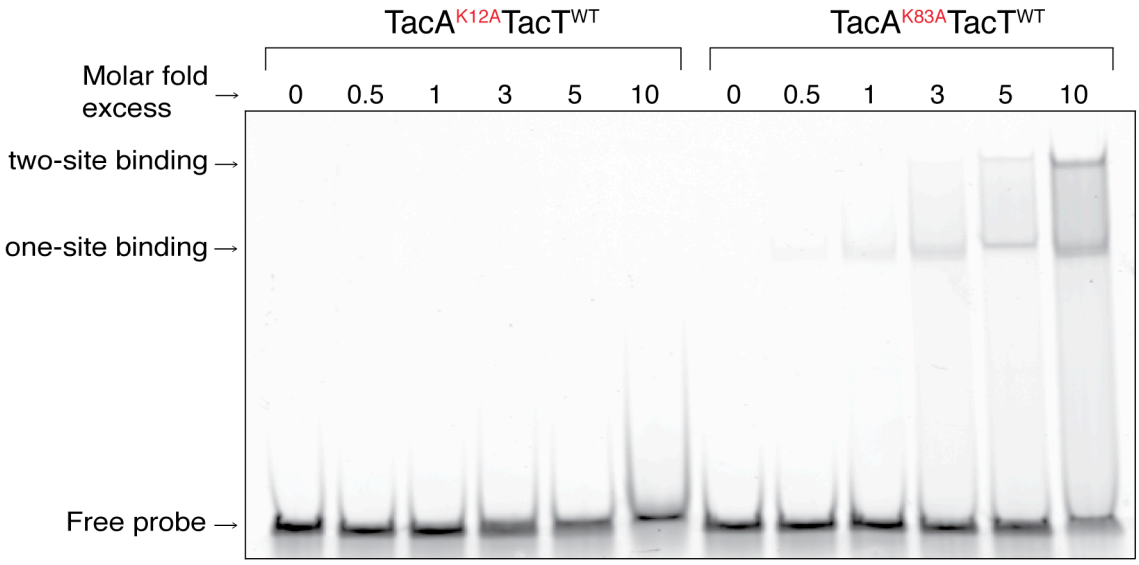

Figure 3.13. Residue K12 of TacA is needed to bind DNA. The ability of lysine variants TacA^{K12A}T^{WT} and TacA^{83}T^{WT} to bind to DNA was analyzed by electrophoretic mobility shift assay. Probe 3 (1.01 pmol) was incubated with increasing molar excess of either TacA^{K12A}T^{WT} or TacA^{K83A}T^{WT} (0.5, 1.01, 3.03, 5.05, 10.1 pmol protein).

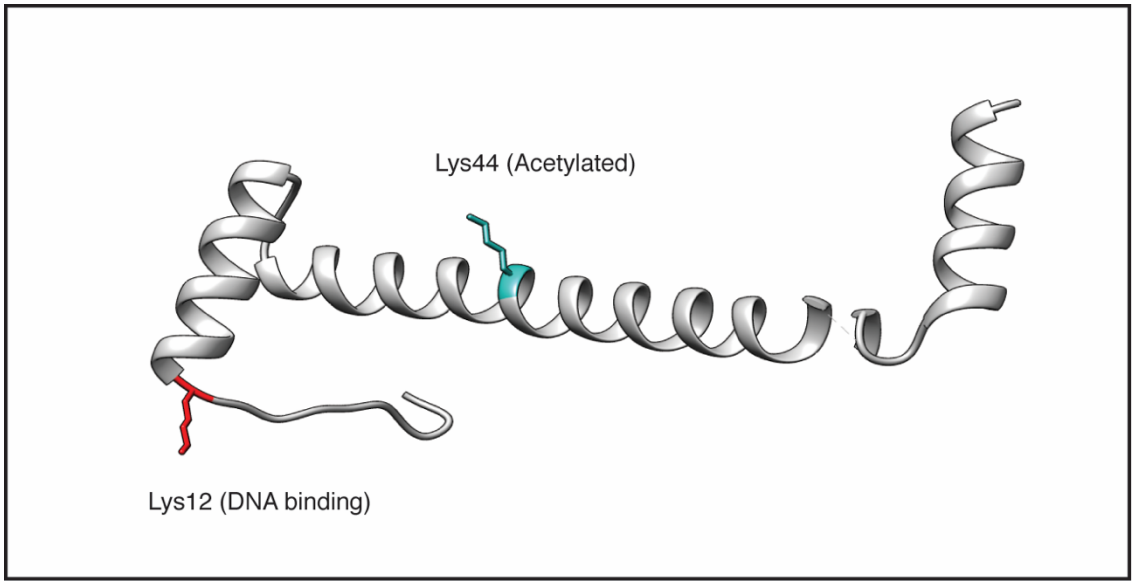

Figure 3.14. PHYRE model representation of antitoxin. Protein sequence of TacA was uploaded to PHYRE protein folding recognition server to analyze predicted structure of the antitoxin. Structure was colored and analyzed using Chimera software. Residue K12, which was required for DNA binding is highlighted in red. Residue K44, which was acetylated by the toxin is highlighted in cyan.

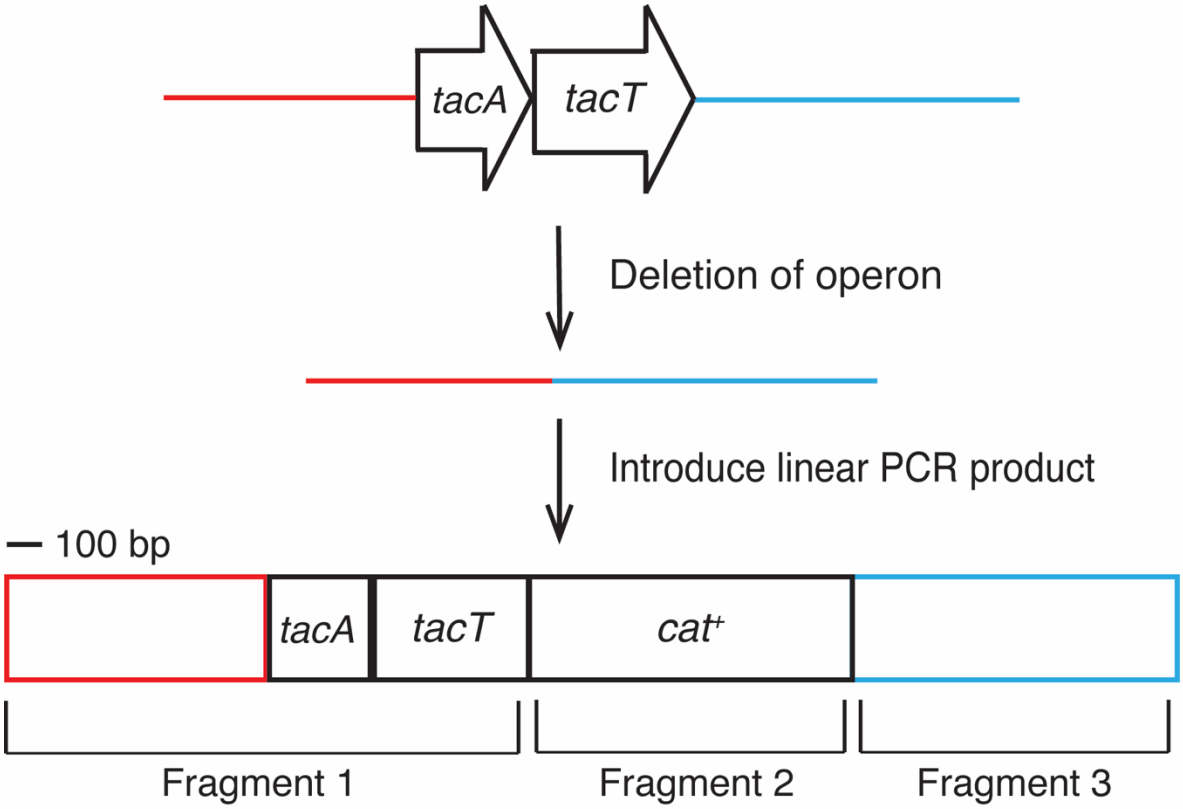

Figure 3.15. Schematic of SOE-qPCR fragment assembly as described in *Materials and methods*. Upstream region in red comprised of 792 bp upstream of *tacAT*. Downstream region ~1 kB of *tacAT*. Fragments were fused together using SOE-qPCR using primers listed in Table S2. DNA sizes to scale.

CHAPTER 4

MODULATION OF THE BACTERIAL COBB SIRTUIN DEACYLASE ACTIVITY BY N-TERMINAL ACETYLATION[3]

[3]Parks A.R., and Escalante-Semerena J.C. 2020. doi:10.1073/pnas.2005296117

Reprinted here with permission from the publisher.

4.1 ABSTRACT

In eukaryotic cells, the *N*-terminal amino moiety of many proteins is modified by *N*-terminal acetyltransferases (NATs). This protein modification can alter the folding of the target protein, affect binding interactions of the target protein with substrates, allosteric effectors or other proteins, or can trigger protein degradation. In prokaryotes, only ribosomal proteins are known to be *N*-terminally acetylated, and the acetyltransferases responsible for this modification belong to the Rim family of proteins. Here we report that, in *Salmonella enterica*, the sirtuin deacylase CobB long isoform ($CobB_L$) is *N*-terminally acetylated by the YiaC protein of this bacterium. Results of *in vitro* acetylation assays showed that $CobB_L$ was acetylated by YiaC; LC/MS/MS was used to confirm these results. Results of *in vitro* and *in vivo* experiments showed that $CobB_L$ deacetylase activity was negatively affected when YiaC acetylated its *N*-terminus. To our knowledge, this is the first report of i) posttranslational control of a bacterial sirtuin deacylase activity, ii) that the Gcn5-related YiaC protein is the acetyltransferase that modifies $CobB_L$, and iii) that YiaC is a *N*α acetyltransferase (NAT). Based on our data we propose the name of NatA, for N-acyltransferase A in lieu of YiaC to reflect the function of the enzyme.

4.2 SIGNIFICANCE. *N*-terminal protein acetylation is poorly understood in bacteria. Herein we report the identification of a *N*α acetyltransferase (NAT) that modulates the activity of a sirtuin deacylase in a human pathogen. This is significant because the alluded enzyme (named NatA, formerly YiaC) is the first

prokaryotic non-Rim-type NAT reported, and *N*-terminal acetylation of a bacterial sirtuin has not been reported. Also significant is the fact that NatA affects the metabolism of acetate, a short-chain fatty acid known to play important role in pathogenesis in the human gut.

4.3 INTRODUCTION

Protein acylation is common in prokaryotes and eukaryotes, and is an effective and rapid means of controlling protein function in response to diverse stimuli [1, 2]. What stands out about protein acylation is the diversity of organic acids used by cells to modify proteins (*e.g.,* acetate, propionate, malonate, succinate, etc), and the large number of acyltransferases that catalyze the modifications [1]. Many of the acyltransferases that modify proteins and small molecules belong to the protein superfamily PF00583, and among this family, many proteins contain the so-called GNAT domain (for GCN5-related acetyltransferase; IPR000182). GNATs acylate free amino groups of proteins or small molecules [1, 2]. For example, there is a subset of well-studied GNATs that modify the epsilon amino group (*N*ε) in lysine side chains [3, 4], while other GNATs modify the alpha amino group (*N*α) of the starting residue of proteins [5-8].

To frame the work reported here, we note that many eukaryotic proteins are acetylated on their *N*-termini, and that the acetyltransferases responsible for these modifications are referred to as *N*-acetyltransferases (hereafter NATs). In general, NATs catalyze the transfer of the acetyl-group from acetyl-CoA (AcCoA) to a

primary amine of a small molecule or the *N*-terminal amino group of a peptide or protein. In eukaryotes, *N*α acetylation has been suggested to alter protein folding, protein-protein interactions, and protein degradation [9-12]. In contrast, little is known about the enzymes that catalyze *N*-terminal protein acetylation in prokaryotes, and what the physiological reasons for such modification may be. Examples of *N*-terminal acetylation of bacterial proteins, where the acetyltransferase is known is the acetylation of ribosomal proteins S18, S5 and L7/L12 by acetyltransferases RimI, RimJ, and RimL [13-15]. These acetyltransferases were thought to have high substrate specificity until recently, when the *Mycobacterium tuberculosis* RimI was shown to acetylate different peptides *in vitro* [16]. Several other proteins are known to be *N*α acetylated in bacteria, such as secretion chaperone SecB of *E. coli* and the virulence factor ESAT-6 (ExsA) of several *Mycobacterium* species. In the case of the ESAT-6 (ExsA) protein, acetylation of its *N*-terminus abolished binding interactions with its protein partner CFP10 (ExsB), and attenuated *Mycobacterium marinum* virulence [17-19]. Even though there has been identification of *N*-terminal acetylation in prokaryotes, the acetyltransferases responsible for the modifications have not been identified. For example, recently published bacterial *N*-terminal acetylomes of *Pseudomonas aruginosa*, *Acinetobacter baumannii*, and *Mycobacterium tuberculosis* showed that roughly ~10% of the proteins were *N*-terminally acetylated [8, 11, 20], but the enzymes responsible for such modifications were not identified.

The *S. enterica* genome contains ~26 GNATs, but the function of about a third of them has yet to be defined. In addition, the *S. enterica* genome possesses one NAD^+-dependent sirtuin deacylase CobB, whose function works in concert with the protein acetyltransferase (Pat) enzyme to reversibly modulate the activity of acetyl-CoA synthetase (Acs) [21, 22]. As shown below in scheme 1, the CobB-dependent deacetylation reaction consumes NAD^+ and yields *O*-acetyl-ADP-ribose (*O*-AADPR) and nicotinamide (Nm).

$$\text{Acs} + \text{AcCoA} \underset{\text{CobB}}{\overset{\text{Pat}}{\rightleftharpoons}} \text{Acs}^{\text{Ac}} + \text{CoA} + \text{AMP} + \text{PP}_\text{i}$$

$$\text{Nm} + \textit{O}\text{-AADPR} \qquad \text{NAD}^+$$

(Scheme 1)

Interestingly, this bacterium synthesizes two biologically active isoforms of CobB, referred to as $CobB_S$ (short isoform) and $CobB_L$ (long isoform), which differ in size by a 37-residue *N*-terminal extension of the catalytic core (Fig. 4.1).

Our group showed that both isoforms of CobB deacetylated their *bona fide* protein substrate, *i.e.,* acetylated acetyl-CoA synthetase (Acs^{Ac}) *in vivo* and *in vitro* [23]. However, the physiological relevance of the two CobB isoforms in *S. enterica* is unknown. Here, we report that the $CobB_L$ sirtuin deacylase isoform of this bacterium is *N*-terminally acetylated, and that the putative YiaC acetyltransferase acetylates the *N*-terminus of $CobB_L$. We also report *in vivo* and *in vitro* evidence that YiaC-dependent *N*-terminal acetylation of $CobB_L$ negative affects its deacetylase activity.

4.4 MATERIALS AND METHODS

Bacterial strains. All strains constructed were derivatives of *Salmonella enterica enterica* sv Typhimurium LT2 (*S. enterica* throughout the paper) (Table S1). Genes were deleted using described protocols [24]. Bacterial cultures were grown at 37ºC with shaking unless stated otherwise.

Strain construction. All strains used in this study were derivatives of strain JE6583 (*metE205 ara-9*). The Δ*cobB* deletion and *yiaC::cat*$^+$ strains were constructed as follows. Pfu Ultra II DNA polymerase (Stratagene) was used to amplify flanking regions of the plasmid pKD3 [24] fused to 36 to 39 bp of internal to the beginning or end of the *cobB* gene. PCRs were resolved on a 1% (w/v) agarose gel and checked for amplification by post-staining with ethidium bromide (0.5 μg /ml) for 10 min. Polymerase chain reaction (PCR) fragments were cleaned with the Wizard SV gel and PCR cleanup kit (Promega), and ~200 ng of product was electroporated into *S. enterica* strain JE6692 (*metE205 ara-9* / pKD46) using a 0.2-cm electroporation cuvette (MidiSci) and a microPulser electroporator (Bio-Rad Laboratories) on Ec2 setting. Cells were incubated at 37ºC with shaking, plated on lysogeny broth (LB) supplemented with 12.5 μg/ml of chloramphenicol, and incubated at 37ºC overnight. Insertion of *cat*$^+$ was confirmed by PCR, then the marker and deletion were moved by P22-mediated transduction into strain JE6583 as described elsewhere [25]. Colonies were streaked to isolation several times under antibiotic selection and phage sensitivity was tested. To resolve the *cat*$^+$ marker of the *cobB*::*cat*$^+$ strain, plasmid pCP20 [24] was introduced into strains of

interest by transformation by electroporation as described elsewhere [26]. The deletion of the gene was confirmed by Sanger sequencing using primers flanking the gene of interest.

Plasmid construction for complementation and overexpression. Plasmids used in this study are listed in Table S2. Primers used in this study were manufactured by Integrated DNA Technologies, Inc. (IDT; Coralville, IA), and are listed in Table S3. The high-efficiency cloning method of BspQI digestion described elsewhere [27] was used to clone the *cobB* or *yiaC* genes into plasmids pCV1 or pCV3 vectors The *cobB* allele encoding the $CobB_S$ isoform and *yiaC* were amplified from *S. enterica* genomic DNA using Phusion High-Fidelity DNA polymerase (New England Biolabs), while the *cobB* allele encoding $CobB_L$ was amplified from plasmid pCOBB19 where quick-change mutagenesis was performed to change M37 and M38 to alanines [primers and plasmid information for the construction of this allele can be found in [23]]. PCR products were analyzed by DNA gel electrophoresis on 1% (w/v) agarose gels after staining with ethidium bromide (0.5 mg/ml). PCR products were cleaned using the Wizard SV gel and PCR cleanup kit (Promega) and cut with BspQI restriction enzyme (New England Biolabs) at 50ºC. *cobB* gene fragments were digested for 30 min, then heat inactivated for 5 min at 80ºC. The *yiaC* gene was cut for one hour and was not heat inactivated.

PCR products of *cobB* for use in protein over production were digested with KpnI (5’ primer overhang) and SalI (3’ overhang) restriction enzymes and cloned into

pTEV6 (a protein over production plasmid with a maltose-binding protein (MBP)-H_6 fusion) [28]. For cloning of *cobB* into the *C*-terminal H_6 fusion vector pET-23a, PCR products were digested with XbaI (5') and XhoI (3'), and for cloning of *cobB* into the *N*-terminal H6 fusion vector pTEV5, *cobB* PCR products were digested with NheI (5') and XhoI (3'). The PCR product of the *yiaC* gene was cut with KpnI (5' overhang) and NotI (3' overhang) restriction enzymes for ligation into plasmid pTEV6. All digests were conducted at 37ºC for one hour using FastDigest enzymes from Thermo Fisher Scientific. PCRs and digested plasmids were cleaned with the Promega Wizard® SV gel and PCR clean-up system; digested PCR fragments were ligated into digested plasmids using T4 DNA ligase (Fisher) for one hour at room temperature, with the exception that PCR fragments were ligated into plasmids at a 3:1 plasmid:fragment ratio.

After transformation into *E. coli* DH5a cells and colony PCR screening of ligated plasmids, correct plasmids were isolated using the Wizard Plus SV miniprep kit (Promega). Sanger DNA sequencing was conducted at the Georgia Genomics and Bioinformatics Core (GGBC) at the University of Georgia to rule out mutations and confirm insertions.

Residues K14 and K16 of $CobB_L$ were changed to alanine by site-directed mutagenesis of *cobB* on plasmid pCOBB72. Primers used for this purpose were designed using Stratagene's QuikChange Primer Design software package (https://www.agilent.com/store/primerDesignProgram.jsp). PCR was conducted using Pfu Ultra II DNA polymerase (Agilent) with primers listed in table S3. Modification to this polymerase protocol included an extension temperature of

68ºC and extension time of 2.5 min per kb. Mutations were confirmed by Sanger sequencing.

***In vivo* growth analyses including culture media and chemicals**. All bacterial strains and plasmids used in this study are listed in Table S1 and S2. Growth studies were conducted as follows: starter bacterial cultures were grown from a single colony in nutrient broth (NB [Difco]) overnight shaking at 37ºC. In the morning, cells were sub-cultured into no-carbon essential (NCE) minimal medium [29] containing sodium acetate (10 mM), $MgSO_4$ (1 mM), Wolfe's trace minerals [30], ampicillin (100 μg/ml), chloramphenicol (10 μg/ml), and L-(+)-arabinose (100 μM). Cultures were inoculated at 1% (v/v), 2 μl of culture was used to inoculate 198 μl of fresh medium and then pipetted into individual wells of a 96-well polystyrene plate (Falcon). Microtiter plates were incubated at 37 ºC shaking continuously inside a PowerWave microtiter plate reader (Bio-Tek Instruments). Density of cells was monitored hourly at 630 nm and data were analyzed using Prism 8 (GraphPad).

Purification of $CobB_S$, $CobB_L$, and YiaC proteins

$CobB_S$ and $CobB_L$. The CobB isoforms were purified as follows: plasmids pCOBB71 and pCOBB72 were electroporated into *E. coli* C41 (λDE3) Δ*pat* (strain JE9314). Cultures of *E. coli* containing plasmids pCOBB71 or pCOBB72 were grown overnight in super broth [31] and were sub-cultured 1:100 (v/v) into two liters each of super broth containing ampicillin (100 μg/ml). Cells were grown shaking at

37°C to an optical density (measured at 600 nm; OD_{600}) of 0.7, induced with isopropyl-β-D-thiogalactopyranoside (IPTG, 0.5 mM), and grown overnight at 37 °C. After 16 h of incubation, cells were pelleted at 6,000 x *g* for 15 min in a Beckman Coulter Avanti J-2 XPI centrifuge equipped with a JLA-8.1000 rotor. Cell pellets were re-suspended in 60 ml of cold buffer A [sodium phosphate buffer (20 mM, pH 7.5) containing NaCl (500 mM), imidazole (20 mM), glycerol (10%, v/v)]. Phenylmethanesulfonyl fluoride (PMSF, 1mM), lysozyme (1 mg/ml), and DNase I (1μg/ml) were added to the cell suspension, and cells were lysed on ice by three, 30-s cycles of sonication (2 sec on, 2 sec off) at 60% amplitude on a Qsonica sonicator. Cell lysates were centrifuged at 40,000 x *g* for 30 min in a Beckman Coulter Avanti J-25I centrifuge equipped with a JA-25.50 rotor centrifuge then filtered with a 0.45 □m filter. Filtered lysates were applied to a 5-ml HisTrap FF column (GE Healthcare) using an ÄKTA fast protein liquid chromatography (FPLC) system (GE Healthcare). The column was washed with 15 column volumes of buffer A, 15 ml of 4% elution buffer B [sodium phosphate buffer (20 mM, pH 7.5) containing NaCl (500 mM), imidazole (500 mM), glycerol (10%, v/v)], followed by a 15-column volumes linear gradient elution from 4% to 100% buffer B, and 5 column volumes of 100% buffer B. Fractions were resolved on a 15% (w/v) sodium dodecyl sulfate-polyacrylamide electrophoresis (SDS-PAGE) gel [32] and elution fractions where MBP-H_6-$CobB_L$ were present were pooled, recombinant tobacco etch virus (rTEV) protease [33] was added at a concentration of 1 mg TEV: 100 mg protein, and dialyzed for 3 h at room temperature in sodium phosphate buffer (20 mM, pH 7.5) containing, NaCl (500 mM), glycerol (10%, v/v), dithiothreitol

(DTT, 1mM). Protein mixtures were dialyzed at 4°C overnight in 20mM sodium phosphate buffer pH 7.5, 500mM NaCl, 10% glycerol (v/v), then at 4°C for 3 h in sodium phosphate buffer (20 mM, pH 7.5), containing NaCl (500 mM), glycerol (10%, v/v), and imidazole (20 mM). After dialysis the protein was filtered and re-loaded onto the 5-ml HisTrap column, and the above protocol was repeated. Fractions were resolved by SDS-PAGE and fractions containing only $CobB_L$ were pooled and dialyzed at 4 °C for 3 h in HEPES buffer (20 mM, pH 7.5), containing NaCl (250 mM), glycerol (10%, v/v), then dialyzed and stored in HEPES buffer (20 mM, pH 7.5), containing NaCl (100mM), glycerol (10%, v/v). Concentrations of $CobB_L$ were determined using a Nanodrop® ND-1000 spectrophotometer (Thermo Fisher Scientific) by measuring the absorbance at 280 nm and using the molar extinction coefficient of the protein (22,460 M^{-1} cm^{-1}).

$CobB_L$ lysine variants, *N*-terminal, H_6-tagged $CobB_L$ (H_6-$CobB_L$) and *C*-terminal, H_6-tagged $CobB_L$ ($CobB_L$-H_6) were purified as follows: plasmids pCOBB151, pCOBB152, pCOBB153, pCOBB50, and pCOBB48 were electroporated into *E. coli* C41(λDE3) □*pat* (strain JE9314, Table 1). *E. coli* cells harboring the aforementioned plasmids were grown overnight in lysogeny broth (LB) [34, 35] and sub-cultured 1:100 (v/v) in the morning into two liters of super broth containing ampicillin (100 μg/ml). Cultures were grown with shaking at 37ºC to an optical density (OD) of 0.7 at 600 nm prior to induction of the gene encoding the protein of interest by the addition of 0.5 mM IPTG. Cultures were incubated at 37ºC for an additional 15 h after induction. Cells were harvested by centrifugation at 6000 x *g* for 15 min in a Beckman Coulter Avanti J-2 XPI centrifuge equipped with a JLA-

8.1000 rotor. Cell pellets were resuspended in 40 ml of Buffer A (sodium phosphate heptahydrate (17 mM), sodium phosphate monohydrate (3 mM), NaCl (0.5 M), imidazole (20 mM), 10% (v/v) glycerol pH 7.5). PMSF (1mM), lysozyme (1 mg/ml), and DNase I (1μg/ml) were added to the cell suspension, and cells were lysed on ice by three, 30-s cycles of sonication (2 sec on, 2 sec off) at 60% amplitude on a Qsonica sonicator. Cell lysates were centrifuged at 40,000 x *g* for 30 min in a Beckman Coulter Avanti J-25I centrifuge equipped with a JA-25.50 rotor centrifuge then filtered with a 0.45-μm filter. Filtered lysates were applied onto 5 ml of HisPur Ni-NTA resin. Resin was washed with 32 ml of buffer A, 15 ml of 4% buffer B (sodium phosphate heptahydrate (17 mM), sodium phosphate monohydrate (3 mM), NaCl (0.5 M), imidazole (0.5 M), 10% (v/v) glycerol pH 7.5), and proteins were eluted off the column and collected in three 8-ml fractions of 100% buffer B; 24 ml total). All fractions collected from purifications were resolved via SDS-PAGE analysis [32] using the PrecisionPlus Protein ladder (Bio-Rad Laboratories) to generate a molecular mass ladder. Proteins were dialyzed against one liter of buffer A containing dithiothreitol (DTT, 1 mM) at room temperature for three hours while the maltose-binding-protein (MBP)-H_6 was cleaved at a ratio of 100 mg of protein:1 mg of H_6- rTEV) protease purified as described elsewhere [36], except the H_6-CobBL and $CobB_L$-H_6 were not cleaved and were dialyzed in storage buffer (listed below) and frozen at -80ºC. Cleaved proteins were dialyzed twice more against 1 L of buffer A at 4ºC. After cleavage, proteins were loaded onto a 5-ml HisPur Ni-NTA column, which was developed by gravity. The rest of the purification steps were as described above. After a second SDS-PAGE analysis

proteins were dialyzed in a step-wise manner to decrease the concentration of sodium chloride. Three 1-L changes were performed with the following buffers: 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid buffer (HEPES, 20 mM pH 7.5) containing glycerol [10% (v/v)], and either 250 mM or 100 mM NaCl. The $CobB_L^{K14A,K16A}$ protein were further dialyzed into HEPES buffer (20 mM, pH 7.5) containing glycerol [10% (v/v)], and NaCl (50 mM), and were applied onto a 5-ml column of phosphocellulose resin. $CobB_L^{K14A,K16A}$ was eluted with a 0.05-1 M step-wise gradient of NaCl. This was a very effective chromatographic step since $CobB_L^{K14A\ K16A}$ proteins eluted off the column when the NaCl concentration reached 1 M. All proteins were dialyzed thrice into the storage HEPES buffer (20 mM, pH 7.5) containing glycerol [10% (v/v)], and NaCl (100 mM) at 4ºC, followed by flash freezing in liquid nitrogen. The concentration of proteins purified was determined using a Nanodrop® ND-1000 spectrophotometer (Thermo Fisher Scientific) by measuring the absorbance at 280 nm and using the molar extinction coefficient of the CobB protein (22,460 M^{-1} cm^{-1}) as determined by ProtParam (https://web.expasy.org/protparam/).

YiaC. Plasmid pYIA6 encoding wild-type YiaC was transformed into *E. coli* C41(λDE3) Δ*pat* (strain JE9314). An overnight culture of *E. coli* harboring pYIA6 was sub-cultured in a 1:100 ratio into 1 L of Terrific Broth (TB; tryptone (12 g/L), yeast extract (24 g/L), glycerol (4 mL/L), $MgCl_2$ (2mM)buffered with KH_2PO_4 (0.17 M) plus K_2HPO_4 (0.72 M) containing ampicillin (100 μg/ml) . Cultures were grown with shaking at 37ºC to an OD_{600} of 0.5, induced with IPTG (0.5 mM), then allowed

to grow with shaking at 37ºC for an additional 18 h. Cell-free lysate was prepared as described above. Filtered crude cell-free extract was applied onto a 1-ml HisBind FF column (GE Healthcare) pre-equilibrated with HEPES buffer (50 mM, pH 7.5) containing NaCl (0.5 M), imidazole (20 mM), glycerol (5%, v/v), *tris*-(2-carboxyethyl)phosphine (TCEP, 1 mM) using an ÄKTA fast protein liquid chromatography (FPLC) system (GE Healthcare). The column was washed with buffer A and MBP-H_6-YiaC was eluted with a linear gradient (20 to 500 mM) of imidazole. Fractions containing YiaC (as determined by SDS-PAGE) were pooled and dialyzed against buffer A containing DTT (1 mM) for 3 h at room temperature with a 10 mg of MBP-H_6-YiaC:1 mg of H_6-rTEV protease ratio. Cleaved proteins were dialyzed twice more at 4ºC against buffer A then re-applied onto a 1-ml HisBind FF column to separate cleaved proteins from MBP-H_6 and H_6-rTEV protease. Cleaved proteins were collected in the flow through, analyzed by SDS-PAGE, pooled and dialyzed against storage buffer (HEPES (20 mM), NaCl (150 mM), TCEP (1 mM), glycerol (20%, v/v) pH 7.5). The concentration of YiaC was determined using a Nanodrop 1000 spectrophotometer (Thermo Scientific) using the molecular mass of 17.2 kDa as determined by size exclusion chromatography and Expasy ProParam, and extinction coefficient (47,440 M^{-1} cm^{-1} [Expasy ProParam]).

Size exclusion chromatography. The molecular mass of YiaC was determined by size exclusion chromatography using a Superose 12 10/300 GL gel filtration column (GE Healthcare Life Sciences) equilibrated as per manufacturer's protocol

using water and HEPES buffer [50 mM, pH 7.0 at 4C] plus NaCl (150 mM). First, gel filtration standards of bovine thyroglobulin (670,000 Da), bovine gamma globulin (158, 000 Da), chicken ovalbumin (44,000 Da), horse myoglobin (17,000 Da), and vitamin B_{12} (1,350 Da)(Bio-Rad Laboratories) were applied onto the column using a 100-μl superloop. The retention times of the standards were used to generate a standard curve using the $\log_{10}$ molecular mass of each standard. Similarly, YiaC protein was eluted from the column, retention time was recorded, and molecular mass determined using the equation y = -0.138x + 5.5442 calculated from the standard curve. The determination of the oligomeric state of YiaC was conducted in triplicate. The reported oligomeric state of YiaC is the average molecular mass of three independent experiments.

***In vitro* acetylation assays.** For experiments shown in figure 2, homogeneous $CobB_L$ or $CobB_S$ proteins (5 μM) were incubated with [1-^{14}C]-AcCoA (20 μM) in HEPES buffer (50 mM, pH 7) containing *tris*-(2-carboxyethyl-ohosphine (TCEP) hydrochloride (1 mM) with or without YiaC protein (3 μM) for 1 h at 37ºC in a total volume of 25 μl. Reactions were stopped by the addition of 5 μl of SDS-loading dye (glycerol (60%, v/v), Tris-HCl [0.3 M, pH 6.8], 12 mM ethylenediaminotetracetic acid (EDTA), sodium dodecyl sulfate (SDS, 12% w/v) plus 2-mercaptoethanol (0.87 mM), bromophenol blue (0.05%, w/v) and reactions were resolved by SDS-PAGE at 200 V for 45 min on a 15% (w/v) polyacrylamide gel with Tris-HCl buffer at pH 8.8 for the resolving gel and Tris-HCl at pH 6.8 for the stacking gel. After electrophoresis, gels were stained with Coomassie dye [Coomassie Brilliant Blue

R (1g/L), isopropanol (25%, v/v), glacial acetic acid (10%)] then exposed overnight to BAS storage phosphor screens (GE Healthcare). Radiolabel transfer from [1-^{14}C]-AcCoA to $CobB_L$ was visualized using a Typhoon Trio Plus variable mode imager on phosphor imaging setting (GE Healthcare). *In vitro* acetylation assays with $CobB_L$ peptides was performed as described above using a ratio of $CobB_L$ peptides (5 μM) to YiaC (3 μM). *In vitro* acetylation assays of lysine variants of $CobB_L$, *N*-terminally H_6-tagged or *C*-terminally H_6-tagged $CobB_L$ were performed as stated above except 3 μM of CobB proteins and 1 μM of YiaC protein was used. Due to low yield of protein purified, only 2 μM of the $CobBL^{K14A\ K16A}$ variant was used.

***In vitro* deacetylation assays**. For experiments shown in figure 8, homogeneous $CobB_L$ protein (12 μM) was incubated with [1-^{14}C]-AcCoA (20 μM) with or without YiaC (8 μM) for one hour in HEPES buffer (50 mM pH 7) containing TCEP (1mM) at 37ºC in a final volume of 50 μl. After a one-hour incubation, ethyl-CoA was added to both reactions to a final concentration of 32 mM, and incubated at 37ºC for one additional hour. Simultaneously, a reaction mixture containing Acs (3 μM), Pat (1 μM), of [1-^{14}C]-AcCoA (20 μM) was incubated at 37ºC for one hour in 200 μl of HEPES buffer (50 mM, pH 7) containing TCEP (1 mM). After incubation, the Acs and Pat reaction mixture was buffer exchanged with an Amicon Ultra 0.5-mL centrifugal filter Ultracel-10K (10 kDa molecular mass cut off) to remove excess AcCoA. Before elution from the filter, proteins were washed on the filter twice with

400 μl of HEPES buffer (10mM, pH 7). Eluted Acs was brought to a concentration of 6 μM in HEPES buffer (10mM, pH 7).

Deacetylation reactions were set up as follows: Acetyl-CoA synthetase acetylated with [^{14}C-1]Ac-CoA (Acs^{Ac*}) (3 μM) was added to a reaction mixture containing $CobB_L$ (3 μM) or *N*-terminally acetylated $CobB_L$ ($^{Ac}CobB_L$) (3 μM) plus dithiothreitol (DTT, 2 mM), NAD^+ (2 mM) in HEPES buffer (50 mM, pH 7). Negative controls of the reactions above were also set up without NAD^+ added. Reactions were allowed to incubate for one hour at 37ºC. Reactions were terminated by the addition of SDS loading dye (ingredients listed under *acetylation assays*). Samples were resolved by SDS-PAGE on a 12% (w/v) polyacrylamide gel developed at 215V for 45 min. The gel was stained with Coomassie Brilliant Blue R as stated under *in vitro protein assays* and imaged in the same manner.

We used purified *S. enterica* protein acetyltransferase (*Se*Pat) enzyme to acetylate *S. enterica* acetyl-CoA synthetase (Acs) using [1-^{14}C]-AcCoA as described elsewhere [3] (Scheme 1). We chose *S. enterica* Acs (*Se*Acs), since acetylated *Se*Acs ($SeAcs^{Ac}$) is a *bona fide* substrate of $CobB_L$ [23]. We also radiolabeled $CobB_L$, and for this purpose we used YiaC. To label $CobB_L$, reaction mixtures containing $CobB_L$ + [1-^{14}C]-AcCoA $\pm$ YiaC were incubated for one hour at 37ºC. To terminate the reaction, an excess of ethyl-CoA was added to the reaction mixture to inhibit YiaC activity. To determine how much ethyl-CoA was needed to ensure YiaC inhibition, we quantified the effect of ethyl-CoA on YiaC activity, and determined that at concentrations higher than 4 mM ethyl-CoA, YiaC was completely inhibited (Fig. 4.10). Given the amount of YiaC used to label

$CobB_L$, we terminated the reaction by adding ethyl-CoA to a final concentration of 32 mM.

***In vitro* acetyl-CoA synthetase activity assay**. Strains JE23602 and JE23603 were inoculated 1:100 into one liter each of minimal medium supplemented with acetate (10 mM) as described above except 50 μM of L-(+)-arabinose was used to induce genes on plasmids. Cells were grown at 37ºC shaking at 160 rpm to OD_{600} of 0.15 (mid-log on low acetate), centrifuged at 6000 x *g* for 15 min in a Beckman Coulter Avanti J-2 XPI centrifuge equipped with a JLA-8.1000 rotor. Pellets were resuspended in 500 μl of HEPES buffer (50 mM, pH 7.5). Cell lysates were sonicated 3 x 30 s with 2 s on, 2 s off intervals using a Fisher Scientific Sonic Dismembrator model 150I. Sonicated lysates were centrifuged in a microcentrifuge at 16,000 x *g* for 20 min at 4 ºC. Supernatants were placed in fresh Eppendorf tubes and the total protein concentration was determined by the Bradford method [37] using a Bio-Rad Laboratories kit and sigma bovine gamma globulin as a protein standard. Activity of Acs protein in lysate extracts was quantified using an NADH-consuming coupled assay [38]. Four micrograms of protein was added to a final volume of 100 μl reactions that contain the following: HEPES buffer (50 mM, pH 7 at 25ºC), ATP (2.5 mM), coenzyme A (CoASH, 0.5 mM), TCEP (1 mM), $MgCl_2$ (5 mM), phosphoenolpyruvate (3 mM), pyruvate kinase (1 U), myokinase (5 U), lactate dehydrogenase (1.5 U), and sodium acetate (0.2 mM). All reactions were started with the addition of NADH (0.1 mM). Negative controls were reactions

without NADH added. The absorbance of NADH was monitored at 340 nm every 30 s for 8 min in a 96-well plate using a Spectramax Plus UV-visible spectrophotometer (Molecular Devices). Acs activity in lysates was tested in biological triplicate and each measurement was determined in nine technical replicates. Specific activity of Acs was calculated as described [39].

Western blot analysis. For supplemental figure 2A, cell-free extracts of strains JE23602 and JE23603 used in acetyl-CoA synthetase activity assays were loaded on 12% SDS-PAGE gels at a concentration of 20, 30, and 40 μg each along with pure Acs protein loaded at 0.125, 0.25, 0.5, 1, 2, and 4 μg. After polyacrylamide gels electrophoresed at 200 V for 45 min, proteins in the gels were transferred to 0.45 μm PVDF membranes (Immobilon® – P membranes from Millipore) using a Bio-Rad Trans-Blot® Turbo™ transfer system at 25 V, 1.0 A for 30 min in 1X Transfer buffer (Tris-HCl (25 mM), glycine (192 mM), and 10% methanol (v/v). Blots were then blocked with phosphate buffer saline-tween-20 (PBST) buffer with 5% (w/v) dehydrated milk powder for one hour at room temperature, then incubated with 1:500 of anti-Acs polyclonal rabbit antibodies (generated by Envigo,PA). Blots were washed thrice with PBST buffer, then incubated for one hour in PBST-5% milk buffer with 1:10,000 anti-rabbit IgG (whole molecule) peroxidase conjugate (Sigma). The blot was then washed thrice with PBST and developed using Thermo Scientific SuperSignal™ West Pico Plus chemiluminescent substrate and imaged using a UVP ChemStudio imaging system from Analytik Jena. Standard curves of the densitometry of pixel bands

were generated for each developed western band for every gel run, and unknown Acs protein concentration was calculated using the equation generated by each curve. Every lysate used in acetyl-CoA synthetase activity assays was tested twice per strain per lysate for anti-Acs antibody determined protein concentration. Average Acs protein concentration as determined by quantitative western blotting is depicted in figure 8B, as well as an example developed blot, standard curve, and accompanying SDS-PAGE gel (supplemental Fig. 4.11).

For figure 9ABC and supplemental figure 5ABCD, strains JE22070, JE25564, and JE25565 were grown in either 1 liter each of minimal medium supplemented with 10 mM acetate and 500 μM of L-(+)-arabinose or were grown in 200 mL of LB supplemented with 1 mM of L-(+)-arabinose. Acetate grown cells were harvested during mid-logarithmic phase, while LB grown cells were harvested at OD_{600} of 0.2, 0.5, and 0.7. Cells were processed and extracts were quantified as described above. 50 μg of cell-free extractsof LB-grown cells or 200 μg of cell-free extracts of acetate-grown cells were resolved by SDS-PAGE on 15% (w/v) polyacrylamide gels. Western blot transfers and washes were conducted as described above except the primary antibodies were anti-CobB polyclonal rabbit antibodies (generated by Envigo, PA) at 1:500 or as an internal protein concentration standard, anti-DnaK antibodies [8E2/2] from Abcam (cat# ab69617) at a dilution of 1:10,000 in 5% milk PBST buffer. Blots containing anti-CobB primary antibodies were washed three times in 5% milk-phosphate buffer saline Tween 20 (PBST) buffer and were incubated for one hour in 1:10,000 of secondary antibody being goat anti-rabbit IgG peroxidase conjugate (Sigma). Secondary antibodies used for

anti-DnaK treated blots were goat anti-mouse IgG (H+L) (cat# 31432, Invitrogen) at a concentration of 1:5,000. After three washes in PBST buffer, blots were developed as described above.

Average Acs, CobB isoform, or DnaK protein concentrations from lysates were calculated using the pixel densitometry calculator of the UVP ChemStudio imaging system of Analytik Jena. The densitometry of pixel bands was calculated from technical duplicates of biological triplicates of each lysate. The volume of pixel density was plotted using Prism 8 by GraphPad and one-way ANOVA was used to compare pixel densities across samples to generate P-values.

Mass spectrometry. *Enzymatic "In Gel" Digestion.* "In gel" digestion and mass spectrometric analysis was done at the Mass Spectrometry Facility [Biotechnology Center, University of Wisconsin-Madison]. Briefly, excised gel pieces were washed twice for two min in MeOH:H_20:NH_4HCO_3 [50% (v/v):50% (v/v):100 mM], dehydrated for two min in acetonitrile (ACN):H_20:NH_4HCO_3 [50% (v/v):50% (v/v):25 mM] then once more for 30 s in 100% ACN, dried in a Speed-Vac for one min, rehydrated completely and reduced in dithiothreitol (DTT, 25 mM) in NH_4HCO_3 (25 mM)] for 30 min at 56 °C, alkylated by solution exchange with iodoacetamide (IAA, 55mM) dissolved in NH_4HCO_3 (25mM) in darkness at room temperature for 30 min, washed once with NH_4HCO_3 (25mM) dehydrated twice for two min in ACN:H_20:NH_4HCO_3 [50% (v/v):50% (v/v):25 mM] then once more for 30 s in 100% ACN, dried in a Speed-Vac again and finally rehydrated with 20 μl of trypsin solution [10 ng/μl trypsin (Promega) in NH_4HCO_3 (25mM) and ProteaseMAX

(0.01% w/v) (Promega)]. Additional 30 μl of digestion solution [NH_4HCO_3, 25mM + and ProteaseMAX, 0.01% w/v (Promega)] was added to facilitate complete rehydration with excess overlay needed for peptide extraction. The digestion was conducted for 3 h at 42°C. Peptides generated from digestion were transferred to a new tube and acidified with trifluoroacetic acid [(TFA, 2.5% (v/v)] to 0.3% v/v final. Degraded ProteaseMAX was removed via centrifugation [max speed, 10 min] and the peptides solid phase extracted (*ZipTip® C18* pipette tips Millipore, Billerica, MA).

NanoLC-MS/MS. Peptides were analyzed by nanoLC-MS/MS using the Agilent 1100 nanoflow system (Agilent) connected to a hybrid linear ion trap-orbitrap mass spectrometer (LTQ-Orbitrap Elite™, Thermo Fisher Scientific) equipped with an EASY-Spray™ electrospray source. Chromatography of peptides prior to mass spectral analysis was accomplished using capillary emitter column (PepMap® C18, 3 μM, 100 Å, 150 x 0.075 mm, Thermo Fisher Scientific) onto which two μl of extracted peptides was automatically loaded. NanoHPLC system delivered solvents A: formic acid (0.1% (v/v), and B: 99.9% (v/v) acetonitrile, 0.1% (v/v) formic acid at 0.50 μL/min to load the peptides (over a 30 min period) and 0.3 μl/min to elute peptides directly into the nano-electrospray with gradual gradient from 3% (v/v) B to 20% (v/v) B over 17 min, followed by five min fast gradient from 20% (v/v) B to 50% (v/v) B and concluded with four min ramp to 95% (v/v) B at which time a one min flush-out took place. As peptides eluted from the HPLC-column/electrospray source survey MS scans were acquired in the Orbitrap with a

resolution of 120,000 followed by MS2 fragmentation of 20 most intense peptides detected in the MS1 scan from 350 to 1800 m/z; redundancy was limited by dynamic exclusion.

Data analysis. Raw MS/MS data was converted to mgf file format using MSConvert (ProteoWizard: Open Source Software for Rapid Proteomics Tools Development). Resulting mgf files were used to search against user defined amino acid sequence database containing *Salmonella enterica* CobB sequences (long and short variants) along with a list of common contaminants using in-house *Mascot* search engine 2.2.07 [Matrix Science] with variable Lysine and protein *N*-terminus acetylation, methionine oxidation, asparagine and glutamine deamidation plus fixed cysteine carbamidomethylation. Peptide mass tolerance was set at 15 ppm and fragment mass at 0.6 Da. All the significant identifications based on the ion scores were manually interrogated to confirm algorithmic assignments. Extracted ion chromatograms were generated to evaluate abundances between samples and raw MS2 files inspected for correctness of identification.

4.5 RESULTS

YiaC acetylates the long isoform of CobB, but not the short isoform. A search for protein substrates for the *S. enterica* putative GCN5-related acetyltransferases led us to discover that the *S. enterica* YiaC protein acetylated *S. enterica* $CobB_L$, but not $CobB_S$. As shown in figure 4.2, when both isoforms of *S. enterica* CobB were incubated with [1-^{14}C]-AcCoA as a function of YiaC, radiolabel was

transferred to $CobB_L$, but not to $CobB_S$ (Fig. 4.2, lanes 4 and 5, respectively). Since $CobB_S$ was not acetylated, we surmised that the site(s) of acetylation were located within the 37-aa *N*-terminal, arginine-rich motif of $CobB_L$ (Fig. 4.1).

YiaC does not acetylate *N*ε amino groups of lysine residues. The *N*-terminal extension of $CobB_L$ contains two lysines (K14, K16), which we investigated as possible acetylation sites. To test this possibility, we changed K14 and K16 to alanine, independently and in combination. The three variants ($CobB_L^{K14A}$, $CobB_L^{K16A}$, and $CobB_L^{K14A,K16A}$) were overproduced, isolated, and incubated with YiaC in the presence of [1-^{14}C]-AcCoA. Surprisingly, YiaC acetylated all three $CobB_L$ variants (Fig. 4.3, lanes 5, 7, 9), suggesting that, under the assay conditions used, YiaC did not modify the *N*ε position of either K14 or K16. We note that the intensity of the signal for acetylated $CobB_L^{K14A,K16A}$ variant was less than the single-aa variants and that the signal intensity was commensurate to the amount of protein loaded on the gel. The yield of $CobB_L^{K14A,K16A}$ variant was lower than those of the single-aa variants.

YiaC modifies the *N*-terminus of $CobB_L$ *in vitro.* To support our hypothesis that $CobB_L$ is *N*α-acetylated, liquid chromatography-tandem mass spectrometry (LC-MS/MS) was conducted. Results of peptide fingerprinting analysis of acetylated $CobB_L$ ($^{Ac}CobB_L$) unequivocally showed that the *N*-terminus of $CobB_L$ was acetylated by YiaC (Fig. 4.4 A, B).

To confirm the LC-MS/MS data, two peptides of the first 50-amino acids of $CobB_L$ were synthesized (Peptide 2.0, VA), one started with unmodified L-Met, the second one started with L-Met^{Ac}. *In vitro* acetylation assays were performed with the above-mentioned peptides as

substrates for YiaC. When [1-^{14}C]-AcCoA was added to the reaction mixture, YiaC acetylated the $CobB_L$ peptides whose *N*-terminal amino group was not modified (Fig. 5, lanes 2, 8), but did not acetylate the $CobB_L$ peptide whose first residue was L-Met^{Ac} (Fig. 4.5, lane 5). Collectively, these data showed that YiaC was a *N*α– protein acetyltransferase (NAT).

***N*-terminally acetylated $CobB_L$ ($^{Ac}CobB_L$) cannot be deacetylated by either isoform of CobB**. To test whether *N*-terminally acetylated CobB could be deacetylated by either form of CobB, acetylation assays with YiaC, $CobB_L$, and [1-^{14}C]-AcCoA were conducted. Radiolabeled, acetylated $CobB_L$ ($^{*Ac}CobB_L$) samples were freed of excess acetyl-CoA and were incubated with either $CobB_L$ or $CobB_S$, with or without NAD^+. Positive controls were also added to show $CobB_L$ and $CobB_S$ deacetylated its *bona fide* substrate, Acs^{Ac}. As seen in figure 4.6, $CobB_L$ and $CobB_S$ deacetylated Acs^{*Ac} (lanes 3 and 5), but did not to deacetylate $^{*Ac}CobB_L$ (lanes 7 and 9). This showed that *N*-terminal acetylation of $CobB_L$ was not reversed by either CobB isoform.

In vitro* and *in vivo* evidence that *N*α acetylation of $CobB_L$ alters its deacetylase activity.** <u>In vitro *evidence*</u>.*** The *N*-terminal acetylation of $CobB_L$ by

YiaC (Fig. 4.2) raised questions regarding the effect of the modification on the enzymatic activity of $CobB_L$. To answer this question, we performed *in vitro* protein deacetylation assays with radiolabeled, acetylated Acs (Acs^{*Ac}) and radiolabeled, *N*-terminally acetylated $CobB_L$ ($^{*Ac}CobB_L$) to enhance our ability to detect deacetylation events. The protocols for the generation of Acs^{*Ac} and $^{*Ac}CobB_L$ are described in the *Materials and Methods* section. Once we had radiolabeled Acs^{*Ac} and $^{*Ac}CobB_L$, the following experiment was performed. $CobB_L$ or $^{*Ac}CobB_L$ was added to a reaction mixture that contained Acs^{*Ac} and NAD^+, and samples (25 μl ea) were incubated for one hour. As seen in figure 4.7A, unacetylated $CobB_L$ deacetylated Acs^{*Ac} as indicated by the disappearance of the signal in the phosphor image corresponding to Acs^{*Ac} (molecular mass ~72 kDa) (Fig. 4.7A, lane 5). In contrast, $^{*Ac}CobB_L$ did not deacetylate Acs^{*Ac} as efficiently as $CobB_L$ (Fig. 4.7A, lanes 3 *vs* 5). Quantitative densitometry of the signals of Acs^{Ac} radiolabel disappearance showed an average of ~50% reduction in $CobB_L$ deacetylase activity when acetylated. The results shown in figure 4.7 are representative of a set of six separate experiments. These data showed that the enzymatic activity of $^{*Ac}CobB_L$ was negatively affected by the modification.

In vivo *evidence*. To verify that acetylation of the *N*-terminus of $CobB_L$ affected its enzymatic activity, we performed *in vivo* experiments to assess the activity of $CobB_L$ as a function of YiaC during growth on minimal medium containing a low concentration of acetate (10 mM). It is known that, under such conditions, CobB function is required to maintain Acs deacetylated, hence active [3, 22, 40]. For this purpose, we moved two plasmids into a *S. enterica* Δ*cobB* strain. One plasmid

encoded either $CobB_L^{(M37A, M38A)}$ or $CobB_S^{(M1A)}$, the second plasmid encoded YiaC or the vector control. Genes *cobB* and *yiaC* were under the control of the L-(+)-arabinose-inducible promoter P_{araBAD} [41]. To get only $CobB_L$ and no $CobB_S$ protein, the plasmid encoding $CobB_L$ contained the natural starting methionine (M1) for $CobB_L$, but had two mutations that change the starting methionine and neighboring methionine for $CobB_S$ (*i.e.,* M37 and M38) to alanines. The *cobB* allele encoding $CobB^{M37A,M38A}$ effectively blocks the synthesis of the short CobB isoform [23]. Conversely, for the cell to synthesize only $CobB_S$, the $CobB_S$-encoding plasmid has a *cobB* allele with the first codon changed to encode alanine so that $CobB_L$ protein cannot be made from M1 resulting in the exclusive synthesis of $CobB_S$ starting at position M38 [23]. The growth behaviors of strains □*cobB* and *cobB*$^+$ harboring empty cloning vectors were used as controls. An additional control used a *cobB*$^+$ strain harboring a plasmid carrying *yiaC*$^+$ (open circles). All strains were grown on minimal medium with acetate (10 mM) as the sole carbon and energy source.

Data presented in figure 4.7B show the growth behavior of strains of interest in the presence of inducer (L-(+)-arabinose, 100 μM). The following observations were made: i) as expected, the Δ*cobB* strain failed to grow on 10 mM acetate (black triangles) because Acs remained acetylated, hence inactive; ii) the phenotype of the Δ*cobB* strain was corrected by ectopic expression of *cobB* alleles that directed the synthesis of functional $CobB_S$ (open squares) or $CobB_L$ (black squares); iii) synthesis of YiaC did not affect the growth of the *cobB*$^+$ strain (open circles *vs* gray squares); iv) synthesis of $CobB_S$ and YiaC by a strain with a genomic deletion of

cobB resulted in a growth behavior that was very similar to that of the *cobB*$^+$ strain that synthesized YiaC (open triangles *vs* gray squares and open circles); and v) synthesis of $CobB_L$ and YiaC by the Δ*cobB* strain prematurely arrested growth (gray circles). These results were consistent with our *in vitro* data, which showed that *N*-terminal acetylation of $CobB_L$ by YiaC had a negative effect on the deacetylase activity of $CobB_L$ (Fig. 4.7A). Again, under the growth conditions used, a reduction in deacetylase activity would prevent the deacetylation (hence re-activation) of Acs, blocking the conversion of acetate into acetyl-CoA with the concomitant negative effect on cell growth.

As predicted by the above results, the specific activity of Acs was different in the Δ*cobB* strain that ectopically synthesized $CobB_L$ and YiaC, and the Δ*cobB* strain synthesized $CobB_L$ but no YiaC ectopically overexpressed (Fig 4.8A). In the above-mentioned strains, we measured a statistically significant reduction of Acs activity when YiaC was over produced. These results were consistent with the data showing that YiaC-dependent acetylation of the $CobB_L$ *N*-terminus reduced its activity, maintaining Acs acetylated, (hence inactive) thus arresting growth due to reduced levels of AcCoA.

To verify that the concentration of Acs was the same in both strains, quantitative western blots using rabbit polyclonal anti-Acs antibodies were performed to determine the amount of Acs protein present in lysates used to assay for acetyl-CoA synthetase activity. Indeed, lysates of strains that either overexpressed *yiaC*$^+$ or carried the empty vector, contained the same amount of Acs (Fig. 4.8B, 4.11 A,B, and C).

Acetylation of $CobB_L$ does not cause $CobB_L$ degradation *in vivo.* In Eukaryotes, some *N*-terminally acetylated proteins are targeted for degradation [42]. Other examples suggest that *N*-terminal acetylation stabilizes acetylated proteins and prevents degradation [43, 44]. To assess whether acetylation of $CobB_L$ altered $CobB_L$ protein levels *in vivo*, we used rabbit polyclonal anti-CobB antibodies to quantify levels of CobB in cell-free extracts of three strains: *yiaC*$^+$ / vector, *yiaC*::*kan*$^+$ / vector, and *yiaC*::*kan*$^+$ / pYiaC. Strains were grown in minimal medium supplemented with acetate (10 mM) and with L-(+)-arabinose (500 μM), or in rich medium (LB) containing L-(+)-arabinose (1 mM). In both cases, growth was monitored at 600 nm. Data presented in figure 4.8A and 4.8B show that during growth in rich medium or in minimal medium with a low concentration of acetate, the concentration of $CobB_L$ did not vary in any of the strains. Pure protein controls of $CobB_L$ and $CobB_S$ were added as positive controls and an anti-DnaK blot was included to ensure that all samples were loaded equally. Statistical analysis of pixel density of the anti-CobB western blots shows no difference in $CobB_L$ concentration in any of the strains tested, as well as no difference in $CobB_S$ or the internal control protein DnaK (Figure 4.9C, 4.14).

4.6 DISCUSSION

In *S. enterica*, the YiaC protein is an *N*α acetyltransferase that controls the activity of the long isoform of the CobB sirtuin deacylase. Data reported here support several conclusions regarding the function of the *S. enterica* YiaC protein.

Firstly, YiaC has *N*α acetyltransferase activity. Our data also show that YiaC modifies the long isoform of the CobB sirtuin deacylase ($CobB_L$) of this bacterium, but not the short isoform of CobB (Fig. 4.2). These results placed the sites of acetylation within the 37-aa, *N*-terminal extension of $CobB_L$ (Fig. 4.1). Secondly, YiaC does not modify *N*ε amino groups of lysine side chains of the $CobB_L$ protein (Figs. 4.3, 4.4). Whether YiaC can modify lysyl residues of other proteins remains to be determined. Thirdly, YiaC appears to have somewhat broad specificity for its target, since it acetylated the *N*α amino group of the glycyl residue that remained fused to the protein after protease treatment to remove the MBP-H_6 tag (Figs. 4.2, 4.4). YiaC also acetylated the *N*α amino group of the *N*-terminal methionine of the *C*-terminally H_6 tagged $CobB_L$, and the *N*-terminal methionine of a synthetic peptide comprised of the first 50 amino acids of $CobB_L$ (Figs. 4.5, 4.13).

We note that repeated attempts to isolate *N*-terminally acetylated $CobB_L$ from cells were unsuccessful, in spite of the fact that several different proteases were used in combination prior to mass spectrometry analysis. We posit that the amino acid composition of the *N*-terminus of $CobB_L$ makes this analysis difficult because it is so rich in arginines. However, when we removed the *N*-terminal tag from MBP-H_6-$CobB_L$, the resulting protein had two additional residues on its *N*-terminus, namely Gly-Thr (GT-$CobB_L$). YiaC acetylated GT-$CobB_L$ *in vitro* (Figs. 4.4, S4), but did not acetylate a protein that had an additional methionine residue, *i.e.,* MGT-$CobB_L$ (data not shown) or the *N*-terminally H_6-$CobB_L$ protein (Fig. 4.13).

Collectively, our results show that YiaC is an *N*α acetyltransferase (NAT) and that the $CobB_L$ isoform is a substrate of it. To the best of our knowledge, YiaC is a

bacterial NAT that does not belong to the Rim family of proteins. Based on our data, we propose to change the name of YiaC to NatA, to reflect the fact that it is a *N*α acetyltransferase.

Can YiaC acetylate the epsilon amino group of lysine side chains? Recently, Christensen et. al. showed that overexpression of *yiaC* in an *Escherichia coli pta patZ acs cobB* deletion strain displayed increased protein acetylation measured by western blot analysis using α-AcK antibodies [45]. These results are interesting because these investigators conducted AcK enrichment and mass spectrometry to identify putative YiaC protein lysine targets. CobB was not identified from the aforementioned proteome because the *cobB* gene was deleted in the strains used, and *N*-terminal acetylation of proteins was not reported. As mentioned above, at present we cannot rule out the possibility that YiaC acetylates lysyl residues as suggested by Christensen *et. al*. Additional work is needed to determine whether YiaC can perform *N*α and *N*ε protein acetylation.

The newly identified activity of YiaC as an *N*-terminal acetyltransferase raises questions about the role of *N*-terminal acetylation in prokaryotic cell physiology.

Concluding remarks. We have shown that *N*-terminal acetylation of $CobB_L$ occurs both *in vitro* and that YiaC-dependent acetylation of $CobB_L$ negatively impacts the deacetylase activity of the enzyme, which in turn negatively affects growth on acetate. How the addition of an acetyl-group to the *N*-terminus of $CobB_L$ impacts its activity is a question of interest, and ongoing studies in our laboratory are

focused on answering this question. To our knowledge, YiaC is the first NAT that does not belong to the Rim protein family of *N*α amino group acetyltransferases. Also, this is the first report of acylation of a prokaryotic sirtuin deacylase. As pointed above, based on the evidence reported here, we propose to change the name of YiaC to NatA (for *N*α-acetyltransferase) to reflect the biochemical activity of the enzyme.

4.7 ACKNOWLEDGEMENTS

We thank Grzegorz Sabat from the Biotechnology Center of The University of Wisconsin-Madison for the performance of the LC/MS/MS analysis. Rachel Burckhardt first observed CobB acetylation.

4.8 CONFLICT OF INTEREST

The authors do not have any conflict of interest to declare.

4.9 FUNDING

This work was supported by NIH grant R35 GM130399 to J.C.E.-S.

4.10 REFERENCES

1. Hentchel, K.L. and J.C. Escalante-Semerena, *Acylation of biomolecules in prokaryotes: a widespread strategy for the control of biological function and metabolic stress.* Microbiol. Mol. Biol. Rev., 2015. **79**: p. 321-346.

2. VanDrisse, C.M. and J.C. Escalante-Semerena, *Protein Acetylation in Bacteria.* Annu. Rev. Microbiol., 2019. **73**: p. 111-132.

3. Starai, V.J. and J.C. Escalante-Semerena, *Identification of the protein acetyltransferase (Pat) enzyme that acetylates acetyl-CoA synthetase in Salmonella enterica.* J. Mol. Biol., 2004. **340**(5): p. 1005-1012.

4. Gardner, J.G. and J.C. Escalante-Semerena, *Biochemical and mutational analyses of AcuA, the acetyltransferase enzyme that controls the activity of the acetyl coenzyme a synthetase (AcsA) in Bacillus subtilis.* J. Bacteriol., 2008. **190**: p. 5132-5136.

5. Takakura, H., et al., *NH2-terminal acetylation of ribosomal proteins of Saccharomyces cerevisiae.* J. Biol. Chem., 1992. **267**: p. 5442-5445.

6. Yoshikawa, A., et al., *Cloning and nucleotide sequencing of the genes rimI and rimJ which encode enzymes acetylating ribosomal proteins S18 and S5 of Escherichia coli K12.* Mol. Gen. Genet., 1987. **209**: p. 481-488.

7. Walsh, C., *Protein N-acetylation*, in *Posttranslational modification of proteins*. 2006, Roberts & Company Publishers: Greenwood Village, CO. p. 151-170.

8. Kentache, T., et al., *Proteomic characterization of Nalpha- and Nepsilon-acetylation in Acinetobacter baumannii.* J Proteomics, 2016. **144**.

9. Nguyen, K.T., et al., *N-terminal acetylation and the N-end rule pathway control degradation of the lipid droplet protein PLIN2.* J. Biol. Chem., 2019. **294**: p. 379-388.

10. Nguyen, K.T., et al., *Control of protein degradation by N-terminal acetylation and the N-end rule pathway.* Exp. Mol. Med., 2018. **50**: p. 91.

11. Thompson, C.R., M.M. Champion, and P.A. Champion, *Quantitative N-terminal footprinting of pathogenic mycobacteria reveals differential protein acetylation.* J. Proteome Res., 2018. **17**: p. 3246-3258.

12. Ree, R., S. Varland, and T. Arnesen, *Spotlight on protein N-terminal acetylation.* Exp. Mol. Med., 2018. **50**: p. 90.

13. Yaguchi, M., et al., *Cooperative control of translational fidelity by ribosomal proteins in Escherichia coli. II. Localization of amino acid replacements in proteins S5 and S12 altered in double mutants resistant to neamine.* Mol. Gen. Genet., 1975. **142**: p. 35-43.

14. Wittmann-Liebold, B. and B. Greuer, *The primary structure of protein S5 from the small subunit of the Escherichia coli ribosome.* FEBS Lett., 1978. **95**(1): p. 91-98.

15. Terhorst, C., B. Wittmann-Liebold, and W. Moller, *50-S ribosomal proteins. Peptide studies on two acidic proteins, A 1 and A 2 , isolated from 50-S ribosomes of Escherichia coli.* Eur. J. Biochem., 1972. **25**: p. 13-19.

16. Pathak, D., et al., *Biochemical evidence for relaxed substrate specificity of Nalpha-acetyltransferase (Rv3420c/rimI) of Mycobacterium tuberculosis.* Sci. Rep., 2016. **6**: p. 28892.

17. Smith, V.F., et al., *Electrospray mass spectrometric investigation of the chaperone SecB.* Protein Sci., 1996. **5**: p. 488-494.

18. Okkels, L.M., et al., *CFP10 discriminates between nonacetylated and acetylated ESAT-6 of Mycobacterium tuberculosis by differential interaction.* Proteomics, 2004. **4**: p. 2954-2960.

19. Mediero, A., M. Perez-Aso, and B.N. Cronstein, *Activation of EPAC1/2 is essential for osteoclast formation by modulating NFkappaB nuclear translocation and actin cytoskeleton rearrangements.* FASEB J., 2014. **28**: p. 4901-4913.

20. Ouidir, T., et al., *Characterization of N-terminal protein modifications in Pseudomonas aeruginosa PA14.* J. Proteomics, 2015. **114**: p. 214-225.

21. Starai, V.J., et al., *Sir2-dependent activation of acetyl-CoA synthetase by deacetylation of active lysine.* Science, 2002. **298**: p. 2390-2392.

22. Starai, V.J., et al., *Short-chain fatty acid activation by acyl-coenzyme A synthetases requires SIR2 protein function in Salmonella enterica and Saccharomyces cerevisiae.* Genetics, 2003. **163**: p. 545-55.

23. Tucker, A.C. and J.C. Escalante-Semerena, *Biologically active isoforms of CobB sirtuin deacetylase in Salmonella enterica and Erwinia amylovora.* J. Bacteriol., 2010. **192**: p. 6200-6208.

24. Datsenko, K.A. and B.L. Wanner, *One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products.* Proc. Natl. Acad. Sci. USA, 2000. **97**: p. 6640-6645.

25. Davis, R.W., D. Botstein, and J.R. Roth, *A manual for genetic engineering: advanced bacterial genetics.* 1980, Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

26. Calvin, N.M. and P.C. Hanawalt, *High-efficiency transformation of bacterial cells by electroporation.* J. Bacteriol., 1988. **170**: p. 2796-2801.

27. VanDrisse, C.M. and J.C. Escalante-Semerena, *New high-cloning-efficiency vectors for complementation studies and recombinant protein overproduction in Escherichia coli and Salmonella enterica.* Plasmid, 2016. **86**: p. 1-6.

28. Rocco, C.J., et al., *Construction and use of new cloning vectors for the rapid isolation of recombinant proteins from Escherichia coli.* Plasmid, 2008. **59**: p. 231-237.

29. Berkowitz, D., et al., *Procedure for identifying nonsense mutations.* J. Bacteriol., 1968. **96**: p. 215-220.

30. Balch, W.E. and R.S. Wolfe, *New approach to the cultivation of methanogenic bacteria: 2-mercaptoethanesulfonic acid (HS-CoM)-dependent growth of Methanobacterium ruminantium in a pressurized atmosphere.* Appl. Environ. Microbiol., 1976. **32**: p. 781-791.

31. Elbing, K. and R. Brent, *Media preparation and bacteriological tools*, in *Curr. Prot. Molec. Biol.*, F.M. Ausubel, et al., Editors. 2002, Wiley Interscience: New York. p. Unit 1.1.3.

32. Laemmli, U.K., *Cleavage of structural proteins during the assembly of the head of bacteriophage T4.* Nature, 1970. **227**: p. 680-685.

33. Blommel, P.G., et al., *Enhanced bacterial protein expression during auto-induction obtained by alteration of lac repressor dosage and medium composition.* Biotechnol. Prog., 2007. **23**: p. 585-598.

34. Bertani, G., *Studies on lysogenesis. I. The mode of phage liberation by lysogenic Escherichia coli.* J. Bacteriol., 1951. **62**: p. 293-300.

35. Bertani, G., *Lysogeny at mid-twentieth century: P1, P2, and other experimental systems.* J. Bacteriol., 2004. **186**: p. 595-600.

36. Blommel, P.G. and B.G. Fox, *A combined approach to improving large-scale production of tobacco etch virus protease.* Protein Expr. Purif., 2007. **55**: p. 53-68.

37. Bradford, M.M., *A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding.* Anal. Biochem., 1976. **72**: p. 248-254.

38. Reger, A.S., J.M. Carney, and A.M. Gulick, *Biochemical and crystallographic analysis of substrate binding and conformational changes in acetyl-CoA synthetase.* Biochemistry, 2007. **46**: p. 6536-6546.

39. Garrity, J., et al., *N-lysine propionylation controls the activity of propionyl-CoA synthetase.* J. Biol. Chem., 2007. **282**: p. 30239-30245.

40. Chan, C.H., et al., *In Salmonella enterica, the sirtuin-dependent protein acylation/deacylation system (SDPADS) maintains energy homeostasis during growth on low concentrations of acetate.* Mol. Microbiol., 2011. **80**: p. 168-183.

41. Guzman, L.M., et al., *Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter.* J Bacteriol, 1995. **177**(14): p. 4121-30.

42. Hwang, C.S., A. Shemorry, and A. Varshavsky, *N-terminal acetylation of cellular proteins creates specific degradation signals.* Science, 2010. **327**: p. 973-977.

43. Jornvall, H., *Acetylation of protein N-terminal amino groups structural observations on alpha-amino acetylated proteins.* J. Theor. Biol., 1975. **55**: p. 1-12.

44. Persson, B., et al., *Structures of N-terminally acetylated proteins.* Eur. J. Biochem., 1985. **152**: p. 523-527.

45. Christensen, D.G., et al., *Identification of novel protein lysine acetyltransferases in Escherichia coli.* MBio, 2018. **9**: p. e01905-18.

Table 4.1. Strains and plasmids used in this study.		
Strain	**Relevant genotype**	**Reference/source[a]**
***S. enterica enterica* serovar Typhimurium LT2 strains**		
JE6583	*metE205 ara-9*	Laboratory strain
Derivatives of JE6583		
JE22352	*cobB*$^{+}$ / pCV1 / pCV3	
JE23752	*cobB*$^{+}$ / pYiaC10 / pCV3	This work
JE23602	Δ*cobB1375* / pCV1 / pCOBB137	This work
JE23603	Δ*cobB1375* / pYiaC10 / pCOBB137	This work
JE23604	Δ*cobB1375* / pCV1 / pCOBB136	This work
JE23605	Δ*cobB1375* / pYiaC10 / pCOBB136	This work
JE23608	Δ*cobB1375* / pCV1 / pCV3	This work
JE25564	*yiaC*432::*kan*$^{+}$ / pCV1	
JE25565	*yiaC*432::*kan*$^{+}$ / pYiaC10	
***E. coli* strains**		
E. coli C41 (λDE3)	*pka12*::*kan*$^{+}$ *ompT hsdS* (r_Bm_B) *gal* λ (DE3)	Laboratory collection

Table 4.2. Plasmids used in this study

Plasmid	Genotype	Description	Source
pCV1	*araC*$^+$ *bla*$^+$	P_{araBAD} expression vector	[27]
pCV3	*araC*$^+$ *cat*$^+$	P_{araBAD} expression vector	[27]
pTEV6	*bla*$^+$	*N*-terminal rTEV-cleavable MBP-His fusion overexpression vector	[28]
pET-23a	*bla*$^+$	*C*-terminal 6xHis fusion overexpression vector	Novagen
pTEV5	*bla*$^+$	*N*-terminal 6xHis fusion overexpression vector	[28]
pCOBB72	*cobB*$^+$ *bla*$^+$	*cobB*$^+$ (encodes $CobB_L$ only, native sequence) cloned into in pTEV6, Ap^r	[23]
pCOBB71	*cobB*$^+$ *bla*$^+$	*cobB1373* (encodes $CobB_S$) *cloned* into in pTEV6, Ap^r	[23]
pCOBB48	*cobB*$^+$ *bla*$^+$	*cobB1372* (encodes $CobB_L^{M37A, M38A}$) in pET-23a Ap^r	
pCOBB50	*cobB*$^+$ *bla*$^+$	*cobB1372* (encodes $CobB_L^{M37A, M38A}$) in pTEV5 Ap^r	
pCOBB151	*cobB*$^+$ *bla*$^+$	*cobB1376* (encodes $CobB_L^{K14A}$) in pTEV6, Ap^r	This work
pCOBB152	*cobB*$^+$ *bla*$^+$	*cobB1377* (encodes $CobB_L^{K16A}$) in pTEV6, Ap^r	This work
pCOBB153	*cobB*$^+$ *bla*$^+$	*cobB1378* (encodes $CobB_L^{K14A\ K16A}$) in pTEV6, Ap^r	This work
pCOBB136	*cobB*$^+$ *cat*$^+$	*cobB1373* (encodes $CobB^{M1A}$) in pCV3, Cm^r	This work
pCOBB137	*cobB*$^+$ *cat*$^+$	*cobB1372* (encodes $CobB_L^{M37A, M38A}$) in pCV3, Cm^r	This work
pYiaC10	*yiaC*$^+$ *bla*$^+$	*yiaC*$^+$ cloned into pCV1, Ap^r	This work

Table 4.3. Primers used in this study.	
Primer Name	**Primer Sequence 5' → 3'**
5' *cobB* Wanner	GGTGCTGCTTTTTTACATCTTACCGACTAATCAAAAAAAGAGGTTGTTATGGTGTAGG
3' *cobB* Wanner	ACGTAACGTGAAATGTAGGCCGGATAAGGCGTTACCGGGCAAACAGCACTACATATGAATATCCTCCTTAG
5' pCOBB71 KpnI	ATAGGTACCATGGAAAACCCAAGAGTATTAGTCC
3' pCOBB71 SalI	GATGTCGACCAGCACTACAGCCCTTTCAGG
5' pCOBB72 KpnI	GTAGGTACCATGCAGTCGCGTCGGTTTCATCG
3' pCOBB72 SalI	GATGTCGACCAGCACTACAGCCCTTTCAGGCAGGGTCAGAATCACCT
5' pCOBB48 XbaI	NNTCTAGAATGGAAAACCCAAGAGTATTAGTCC
3' pCOBB48 XhoI	NNNCTCGAGCAGCCCTTTCAGGAATTTATCAA
5' pCOBB50 NheI	GTAGCTAGCATGCAGTCGCGTCGGTTTCATCG
3' pCOBB50 XhoI	CATCTCGAGCTACAGCCCTTTCAGGAATTTATC
5' pCOBB151 and pCOBB153 quick change primer 1	TCACGCAACAGACGTTTATTTGCACGAAAGCGGCTTAATCGATG
5' pCOBB151 and pCOBB153 quick change primer 2	CATCGATTAAGCCGCTTTCGTGCAAATAAACGTCTGTTGCGTGA
5' pCOBB152 and pCOBB153 quick change primer 1	CACGCAACAGACGTGCATTTTTACGAAAGCGGCTTAATCGATG
5' pCOBB152 and pCOBB153 quick change primer 2	CATCGATTAAGCCGCTTTCGTAAAAATGCACGTCTGTTGCGTG
5' pCOBB136 BspQI	NNGCTCTTCNTTCATGGAAAACCCAAGAGTATTAGTCC
5' pCOBB137 BspQI	NNGCTCTTCNTTCATGCAGTCGCGTCGGTTTCA
3' pCOBB136 and pCOBB137 BspQI	NNGCTCTTCNTTACTACAGCCCTTTCAGGAATTTATCAA
5' pYiaC10 BspQI	NNGCTCTTCNTTCATGATTCGCAAATCCCAGAGTGAAGAC
3' pYiaC10 BspQI	NNGCTCTTCNTTATTACGGCGTTTGATCCGCCTGCCAAC

Table 4.4. Homologues of YiaC and CobB isoforms			
Genus species	YiaC homologue	CobB long homologue	CobB short homologue
Klebsiella pneumoniae	Yes	Yes	Yes
Citrobacter koseri	Yes	Yes	Yes
Escherichia coli	Yes	Yes	Yes
Edwardsiella ictaluri	Yes	Yes	Yes
Sodalis glossinidius	Yes	Yes	Yes
Proteus mirabilis	Yes	Yes	Yes
Shigella flexneri	Yes	Yes	Yes
Serratia proteamaculans	Yes	Yes	Yes
Enterobacter sp 638	Yes	Yes	Yes
Erwinia amylovora	No	Yes	No

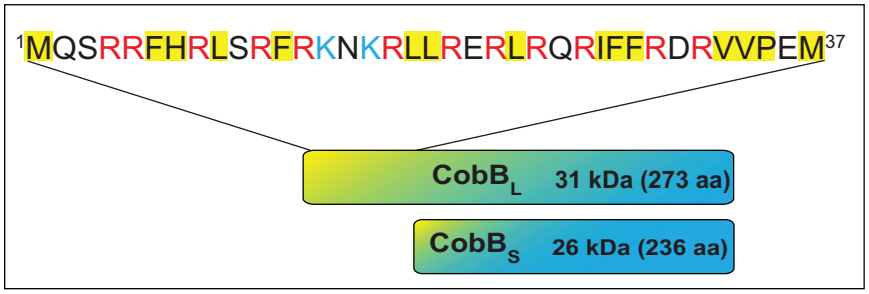

Figure 4.1. Biologically active CobB sirtuin deacylase isoforms of *S. enterica*. *N*-terminal amino acid sequence of $CobB_L$. Yellow highlighted residues represent hydrophobic amino acids; red residues are arginines; blue residues are lysines.

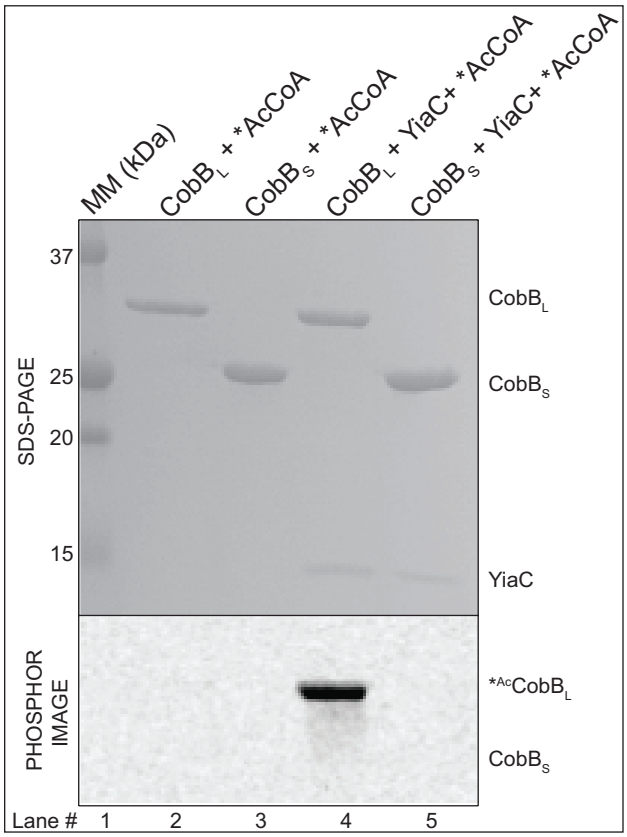

Figure 4.2. YiaC acetylates CobB$_L$ but not CobB$_S$. YiaC-dependent acetylation of CobB isoforms was assessed after incubation of the proteins with [1-^{14}C]-acetyl-CoA (20 μM) for one hour at 37°C. Detailed conditions of the assay are described under *Materials and Methods*. Controls included incubation of the CobB isoforms with [1-^{14}C]-acetyl-CoA in the absence of YiaC. Proteins were resolved by SDS-PAGE and visualized by Coomassie Brilliant Blue R stain (top image) using Precision Plus protein (Bio-Rad) standards as molecular mass markers (kDa). Radiolabel signal was visualized by phosphor imaging (bottom image). Asterisks = radiolabeled acetyl moieties.

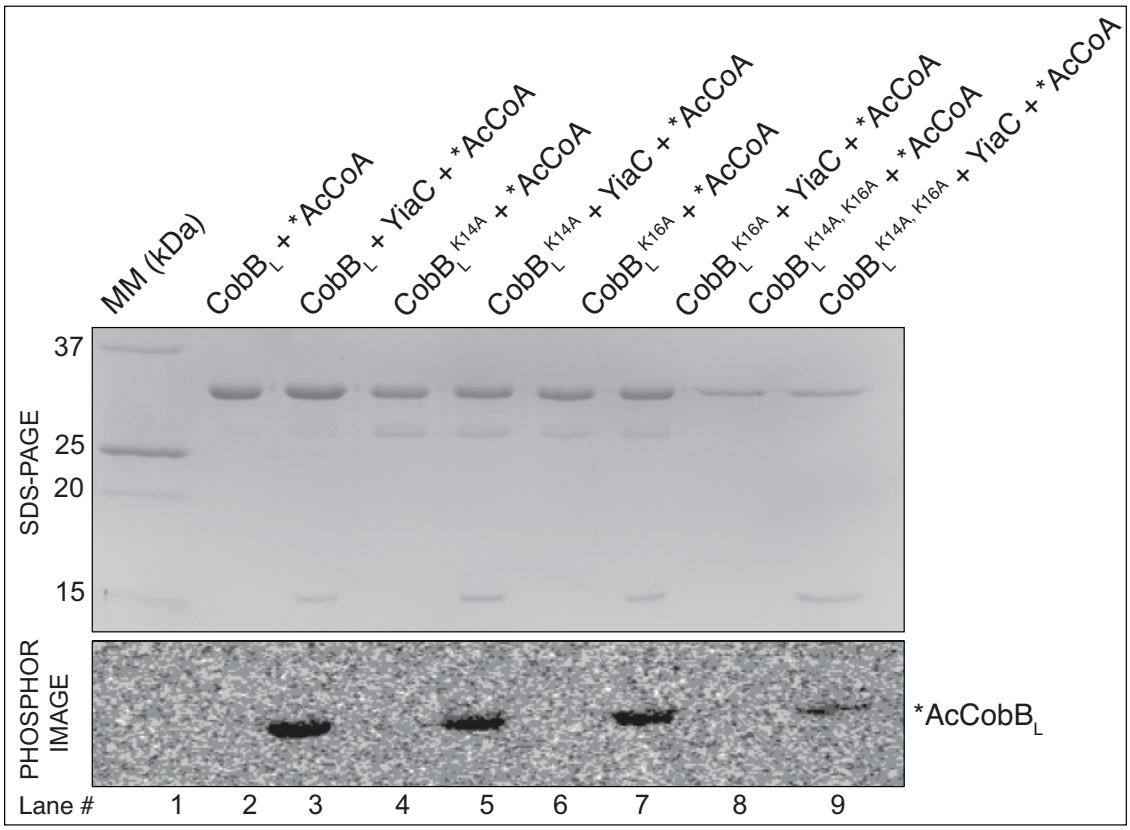

Figure 4.3. YiaC does not acetylate lysyl residues of CobB$_L$. To query the site of modification, lysine variants of CobB$_L$ were purified and incubated with YiaC and [1-^{14}C]-acetyl-CoA. The goal was to determine whether or not YiaC was a *N*α acetyltransferase. This experiment was conducted as described for figure 2 where CobB$_L$ variants (3 μM, except CobB$_L^{K14A\ K16A}$ was at 2 μM) were incubated with [1-^{14}C]-acetyl-CoA and either with or without YiaC (1 μM). Lanes 4 and 5 contained CobB$_L^{K14A}$, lanes 6 and 7 contained CobB$_L^{K16A}$, and lanes 8 and 9 contained CobB$_L^{K14A\ K16A}$. Results of control experiments using CobB$_L$ plus AcCoA with or without YiaC are shown in lanes 2 and 3, respectively. Asterisks = radiolabeled acetyl moieties.

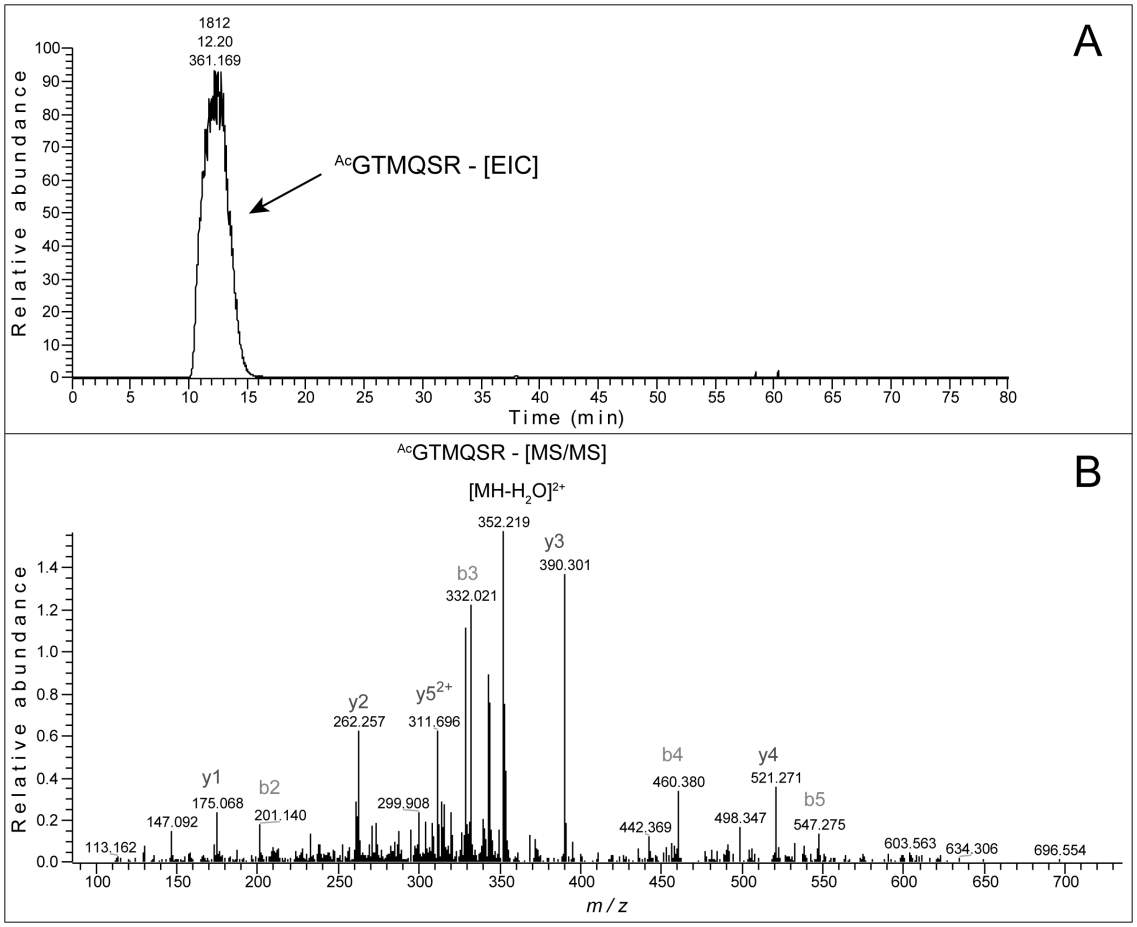

Figure 4.4. Mass spectrometry analysis of acetylated $CobB_L$. $CobB_L$ (5 μM) was incubated with AcCoA (1 mM) with and without YiaC (3 μM) at 37ºC for 1h. Reaction mixture components were resolved by SDS-PAGE. $CobB_L$ was excised from the gel and sent to the UW-Madison Biotechnology Center Mass Spectrometry Facility for LC-MS/MS analysis. A. Chromatographic trace for the selected ion of interest (^{Ac}GTMQSR). EIC, Extracted Ion Chromatogram. B. Mass spectrum of $^{Ac}CobB_L$, *b* ions are the series of fragments that extend from the *N* terminus; *y* ions are the series of fragments that extend from the *C* terminus. MASCOT software (http://www.matrixscience.com) was the online search engine used to identify peptides on the basis of their masses.

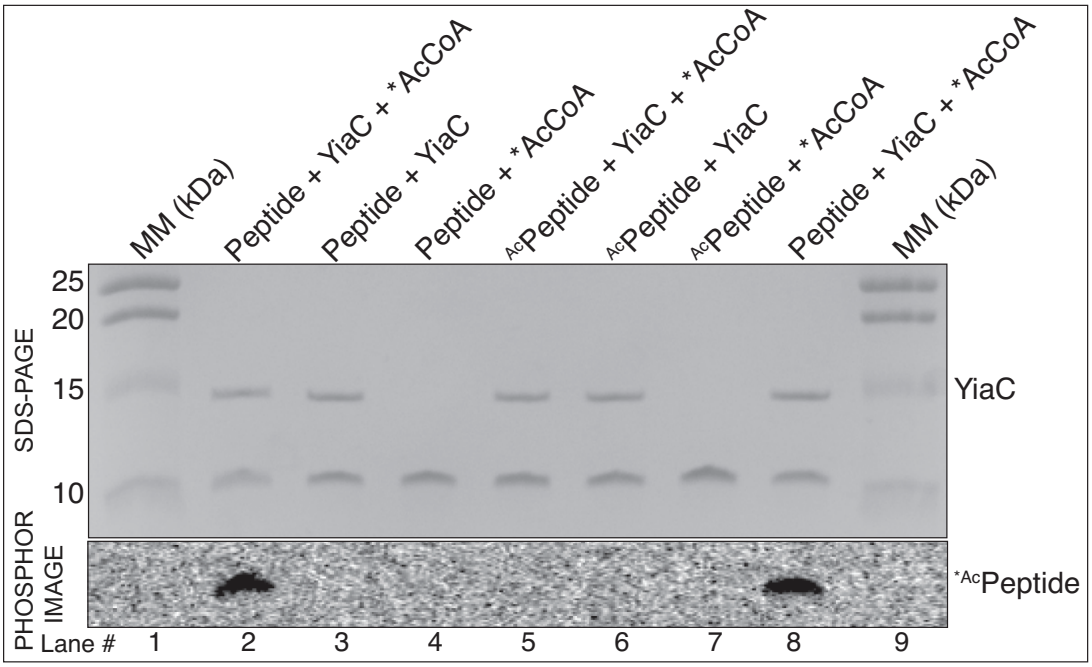

Figure 4.5. YiaC acetylates the *N*-terminal methionine of $CobB_L$. SDS-PAGE and phosphor imaging analysis of unacetylated or *N*-terminally acetylated synthetic peptides of amino acids spanning the first 50 residues of $CobB_L$ were incubated with YiaC without [1-^{14}C]-AcCoA (lanes 2, and 6), with [1-^{14}C]-AcCoA but no YiaC (lanes 4 and 7), or with YiaC and [1-^{14}C]-AcCoA (lanes 2, 5, 8). Lanes 1 and 9 are the Precision Plus protein standard molecular mass markers (kDa). Asterisks = radiolabeled acetyl moieties.

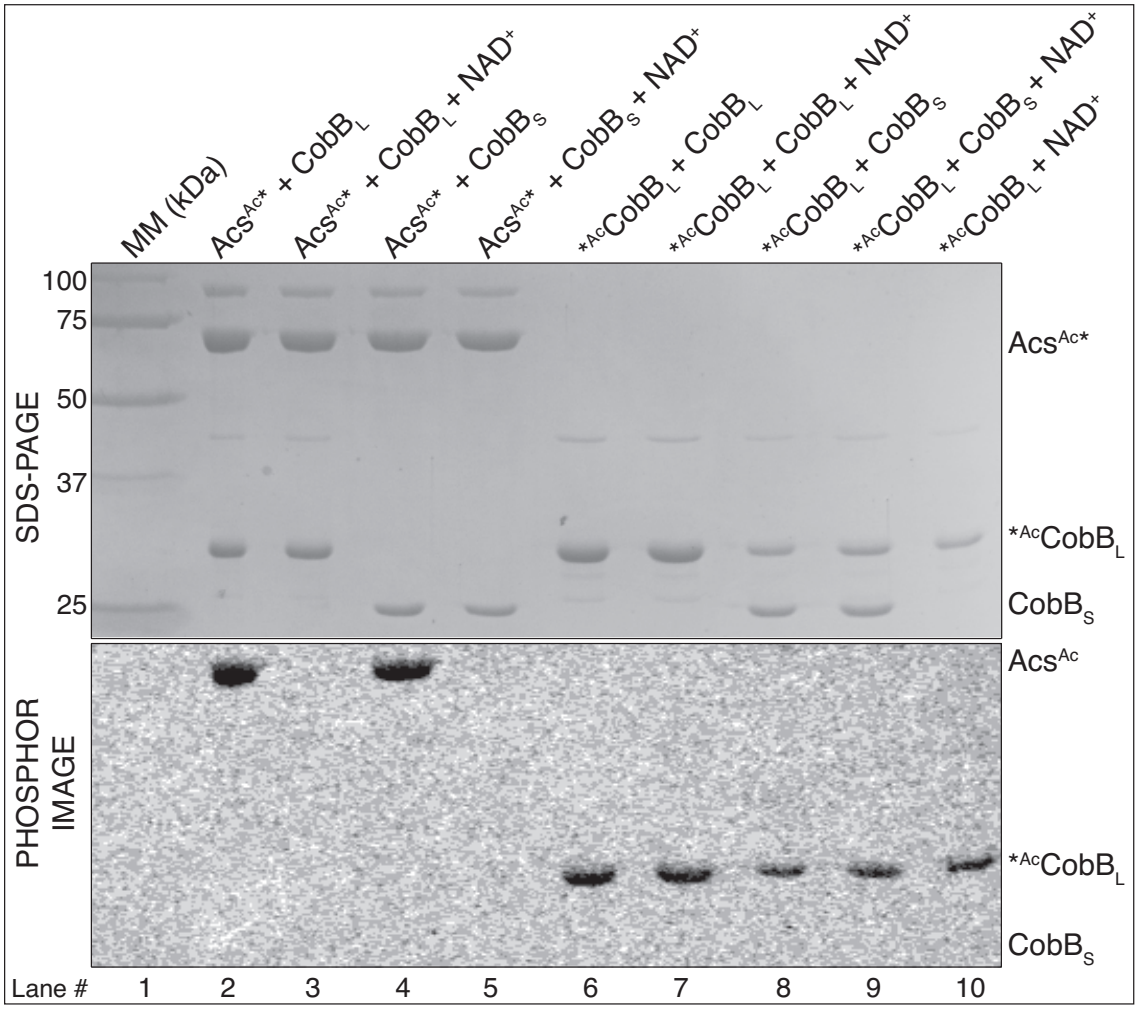

Figure 4.6. *N*α acetylated proteins are not substrates for CobB. [1-^{14}C]-^{Ac}CobB$_L$ was incubated with either CobB$_L$ or CobB$_S$ with (lanes 7, 9) and without NAD$^+$ (lanes 6, 8). Positive controls of unlabeled CobB$_L$ and CobB$_S$ were also tested for their ability to deacetylate [1-^{14}C]- Acs^{Ac} (lanes 1-5). Samples were resolved by SDS-PAGE and transfer of [1-^{14}C]-label was revealed by phosphor image analysis. Asterisks = radiolabeled acetyl moieties.

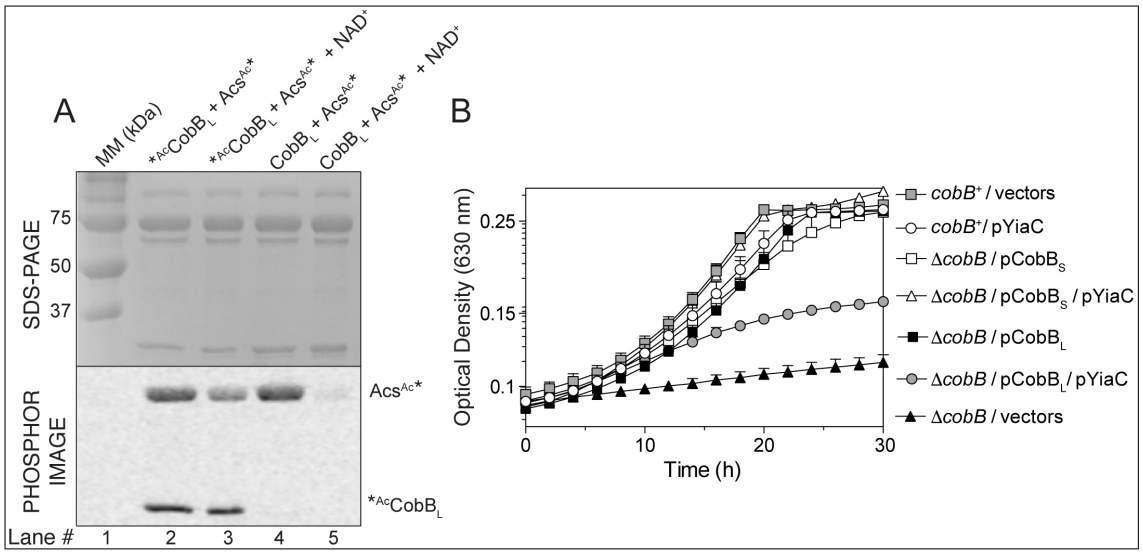

Figure 4.7. Acetylation of the *N* terminus of $CobB_L$ negatively affects its deacetylase activity *in vitro* and *in vivo.* A. To assess the enzymatic activity of $^{Ac}CobB_L$ *in vitro*, Acs protein radiolabeled with [1-^{14}C]-AcCoA (Acs^{Ac*}) was incubated with either $CobB_L$ (lane 5) or $^{*Ac}CobB_L$ (lane 3) and NAD^+ to visualize the removal of the radiolabeled acetyl-group from Acs. $CobB_L$ was acetylated with YiaC using [1-^{14}C]-AcCoA as substrate. Radiolabeled, acetylated $CobB_L$ ($^{*Ac}CobB_L$) helped visualize the mobility of $^{Ac}CobB_L$ on the phosphor image. Negative controls included reactions listed above except no NAD^+ was added (lanes 2 and 4). B. All strains were grown on NCE minimal medium supplemented with acetate (10 mM) as the sole source of carbon and energy. The strains used were: Δ*cobB* / pCV1 / pCV3 (black triangles; negative control), Δ*cobB* / p*cobB*$^+$ ($CobB_S$) / pCV1 (open squares; complementation), Δ*cobB* / p*cobB*$^+$ ($CobB_L$) / pCV1 (black squares; complementation), *cobB*$^+$ / pCV1 / pCV3 (grey squares; wild-type control), *cobB*$^+$ / p*yiaC*$^+$ / pCV3 (open circles), Δ*cobB* / p*cobB*$^+$ ($CobB_S$) / p*yiaC*$^+$ (open triangles), □*cobB* / p*cobB*$^+$ ($CobB_L$) / p*yiaC*$^+$ (grey circles). Cloning vectors pCV1 and pCV3 contain the same arabinose-inducible promoter [27, 41]. Maps of cloning vectors pCV1 and pCV3 can be found in [27]. The concentration of arabinose used was 100 μM. Each strain was grown in biological and technical triplicate and analyses were conducted three independent times. Error bars represent standard deviation (SD) of technical triplicates. Asterisks = radiolabeled acetyl moieties.

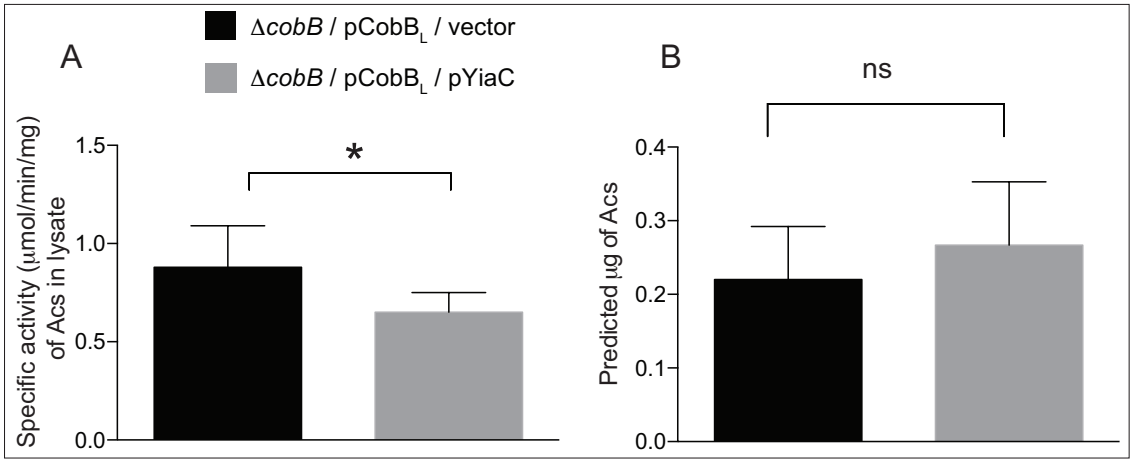

Figure 4.8. Acs activity is decreased when YiaC is over produced in *S. enterica*. A. Strains of *S.enterica* with *cobB* deleted and with only $pCobB_L^{M37A, M38A}$ *in trans* with or without pYiaC were grown to mid-logarithmic phase with 10mM acetate as the sole carbon and energy source and were lysed and tested for Acs specific activity from 4 μg of lysate. Specific activity (μmol AMP min^{-1} mg^{-1}) was calculated using a continuous spectrophotometric assay described in the methods section. The activity of Acs decreased in lysates of the strain with pYiaC overexpressed. The experiment was performed in biological triplicate with nine technical replicates each. B. Acs protein concentration is the same in both strains tested for Acs activity based on quantitative anti-Acs western blot analysis of lysates used in experiment 8A. Error bars represent unpaired t test with equal standard deviation. *, P = 0.03. Figure 8B ns P value = 0.5118.

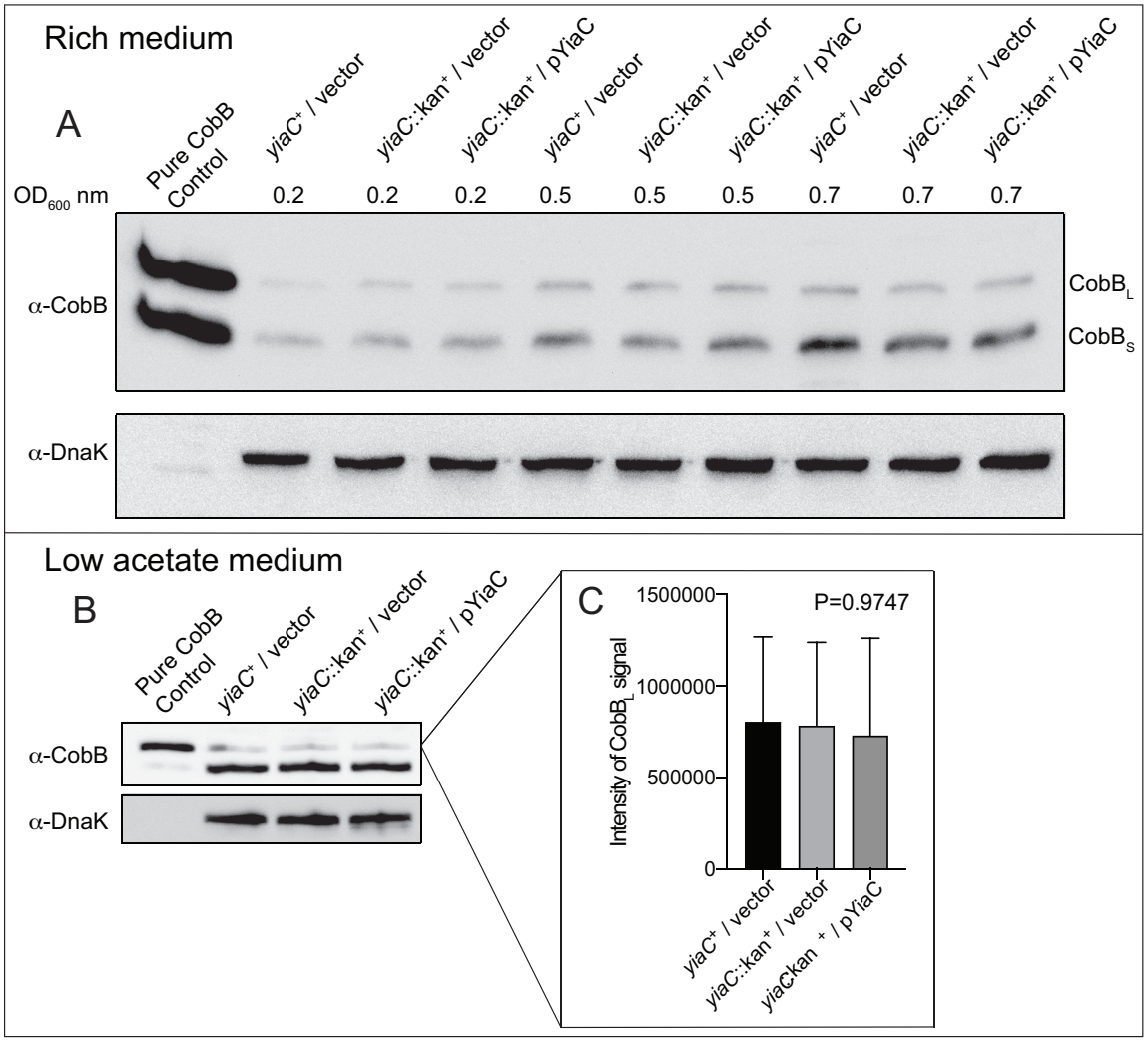

Figure 4.9. YiaC does not alter CobB protein levels *in vivo.* The concentration of CobB isoform protein in *S. enterica* strains where *yiaC* was deleted or complemented by pYiaC were compared to the parent strain (*yiaC*$^+$ / vector) by western blotting. A. Western blots of cells grown on LB medium with 1 mM of L-(+)-arabinose using anti-CobB or anti-DnaK antibodies. Cells were harvested at three different optical densities of 0.2, 0.5, and 0.7. B**.** Western blots of CobB protein levels or DnaK protein levels from cells grown with 10 mM acetate minimal medium supplemented with 500 μM of L-(+)-arabinose. Anti-DnaK western blotting was used as positive controls to show that all lanes were loaded equally and to use as a standard for densitometry calculations. C. The pixel densitometry $CobB_L$ only of the anti-CobB western blot from figure 4.9B showing the concentration of $CobB_L$ in all strains are the same. Calculations were conducted using one-way ANOVA and differences of $CobB_L$ protein concentrations are not significant with a P-value of 0.9747.

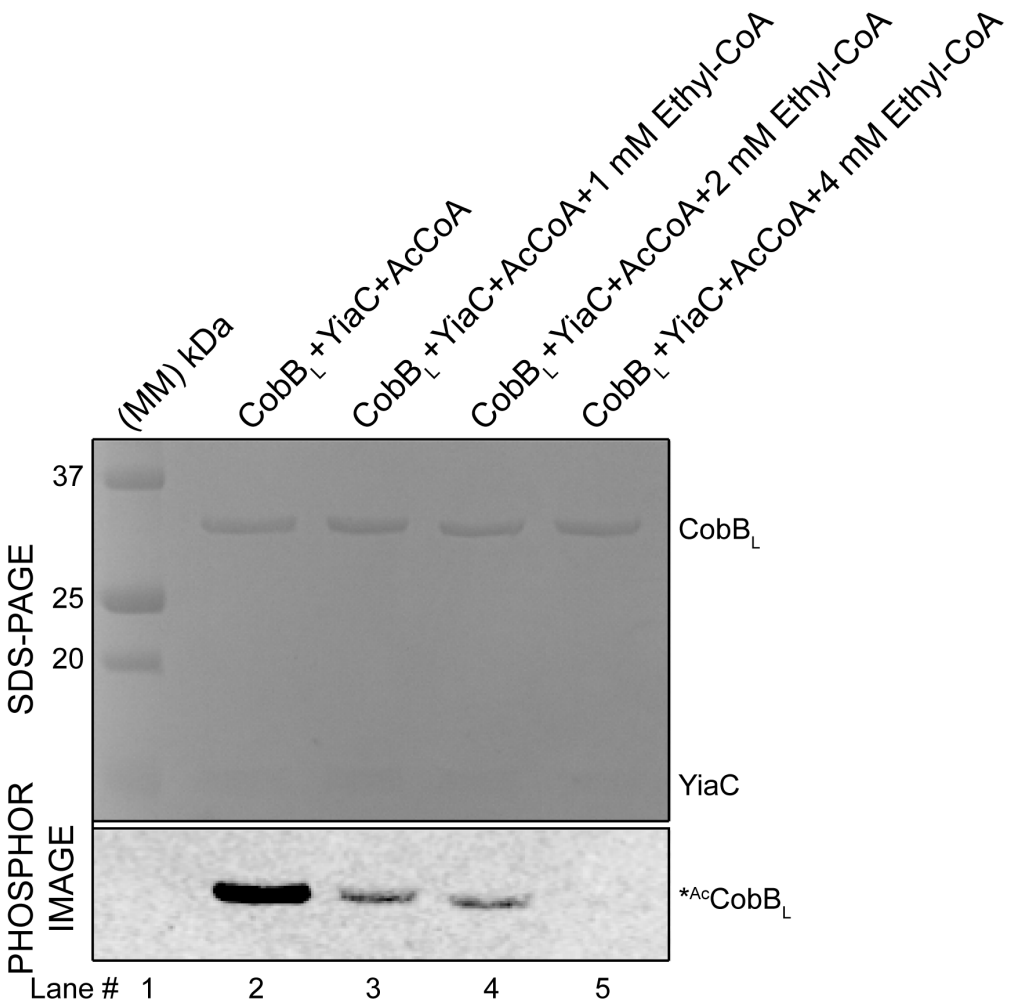

Figure 4.10. Titration of the inhibitory effect of ethyl-CoA on YiaC activity. To assess the concentration of ethyl-CoA required to inactivate purified YiaC protein, ethyl-CoA (1-4 mM) were incubated with the acetyltransferase and target substrates during radiolabeled acetylation assays. Asterisks = radiolabeled acetyl moieties.

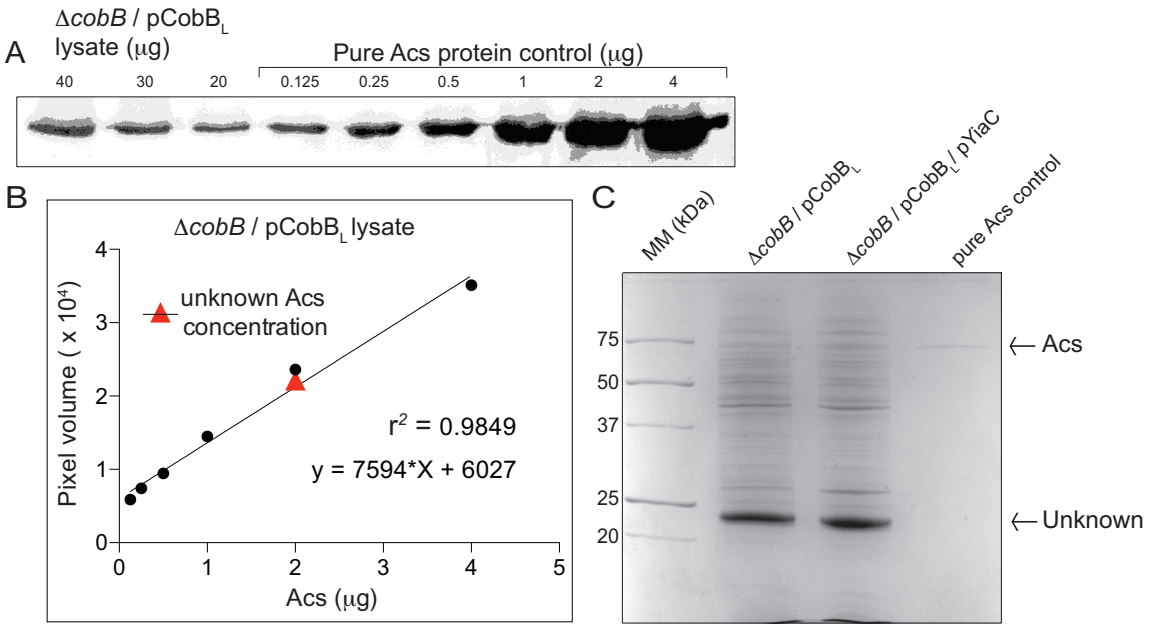

Figure 4.11. Acs protein concentrations are the same in strains carrying the pCobB$_L$ plasmid or the empty cloning vector (vector) *in trans*, relative to a strain carrying the pCobB$_L$ and pYiaC plasmids. A. Representative quantitative western blot using rabbit polyclonal anti-Acs antibodies. These western blots were performed to calculate the concentration of Acs in lysates of *S. enterica* Δ*cobB* strains that synthesized CobB$_L$ or CobB$_L$ and YiaC *in trans*. Concentrations of Acs in cell lysates were compared to a standard curve generated with pure Acs protein. **B.** Linear range of the amount of Acs protein as determined by western blot analysis in the above-mentioned strains. The red triangle represents the amount of protein measured in cell-free extracts of strain Δ*cobB* / pCobB$_L$ or Δ*cobB* / vector in triplicates in three biological replicates. In each instance a standard curve like the one shown was generated. Densitometry was used to quantify Acs. **C.** Accompanying polyacrylamide gel showing that equal amounts of protein present in lysates were loaded into gels.

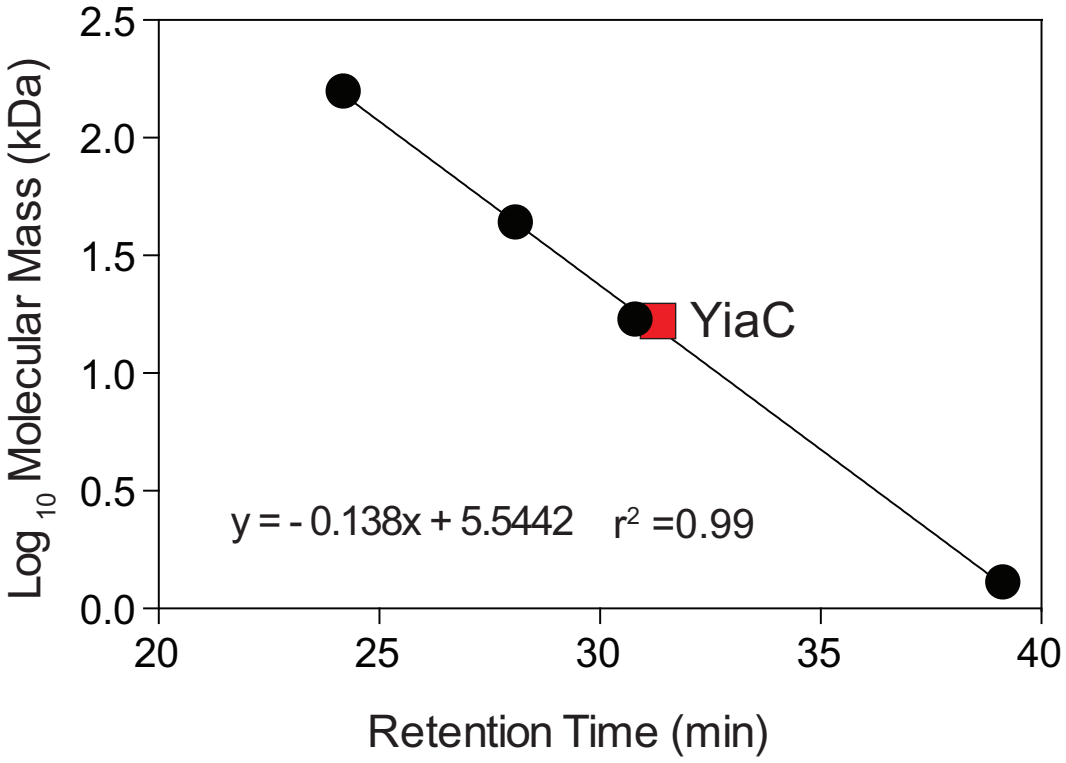

Figure 4.12 Oligomeric state of YiaC in solution. Size exclusion chromatography of purified YiaC protein. The molecular mass of YiaC was determined using a Superose 12 10/300 GL column as described under Materials and Methods. The Bio-Rad Laboratories molecular mass standards were comprised of bovine thyroglobulin (670,000 Da), bovine gamma globulin (158, 000 Da), chicken ovalbumin (44,000 Da), horse myoglobin (17,000 Da), and vitamin B_{12} (1,350 Da). YiaC protein was determined to be a monomer at 17.2 kDa.

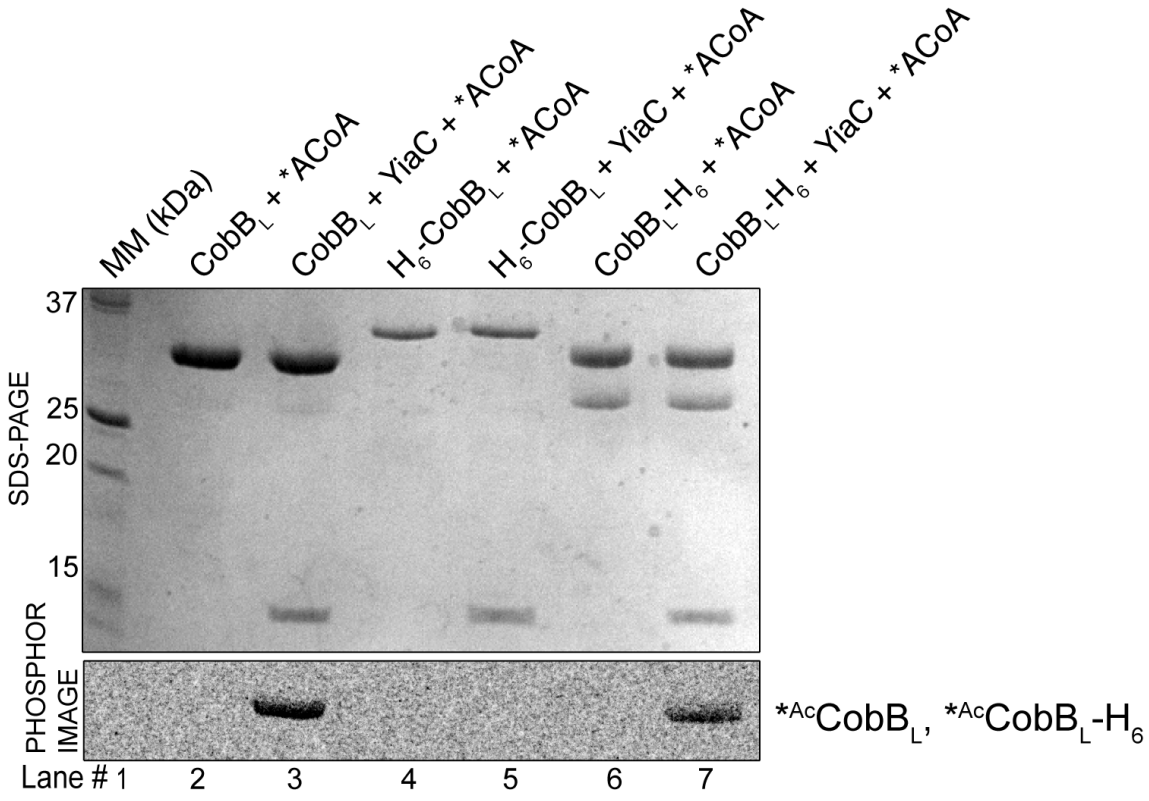

Figure 4.13. YiaC acetylates *C*-terminally His-tagged CobB$_L$, not *N*-terminally H$_6$-tagged CobB$_L$. The ability for YiaC to acetylate **H$_6$**-tagged variants of CobB$_L$ was assessed by *in vitro* radiolabeled acetylation assay as described under *Materials and Methods*. Controls included incubation of 3 μM CobB$_L$ proteins with [1-^{14}C]-acetyl-CoA (20 μM) only (lanes 2, 4, 6), or with [1-^{14}C]-acetyl-CoA (20 μM) and YiaC (1 μM) (lanes 3, 5, 7). Proteins were resolved by SDS-PAGE and transfer of radiolabel was visualized by phosphor imaging.

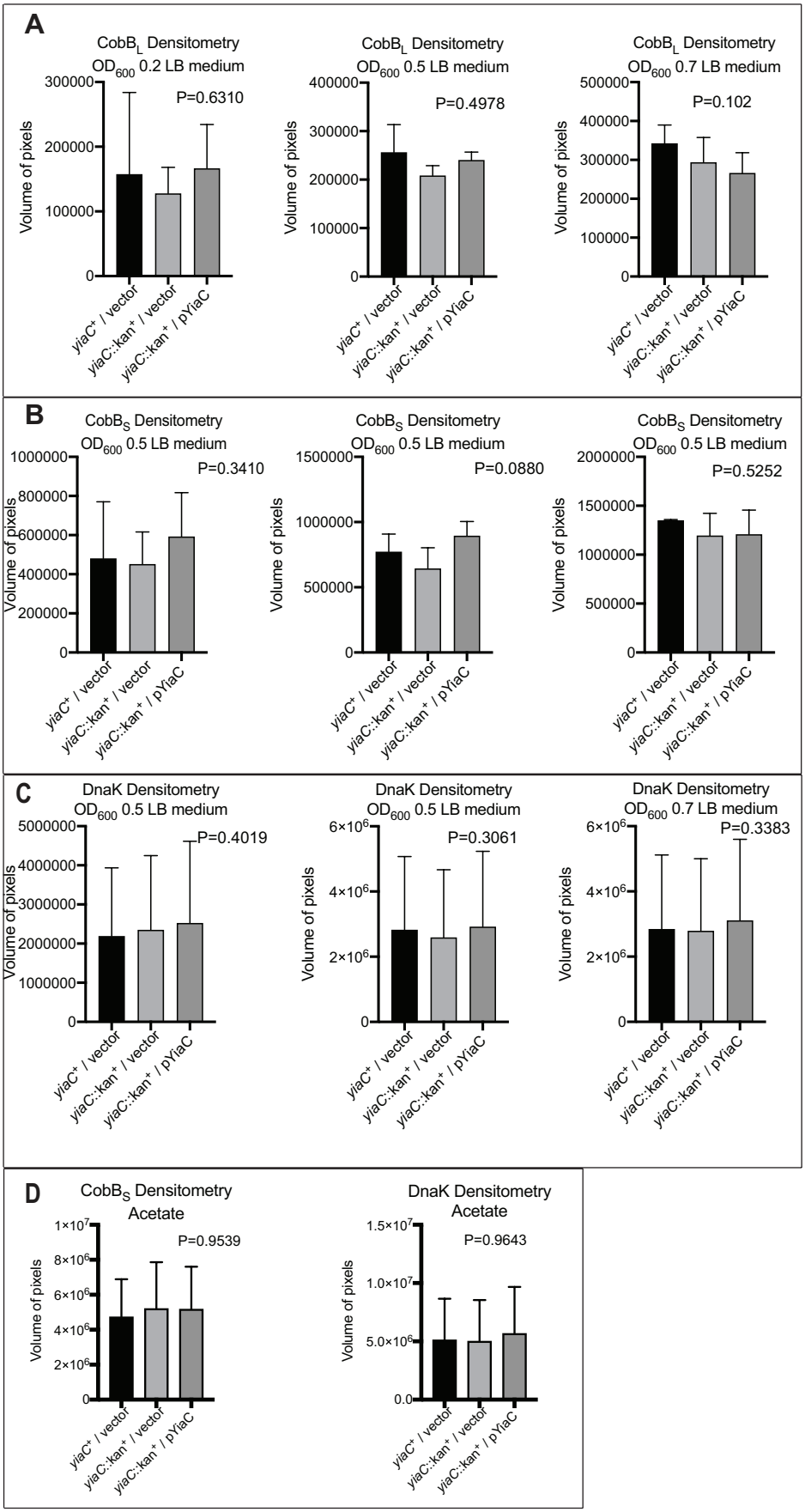
A
CobB$_L$ Densitometry
OD$_{600}$ 0.2 LB medium
P=0.6310
CobB$_L$ Densitometry
OD$_{600}$ 0.5 LB medium
P=0.4978
CobB$_L$ Densitometry
OD$_{600}$ 0.7 LB medium
P=0.102
Volume of pixels
yiaC+ / vector
yiaC::kan+ / vector
yiaC::kan+ / pYiaC
B
CobB$_S$ Densitometry
OD$_{600}$ 0.5 LB medium
P=0.3410
CobB$_S$ Densitometry
OD$_{600}$ 0.5 LB medium
P=0.0880
CobB$_S$ Densitometry
OD$_{600}$ 0.5 LB medium
P=0.5252
C
DnaK Densitometry
OD$_{600}$ 0.5 LB medium
P=0.4019
DnaK Densitometry
OD$_{600}$ 0.5 LB medium
P=0.3061
DnaK Densitometry
OD$_{600}$ 0.7 LB medium
P=0.3383
D
CobB$_S$ Densitometry
Acetate
P=0.9539
DnaK Densitometry
Acetate
P=0.9643

Figure 4.14. Protein levels of the CobB isoforms do not change when YiaC is deleted or over produced. Quantification of $CobB_L$ (**A**), $CobB_S$ (**B**), or DnaK (**C**) protein levels using densitometry of anti-CobB western blots of 50 μg of lysate from cells grown to three different optical densities on rich medium (LB) supplemented with 1 mM L-(+)-arabinose. D. Quantification of $CobB_S$ and DnaK protein levels from 200 μg of lysate from cells grown on minimal medium supplemented with 10 mM acetate and 500 μM L-(+)-arabinose. Strains tested for all western blots are *yiaC*$^+$ / vector, *yiaC*::*kan*$^+$ / vector, or the complemented strain, *yiaC*::*kan*$^+$ / pYiaC. All western blots were conducted in technical duplicates of biological triplicates. Analysis was conducted using one-way analysis of variance (ANOVA) with Prism 8 GraphPad software.

CHAPTER 5

THE LONG ISOFORM OF *S. ENTERICA* COBB ($COBB_L$) SIRTUIN DEACETYLASE BINDS MULTIPLE NUCLEIC ACID AND LIPID SUBSTRATES[4]

[4] Parks A.R.[a], Zhang T., Xu B.Q., and Escalante-Semerena J.C. To be submitted to BioRxiv.

5.1 ABSTRACT

The genome of the enteric bacterium *Salmonella enterica* subsp. *enterica* sv Typhimurium strain LT2 (hereafter *S. enterica*) contains one NAD^+-dependent sirtuin deacylase-encoding gene, CobB. Interestingly, this bacterium synthesizes two biologically active isoforms, $CobB_S$ and $CobB_L$, which differ in size by a 37-residue *N*-terminal extension of the catalytic core. The physiological relevance of possessing two isoforms in *S. enterica* is unknown. Here we provide the first evidence that the *N*-terminal extension of $CobB_L$ enables the long isoform the ability to bind to both DNA and RNA substrates. $CobB_L$ binds to dsDNA, ssDNA, DNA with a 2-base pair gap, cruciform DNA, and RNA, while $CobB_S$ does not have any nucleic acid binding capability of the substrates tested herein. In this study, through the use of atomic force microscopy (AFM), the direct binding of $CobB_L$ to plasmid DNA was visualized. We have also identified growth phenotypes when expression of *cobB* alleles encoding one isoform or the other is increased in a *S. enterica* indicator strain. The alluded phenotypes are readily observed during growth with ethanolamine as the sole carbon and energy source. This newly discovered capability of the long isoform of CobB could help elucidate the biological role of the long isoform as well as shed light on unknown roles of sirtuins in bacterial cell physiology.

5.2 INTRODUCTION

In yeast, the NAD^+-dependent Sir2 protein deacetylase, ySir2p (sirtuin) regulates transcriptional silencing by modulating the acetylation state of lysine-rich histone

tails [1, 2]. Studies of sirtuin function in bacteria, archaea, and eukaryotes have shown that sirtuins are involved in diverse cellular processes. In eukaryotes, sirtuins participate in the regulation of gene expression, metabolism, stress response, and genome integrity maintenance[3-6]. Human cells synthesize seven sirtuin (SIRT1-7) that contain *N*-terminal and or *C*-terminal extensions that enhance their biochemical activities. This includes acting as regulatory regions for post-translational modification, cellular localization [7-9], enzymatic activity [10], or act as extensions that increase their affinity for nucleic acids [11]. For example, a key feature of hSIRT6 activity is that its deacetylase activity is enhanced through binding to nucleosomes, and displays weak deacetylation activity on free acetylated histones *in vitro* [12]. Similarly, the deacetylase activity of hSIRT7 is only activated in the presence of RNA or DNA, and hSIRT7 will not deacylate any non-DNA associated acylated histones *in vitro* [11]. Also, this activity only occurs when the arginine- and lysine-rich *N*- and *C*-terminal domains of SIRT7 are intact [11].

In *S. enterica*, CobB sirtuin activity is needed to regulate the activities of AMP-forming acyl-CoA synthetases (*e.g.,* acetyl- and propionyl-CoA synthetase, Acs and PrpE, respectively) [13]. CobB sirtuin works in concert with the protein acetyltransferase (Pat) that acylates (deactivates) Acs and PrpE, while CobB deacylates (reactivates) [14-16]. Our group has shown that the enteric bacterium *S. enterica* possesses two biologically active isoforms of the sirtuin CobB synthesized from alternative start codons [17]. The long isoform ($CobB_L$) is 37 residues longer than the short isoform ($CobB_S$). The amino acid composition of this

extension is striking: 12 Arg (red letters), two Lys (black highlight), 15 hydrophobic residues (yellow highlight) (^{1}MQSRRFHRLSRFRKNKRLLRERLRQRIFFRDRVVPEM^{37}). In contrast, the short isoform ($CobB_S$) has only the catalytic core. CobB function was originally shown to compensate for the absence of the nicotinate mononucleotide:5,6-dimethylbenzimidazole (DMB) phosphoribosyltransferase (CobT), an enzyme involved in the late steps of coenzyme B_{12} biosynthesis [18]. Previous work also demonstrated that both isoforms of CobB can compensate for the absence of CobT under conditions that demand coenzyme B_{12} biosynthesis [17]. It should be noted that CobB function is not required for coenzyme B_{12} biosynthesis as long as the CobT enzyme is functional. However, this serendipitous ability of CobB to substitute for CobT provides the means to test CobB function *in vivo* under conditions that, as far as we know, are independent of the known role of CobB as a protein deacetylase [14].

Atomic Force Microscopy (AFM) is an ideal tool to image bio-samples because it does not require staining the sample and allows for imaging under near-physiological conditions. [19, 20] Due to the high resolution and easy sample preparation, AFM has been applied widely for imaging DNA, [21-23] protein, [24] and other biomaterials and cells. [25] Further, Magnetic AC mode (MAC mode) AFM, a novel type of tapping mode AFM in which the cantilever is driven directly by a magnetic field, is a powerful tool for imaging with high spatial resolution and better signal-to-noise in liquid environment and therefore may

largely extend the application of AFM to soft living samples, especially those samples that are sensitive to cantilever forces [24, 25].

The physiological need for two biologically active CobB isoforms in *S. enterica* is not known. The work proposed herein queries the physiological roles of the two CobB isoforms, and the involvement of the arginine-rich *N*-terminal extension of $CobB_L$. Here we use electrophoretic mobility shift assays (EMSAs), and atomic force microscopy (AFM) to show that, unlike $CobB_S$, $CobB_L$ non-specifically binds to a variety of nucleic acid substrates. We have also used the ability of CobB to compensate for the absence of CobT to identify a growth condition in which growth of *S. enterica* is altered depending on which CobB isoform is present in the cell. Our *in vivo* data provides the means to discern alternative functions for the CobB isoforms.

5.3 RESULTS AND DISCUSSION

$CobB_L$ is a DNA-binding protein, $CobB_S$ is not. With the high amount of positively charged side chains in the *N*-terminus of $CobB_L$ in mind, we posited that $CobB_L$ would bind to nucleic acids. Electrophoretic mobility shift assays (EMSAs) were performed to test this idea. A double stranded DNA (dsDNA) 5' 6-FAM-labeled probe of the *tacA* promoter from *S. enterica* was incubated with varying fold-excess of either purified $CobB_S$ or $CobB_L$ protein, then run on a native PAGE gel as described under *Experimental Procedures*. Under the conditions used, $CobB_L$ bound to the probe at 10, 15, 20, and 30-fold excess, while $CobB_s$ did not

bind to the DNA at any fold excess (Fig. 5.1A). This result demonstrated that the *N*-terminal extension of $CobB_L$ allowed the protein to bind DNA.

$CobB_L$ binds different DNA substrates. We tested whether or not $CobB_L$ could bind to different nucleic acid substrates. For this purpose, EMSAs were performed with four different DNA probes, i) ssDNA, ii) the dsDNA equivalent of the single-stranded DNA (ssDNA), iii) dsDNA with a 2-nt gap, and iv) cruciform DNA. Gapped DNA and ssDNA were generated as described [26, 27]. $CobB_L$ bound to the ssDNA and dsDNA substrates at 10, 20, and 50-fold excess of $CobB_L$ protein to probe (Fig. 5.1B). Binding of $CobB_L$ to the dsDNA with a 2-nt gap was observed at 20 and 50-fold excess protein to probe, with minimal probe shift seen at 10-fold $CobB_L$ excess (Fig. 5.1B).

Cruciform DNA structures consist of a branch point where a stem and loop form. Branch points and stems can also be found in a related DNA structure, Holliday junctions, which are involved in a wide range of processes such as recombination, replication, gene expression regulation, and double-stranded break repair [28]. These structures of DNA are also targets to a large variety of proteins including junction-resolving enzymes, DNA-repair proteins, nucleoid-associated proteins such as HU, and topoisomerases [27, 28]. A synthetic 6-FAM labeled cruciform DNA based on non-symmetric sequences that allow the structure to be stable was generated and the ability of $CobB_L$ to bind was assessed using EMSAs. $CobB_L$ bound to the cruciform probe at 5- to 50-fold molar excess of protein to probe (Fig 1C). The nucleoid-associated protein HU□$_2$ has been shown to preferentially bind

to cruciform DNA and was used as positive control (Fig 1C) [27]. We speculate that the DNA-binding capability of $CobB_L$ could enhance $CobB_L$'s interactions with other DNA-binding proteins, perhaps to deacetylate them. This DNA-binding ability could also be an unknown regulatory mechanism of $CobB_L$, allowing it to stand in as a nucleoid-associated type protein.

$CobB_L$ binds RNA. To assess whether or not $CobB_L$ could recognize RNA as substrate, competition assays were performed with unlabeled RNA (total RNA from *S. enterica* or yeast tRNA) and 6-FAM-labeled *tacA* probe. In addition to a 1.01-pmol *tacA* promoter DNA probe-only control, every reaction contained 1.01 pmol (50 ng) of *tacA* promoter DNA probe and 20.2 pmol of $CobB_L$. A sweep of increasing concentrations (1, 10, 50, 100, 200 ng) of unlabeled total RNA from *S. enterica* or unlabeled yeast tRNA was added and the reactions were incubated at room temperature for 40 min. Non-denaturing PAGE analysis showed that the labeled DNA probe did not shift in the presence of 100 or 200 ng of unlabeled RNA, and distinct competitive binding was seen in the presence of 10 ng and 50 ng of either source of unlabeled RNA (Fig. 5.1D). To verify that $CobB_L$ was directly binding RNA, unlabeled 51-bp *E. coli hisL* RNA (5.9 pmol) was incubated with increasing amounts of $CobB_L$ protein. Samples were analyzed by non-denaturing PAGE and gels were stained with Sybr Gold for 15 min. Figure 1E shows that $CobB_L$ bound to *hisL* RNA from 2- to 50-fold molar excess of protein to RNA. Peptides containing arginine-rich motifs (a.k.a. ARMs) similar to $CobB_L$'s have been suspected to play roles in RNA-protein interactions [29, 30]. Recently the

emergence of metabolic enzymes moonlighting as RNA-binding proteins has brought about a new hypothesis to understand metabolic enzyme-RNA interactions called REM (RNA, enzymes, metabolites). REM seeks to elucidate the connection between metabolic enzymes that bind RNA, and to understand how RNA binding may link gene expression to intermediary metabolism. The possibility that $CobB_L$ may play a role in modulating the effect of posttranslational modifications on enzyme-RNA interactions is of interest because such a function has not been demonstrated in bacteria (for reviews on this topic see [31, 32]).

A paradigm of the REM hypothesis is the RNA-binding capability of the glycolytic enzyme glyceraldehyde-3-phosphate dehydrogenase (GAPDH). GAPDH has been shown to bind to AU-rich elements (AREs) of mRNA and tRNA [33], [34]. RNA and the enzyme's cofactor, NAD^+, compete for binding to the Rossmann fold domain of GAPDH [35]. CobB also has a Rossmann-fold domain for NAD^+ binding, but as demonstrated with the $CobB_S$ isoform, it is not enough to facilitate nucleic acid binding (Fig. 5.1A). The RNA binding activity of $CobB_L$ could link metabolism and post-translational regulation to gene expression in bacteria.

Recent work showed that the human sirtuin 7 (SIRT7) must bind to nucleic acid with its *N*- and *C*-terminal extensions to be catalytically active, and that RNA serves as a better substrate for promotion of enzymatic activity rather than DNA [11]. RNA binding to $CobB_L$ could help modulate the deacylation of proteins that interact with nucleic acids. The nucleic acid-binding activity of $CobB_L$ would circumvent the need for the enzyme to find its substrate through diffusion increasing the efficiency of deacylation or any other activity yet to be defined for $CobB_L$.

$CobB_L$ binds plasmid DNA as seen by atomic force microscopy. As shown in Figure 2 (A), AFM images clearly show individual pBAD24 plasmid DNA molecules. In ambient conditions the presence of water layers increases the lateral dimensions of the molecules while keeping the height lower than expected. The measured molecule width of about 10-20 nm is consistent with the effect of tip convolution. Moreover the stiffness of the molecule at room temperature is small compared with that of the tip/cantilever and the overall tapping interaction results in a measured height in the range of 0.5-0.7 nm as shown in figure 2 (C). Figure 2 (B) is the AFM image of $CobB_L$-DNA complex where $CobB_L$ molecules bound to pBAD24 DNA that is visualized as lined up protein molecules along the strands of DNA, leaving very little parts of DNA not bound to by $CobB_L$ (Insert in Fig. 5.2 (B). Measurement of the heights of two profiles (labeled 1 and 2) are 1.0-1.3 nm for $CobB_L$-DNA and 0.5-0.7 for unbounded DNA, which agrees with the height of the bare DNA molecules (Fig. 5.2 (A) and (C). It is worth noting that because $CobB_L$ protein was used in high excess, some of the protein aggregated on the MICA surface (Fig. 5.2B). To make sure $CobB_L$ specifically/selectively binds to DNA, we performed AFM imaging of $CobB_S$-DNA complex. As Figure 2 (C) shows, $CobB_S$ does not bind to DNA, but deposits on the MICA surface and DNA then lies freely among these protein molecules. Therefore, we unambiguously demonstrated the DNA-$CobB_L$ interaction by visualizing and comparing the conformation and structure changes using advanced AFM imaging.

Growth conditions that demand CobB$_L$ function. *S. enterica* can grow with ethanolamine as carbon and energy source when the medium is supplemented with B_{12} or its precursors cobinamide (Cbi) and 5,6-dimethylbenzimidazole (DMB) [36]. The conversion of (CN_2)Cbi + DMB to coenzyme B_{12} (CoB_{12}) requires the function of four enzymes (*i.e.,* CobUSTC) [37], one of which (*i.e.,* CobT) is relevant to the work reported here because CobB can restore CoB_{12} synthesis using Cbi + DMB in a Δ*cobT* strain [18].

We observed differences in the behavior of Δ*cobB* Δ*cobT* strains that synthesized *cobB* alleles encoding for CobB$_L^{(M37A\ M38A)}$ or CobB$_S^{(M1A)}$ during growth on ethanolamine medium where Cbi + DMB is needed to synthesize CoB_{12}. As shown in figure 3A, the Δ*cobT* Δ*cobB* strain that synthesized CobB$_L$ under the control of an arabinose-inducible promoter struggled to grow as compared to the *cobB*$^+$ *cobT*$^+$ strain, while the strain expressing CobB$_S$ did not grow (Fig. 5.3A). When arabinose induction was increased 5-fold to 500 μM, the strain that synthesized CobB$_L$ failed to grow, while the strain that synthesized only CobB$_S$ grew, albeit less well than the control strain (Fig. 5.3B). CoB_{12}-dependent growth supported by CobB$_S$ improved with increasing concentration of inducer (1mM) (Fig. 5.3C). This positive effect of increasing concentrations of inducer contrasted sharply with the negative effect that higher CobB$_L$ levels had on cell growth (Fig. 5.3A-C). When growth analysis was conducted where cells were provided CoB_{12}, we see that when high arabinose induction (1mM) is applied, a lag in growth only occurs in the strain expressing CobB$_L$ (Fig. 5.4A and B). Our group has previously shown that CobB$_S$ is present at 5 to 10-fold higher concentration than CobB$_L$ when

S. enterica is grown on a variety of carbon sources [17]. This difference in protein levels may be needed to prevent deleterious effects by $CobB_L$. To demonstrate that the N-terminal arginine residues are playing a role in the interaction of $CobB_L$ with an unknown physiologically relevant substrate, arginine to alanine substitutions were constructed of the arginines present in the N-terminal extension of the $CobB_L$ coding region. These variants were overexpressed in the condition where the ribosyltransferase activity is required. As shown in Figure 5.5A, under low (0.1mM) arabinose induction the more arginine to alanine substitutions that exist in the N-terminal extension of $CobB_L$, the more the growth phenotype of overexpression of CobBL mimics the growth complementation seen of $CobB_S$. The variant of $CobB_L$ that has R11A and R26A changes phenocopies the complementation seen of the native $CobB_L$ isoform, however the $pCobB_L^{R11A\ R17A\ R26A}$ and the $pCobB_L^{R11A\ R17A\ R22A\ R26A}$ variants phenocopy the short isoform of CobB. This also holds true for higher levels of arabinose (0.5mM), where variants of $CobB_L$ that contain more than two arginine-to-alanine changes are able to complement growth much like the complementation seen of $CobB_S$, while the $pCobB_L^{R11A\ R26A}$ variant does not complement growth well much like $CobB_L$ native with high arabinose induction (Fig 5B). Little is known about the regulation of bacterial sirtuins. In prokaryotes, RLA is involved in the regulation of major cellular processes, including enzyme activity, energy homeostasis, protein-DNA binding, protein stability, and protein-protein interactions [38]. The work presented herein provides new knowledge regarding the role of sirtuins in prokaryotes. The unexpected result that $CobB_L$ binds to DNA and RNA raises questions regarding

the specificity of $CobB_L$ for different nucleic acid substrates, the potential for interactions with other DNA-binding proteins, and its physiological role as a nucleic acid binding protein. Addressing these questions will broaden our understanding of the roles that sirtuins play in the bacterial cell.

CobB interacts with liposomes. When thinking about other negatively charged molecules present in the cell that could be interacting with the arginines and lysines of CobBL's N-terminus, we sought to determine if CobBL is able to interact with the lipid bilayer. To test this idea, liposomes were generated through the mixing of POPC, POPE, POPS, and Rh-DHPE (Lissamine™ Rhodamine B 1,2-Dihexadecanoyl-*sn*-glycero-3-phosphoethanolamine) (as described in experimental procedures). Once the liposomes were prepared, CobBL or CobBS proteins were added to liposomes and allowed to incubate together at 4C for 1 hour. The proteoliposomes were then subjected to ultracentrifugation. If $CobB_L$ was able to interact with the lipids of the liposome, we postulated that western blot analysis of the liposomes using anti-CobB antibodies should yield signal at the fraction where the liposomes exist. If CobB cannot interact with the liposomes, we do not expect a signal to be seen and instead the CobB protein would have sedimented out of the liposome fraction during ultracentrifugation. As shown in figure 5.6, a large signal can be seen of $CobB_L$ protein existing in the liposome fraction. However, a signal also exists for $CobB_S$ in the liposome fraction. This could be for several reasons, such as too much $CobB_S$ protein was using which

saturated our signal, or CobB can interact with lipids without the requirement of the *N*-terminal extension.

5.4 EXPERIMENTAL PROCEDURES

Protein purification. HUβ_2 was purified as previously described[39] $CobB_S$ was purified as described elsewhere [17]. $CobB_L$ was purified as follows: plasmid pCOBB72 was electroporated into *E. coli* C41 (λDE3) Δ*pat* (strain JE9314). Cultures of *E. coli* containing plasmid pCOBB72 were grown overnight in super broth [40] and were sub-cultured 1:100 (v/v) into two liters of super broth containing ampicillin (100 μg/ml). Cells were grown shaking at 37 °C to an optical density (measured at 600 nm; OD_{600}) of 0.7, induced with isopropyl-β-D-thiogalactopyranoside (IPTG, 0.5 mM), and grown overnight at 37 °C. After 16 h of incubation, cells were pelleted at 6,000 x *g* for 15 min in a Beckman Coulter Avanti J-2 XPI centrifuge equipped with a JLA-8.1000 rotor. Cell pellets were re-suspended in 60 ml of cold buffer A containing sodium phosphate buffer (20 mM, pH 7.5) containing NaCl (500 mM), imidazole (20 mM), glycerol (10%, v/v). Phenylmethanesulfonyl fluoride (PMSF, 1mM), lysozyme (1 mg/ml), and DNase I (1μg/ml) were added to the cell suspension, and cells were lysed on ice by three, 30-s cycles of sonication (2 sec on, 2 sec off) at 60% amplitude on a Qsonica sonicator. Cell lysates were centrifuged at 40,000 x *g* for 30 min in a Beckman Coulter Avanti J-25I centrifuge equipped with a JA-25.50 rotor centrifuge then filtered with a 0.45 μm filter. Filtered lysates were applied to a 5-ml HisTrap FF column (GE Healthcare) using an ÄKTA fast protein liquid chromatography (FPLC)

system (GE Healthcare). The column was washed with 15 column volumes of buffer A, 15 ml of 4% elution buffer B [sodium phosphate buffer (20 mM, pH 7.5) containing NaCl (500 mM), imidazole (500 mM), glycerol (10%, v/v)], followed by a 15-column volumes linear gradient elution from 4% to 100% buffer B, and 5 column volumes of 100% buffer B. Fractions were resolved on a 15% sodium dodecyl sulfate-polyacrylamide electrophoresis (SDS-PAGE) gel [41] and elution fractions where MBP-His_6-CobB were present were pooled, recombinant tobacco etch virus (rTEV) protease [42] was added at a concentration of 1 mg TEV: 100 mg protein, and dialyzed for 3 h at room temperature in sodium phosphate buffer (20 mM, pH 7.5) containing, NaCl (500 mM), glycerol (10%, v/v), dithiothreitol (DTT, 1mM). Protein mixtures were dialyzed at 4°C overnight in 20mM sodium phosphate buffer pH 7.5, 500mM NaCl, 10% glycerol (v/v), then at 4°C for 3 h in sodium phosphate buffer (20 mM, pH 7.5), containing NaCl (500 mM), glycerol (10%, v/v), and imidazole (20 mM). After dialysis the protein was filtered and re-loaded onto the 5-ml HisTrap column, and the above protocol was repeated. Fractions were resolved by SDS-PAGE and fractions containing only $CobB_L$ were pooled and dialyzed at 4 °C for 3 h in HEPES buffer (20mM, pH 7.5), containing NaCl (250 mM), glycerol (10%, v/v), then dialyzed and stored in HEPES buffer (20mM, pH 7.5), containing NaCl (100mM), glycerol (10%, v/v). Concentrations of $CobB_L$ were determined using a Nanodrop® ND-1000 spectrophotometer (Thermo Fisher Scientific) by measuring the absorbance at 280 nm and using the molar extinction coefficient of the protein (22,460 M^{-1} cm^{-1}).

Construction of DNA substrates. Electrophoretic mobility shift assays (EMSAs) using DNA probes were performed with 6-carboxyfluorescein (6-FAM) covalently attached at the 5' end or 3' end. Primers used in this study were manufactured by Integrated DNA Technologies (IDT, Coralville, IA). The 75-base pair *tacA* probe was generated by PCR amplification using 5' *tacA* EMSA primer (5' 6-FAM label) and 3' *tacA* EMSA primer (listed in Table S2) and genomic DNA from *S. enterica* strain JE10079 (*ara-9*). Sequences of the cruciform, 2-nt gapped DNA, ssDNA and dsDNA were taken from published work and can be found in Table 5.2. Formation of cruciform DNA annealed at a 1:1:1:1 ratio of primers A through D. Linear dsDNA was formed by annealing a 1:1 ratio of 5' 6-FAM labeled primer A with primer E, while 2-nt gapped DNA was formed by annealing a 1:1:1 ratio of primers F, G, and 3' 6-FAM labeled H. Linear ssDNA was 5' 6-FAM labeled primer A by itself. All primer mixtures were heated to 70 ºC in a Thermomixer R (Eppendorf) for 5 min and cooled to room temperature over a 2 h period. PCR products and annealed oligonucleotides were checked for amplification/annealing on agarose gels (1%, w/v) and were purified from the gel using the Promega Wizard SV gel and PCR Cleanup System. The concentration of probes was quantified using a Nanodrop® ND-1000 spectrophotometer (Thermo Fisher).

DNA binding assays. DNA binding assays were performed in 25-μl reactions mixtures containing either 25 ng or 50 ng (or 0.5 pmol-3 pmol) of DNA probe, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) buffer (10 mM, pH 7.5), containing glycerol (10%, v/v), and when stated purified $CobB_S$, $CobB_L$, or $HU\beta_2$

protein in molar fold excess to DNA at concentrations indicated in the figure legends. Binding reactions were incubated for 40 min at room temperature. After incubation, glycerol (27 μmol; 5 μl of a 50% v/v solution) was added to each reaction and samples were run at 4 °C on non-denaturing Criterion *tris*(hydroxymethyl)aminomethane (Tris)-HCl buffer (375 mM, pH 8.6) 7.5% (w/v) polyacrylamide gel (BioRad) at 120 V with cold 0.5X Tris-borate-ethylenediaminetetraacetic acid (EDTA) (TBE) buffer. Gels were imaged using a Typhoon Trio+ variable mode imager (GE Healthcare) at the wavelength 488 nm, and quantified with ImageQuant v5.2 software.

RNA binding assays. RNA competition assays were performed in 25-μl reaction mixtures with HEPES buffer (10 mM, pH 7.5), glycerol (10%, v/v), 5' 6-FAM (1 pmol, 50 ng) labeled *tacA* promoter DNA, $CobB_L$ protein (20 pmol), Rnase OUT™ recombinant ribonuclease inhibitor (0.25 μl) (Invitrogen), and when stated total RNA or yeast tRNA (Sigma) at concentrations indicated in figure legend. Total RNA from *S. enterica* was isolated as described previously {VanDrisse, 2017 #26732}. After reaction mixtures were incubated for 40 min at room temperature, glycerol (27 μmol; 5 μl of a 50% v/v solution) was added and samples were analyzed by non-denaturing PAGE and evaluated as described above. For *hisL* RNA binding assays, 5.9 pmol (200 ng) of unlabeled *hisL* RNA was incubated with Rnase inhibitor and increasing fold-excess of $CobB_L$ protein. Reaction buffer, incubation conditions, and PAGE conditions were the same as described above. Gels were post-stained in the dark with 1X Sybr Gold nucleic acid stain (Invtrogen)

for 15 min. Gels were imaged using a Typhoon Trio+ variable mode imager (GE Healthcare) at wavelength 488nm (Blue) with filter 555 SP at 555 volts and evaluated with ImageQuant v5.2 software.

Technical details of AFM method used in the DNA & CobB-DNA imaging. Here, an Agilent 5500 Controller combined with a 50um by 50um Agilent multipurpose AFM scanner (S/N: 325-00388, with 0.5 angstroms RMS of z noise specification) was used to obtain images in an area as large as 10um by 10um to locate an area of interest of 2μm by 2μm to scan. The whole system was shielded from environmental interference by a PicoPlusIsolation Chamber. Silicon cantilevers tip with spring constant of around 0.1 N/m were used for experiments. All the images were taken under TopMAC mode with a PicoTREC controller (Agilent Technologies, Santa Clara, CA), which is based on a single, low-force touch of the cantilever tip mentioned above. The AFM images were obtained with 512 x 512points and processed using Gwyddion software. [43, 44].

Sample preparations of pCV1 plasmid DNA and CobB. Plasmid pCV1 was dissolved in 20 mM HEPES buffer (pH 7.5) with glycerol (10%, v/v) to a concentration of 62.6 ng/mL. To acquire images of plasmid only controls by AFM, the DNA sample was diluted 50-fold with 10 mM HEPES to the final concentration of 1.25 ng/mL. diluted DNA 20uL was deposited on freshly cleaved mica and wait for 30 min to perform the TopMAC imaging.

For DNA and CobB protein binding experiment, $CobB_S$(2.03 mg/ml) or $CobB_L$(1.71 mg/ml) protein were frozen in a solution of HEPES buffer (20 mM) with glycerol (10% v/v) and NaCl (100 mM). frozen protein was left on ice until completely thawed. CobB proteins were diluted 2,000-fold with HEPES (10 mM) to a concentration of 855 ng/mL, and 10μL of diluted protein was mixed with 10μL 1.25 ng/mL DNA. The mole ratio of $CobB_L$ to DNA is about 50,000 : 1. The excess amount of $CobB_L$ is to ensure a quick reaction due to the slow diffusion process. The binding reaction incubated at room temperature for 40 min. 20 μL of the binding reaction was pipetted on the freshly cleaved mica and allowed to set for 30 min prior to performing TopMAC imaging.

Bacterial strains. All strains used in this work were derivatives of *Salmonella enterica* subsp. *enterica* serovar Typhimurium LT2 that were constructed the Wanner in-frame gene deletion method [45]. All of the bacterial strains and plasmids used in this study are listed in Table 5.1 in the supplemental material.

Growth analysis. Bacteria harboring empty vector, $pCobB_S$ or $pCobB_L$ were growth shaking at 37°C. Bacterial strains were grown in nutrient broth (NB [Difco]), or no-carbon essential (NCE) minimal medium [46]. Growth analysis where ethanolamine was the sole carbon source were conducted in NCE medium with MgSO4 (1 mM), Wolfe's trace minerals (1X) [47], ethanolamine (90mM), ampicillin (100 μg/ml), 5,6-dimethylbenzimidazole (DMB, 150 μM), dicyanocobinamide $(CN)_2Cbi$, 150 nM), and L-(+)-arabinose as indicated for induction of gene

expression. Cyanocobalamin (CNCbl) (150μM) was added in place of $(CN)_2$Cbi + DMB when indicated. Bacterial cultures were grown overnight at 37 ºC in NB. Microtiter plates (96-well) containing 190 μl of ethanolamine medium per well were inoculated in technical triplicate with 10 μl of starter cultures (5% , v/v), and were incubated at 37 °C in a PowerWave microtiter plate reader (Bio-Tek Instruments) under continuous medium shaking; data points were recorded every 30 min. The optical density was read at 630 nm and data points were analyzed using Prism 6 software (GraphPad).

Liposome preparation and Liposome Flotation Assay. The lipids used in the study were POPS, POPE, POPC, and Rh-DHPE (Lissamine™ Rhodamine B 1,2-Dihexadecanoyl-*sn*-glycero-3-phosphoethanolamine). Proteoliposomes were prepared using detergent-mediated reconstitution as described elsewhere [48]. POPC:POPS:POPE:Rh-DHPE were combined at a molar ratio of 80:10:9.5:0.5 to yield 1 μmol lipid. A stream of N_2 gas was used to dry lipids and dried lipids were spun in an Eppendorf VacFuge at 30 ºC for 1 h. The dried lipid film was reconstituted to a final lipid concentration of 2 mM in 500 μl of HEPES buffer (20 mM, pH 7.4) containing CHAPS (40 mM), glycerol (10% [v/v]), and NaCl (0.5 M). Homogeneously purified $CobB_L$ and $CobB_S$ proteins were added at a lipid-to-protein molar ratio of 1000:1 (2mM lipid to 2uM CobB protein) and incubated with rotation on a Benchmark RotoBot at 4 °C for 1 h. Detergent removal and proteoliposome formation was achieved dialysis three times in 1 liter each of HEPES buffer (20 mM, pH 7.4) containing NaCl (0.5 M) and glycerol (10% [v/v])

using Slide-A-Lyzer dialysis cassettes (20-kDa molecular mass cutoff). After dialysis, the proteoliposome suspension was prepared for the liposome flotation assay by application to a Histodenz (Sigma) gradient by mixing the proteoliposomes with an equal volume with HEPES buffer (20 mM, pH 7.4) containing NaCl (0.15 M), glycerol (10% [v/v]), Histodenz (80% [w/v]) and transferred to a Beckman Coulter Ultracentrifuge tube (polyallomer, 13 x 51 mm). A 4-mL overlay of HEPES (20 mM, pH 7.4) containing Histodenz (30% [w/v]), NaCl (0.15 M), glycerol (10% [v/v]) was applied followed by a 200-μL overlay of HEPES buffer (20 mM, pH 7.4) containing NaCl (0.15 M), glycerol (10% [v/v]). The gradient was centrifuged at 4 °C for 3 h at 268,000 x *g* in a 4C Beckman Coulter Optima™ MAX-XP Ultracentrifuge using an MLS-50 rotor. Lipid concentration of the reconstituted liposomes was determined through generation of a standard curve using the fluorescence of Rh-DHPE, measured at an excitation wavelength of 540 nm and emission wavelength of 596 nm using a BioTek Gemini fluorescent plate reader.

Western blot analysis of floated proteoliposomes. To assess whether the CobB isoforms floated with the liposomes or sedimented in the ultracentrifugation process, western blot analysis was conducted using anti-CobB antibodies on the liposomes isolated from ultracentrifugation. The same quantity of proteoliposomes (as determined by a standard curve of Rh-DHPE) that were floated with either: 1) CobBL protein 2) CobBS protein 3) negative control of empty liposome were loaded onto 15% polyacrylamide SDS-PAGE gels and electrophoresed at 200V

for 45 min. Proteins in the gels were then transferred to a 0.45 μm PVDF membranes (Immobilon® – P membranes from Millipore) using a Bio-Rad Trans-Blot® Turbo™ transfer system at 25 V, 1.0 A for 30 min in 1X Transfer buffer (Tris-HCl (25 mM), glycine (192 mM), and 10% methanol (v/v). Membranes were blocked for 1 hour at room temperature on a rocker with phosphate buffered saline containing 1% tween-20 (PBST) and 5% dehydrated milk powder, then membranes were incubated overnight rocking at 4C with a dilution 1:500 of anti-CobB polyclonal rabbit antibodies (generated by Envigo, PA) in 1X PBST buffer containing 5% milk powder. The following morning, membranes were washed three times with PBST, then incubated for 1 hour rocking at room temperature with PBST-5% milk buffer containing 1:10,000 anti-rabbit IgG (whole molecule) peroxidase antibodie conjugate (Sigma). The membranes were then washed three times again with PBST and developed using Thermo Scientific SuperSignal™ West Pico Plus chemiluminescent substrate and imaged with a Analytik Jena UVP ChemStudio imaging system.

5.5 ACKNOWLEDGEMENTS

We would like to thank Ran Zhang for assisting with the atomic force microscopy.

5.6 REFERENCES

1. Imai, S., et al., *Sir2: an NAD-dependent histone deacetylase that connects chromatin silencing, metabolism, and aging.* Cold Spring Harb Symp Quant Biol, 2000. **65**: p. 297-302.
2. Landry, J., J.T. Slama, and R. Sternglanz, *Role of NAD(+) in the deacetylase activity of the SIR2-like proteins.* Biochem Biophys Res Commun, 2000. **278**(3): p. 685-90.
3. Brachmann, C.B., et al., *The SIR2 gene family, conserved from bacteria to humans, functions in silencing, cell cycle progression, and chromosome stability.* Genes Dev, 1995. **9**(23): p. 2888-902.
4. Guarente, L., *Sirtuins in aging and disease.* Cold Spring Harb Symp Quant Biol, 2007. **72**: p. 483-8.
5. Starai, V.J., et al., *Short-chain fatty acid activation by acyl-coenzyme A synthetases requires SIR2 protein function in Salmonella enterica and Saccharomyces cerevisiae.* Genetics, 2003. **163**(2): p. 545-55.
6. Kwon, H.S. and M. Ott, *The ups and downs of SIRT1.* Trends Biochem Sci, 2008. **33**(11): p. 517-25.
7. Ahuja, N., et al., *Regulation of insulin secretion by SIRT4, a mitochondrial ADP-ribosyltransferase.* J. Biol. Chem., 2007: p. 33583-92.
8. Schwer, B., et al., *The human silent information regulator (Sir)2 homologue hSIRT3 is a mitochondrial nicotinamide adenine dinucleotide-dependent deacetylase.* J Cell Biol, 2002. **158**(4): p. 647-57.

9. Matsushita, N., et al., *Distinct regulation of mitochondrial localization and stability of two human Sirt5 isoforms.* Genes Cells, 2011. **16**(2): p. 190-202.

10. Kang, H., et al., *Peptide switch is essential for Sirt1 deacetylase activity.* Mol. Cell, 2011: p. 203-13.

11. Tong, Z., et al., *SIRT7 Is activated by DNA and deacetylates histone H3 in the chromatin context.* ACS Chem. Biol., 2016. **11**: p. 742-747.

12. Gil, R., et al., *SIRT6 exhibits nucleosome-dependent deacetylase activity.* Nucleic Acids Res., 2013. **41**: p. 8537-8545.

13. Starai, V.J., et al., *A link between transcription and intermediary metabolism: a role for Sir2 in the control of acetyl-coenzyme A synthetase.* Curr. Opin. Microbiol., 2004. **7**: p. 115-119.

14. Starai, V.J., et al., *Sir2-dependent activation of acetyl-CoA synthetase by deacetylation of active lysine.* Science, 2002. **298**: p. 2390-2392.

15. Starai, V.J. and J.C. Escalante-Semerena, *Acetyl-coenzyme A synthetase (AMP forming).* Cell. Mol. Life Sci., 2004. **61**: p. 2020-2030.

16. Starai, V.J. and J.C. Escalante-Semerena, *Identification of the protein acetyltransferase (Pat) enzyme that acetylates acetyl-CoA synthetase in Salmonella enterica.* J Mol Biol, 2004. **340**(5): p. 1005-12.

17. Tucker, A.C. and J.C. Escalante-Semerena, *Biologically active isoforms of CobB sirtuin deacetylase in Salmonella enterica and Erwinia amylovora.* J. Bacteriol., 2010. **192**: p. 6200-6208.

18. Tsang, A.W. and J.C. Escalante-Semerena, *CobB, a new member of the SIR2 family of eucaryotic regulatory proteins, is required to compensate for*

the lack of nicotinate mononucleotide:5,6-dimethylbenzimidazole phosphoribosyltransferase activity in cobT mutants during cobalamin biosynthesis in Salmonella typhimurium LT2. J. Biol. Chem., 1998. **273**(48): p. 31788-31794.

19. Li, Q., et al., *AFM-based force spectroscopy for bioimaging and biosensing.* RSC Advances, 2016. **6**(16): p. 12893-12912.
20. Senapati, S. and S. Lindsay, *Recent Progress in Molecular Recognition Imaging Using Atomic Force Microscopy.* Accounts of Chemical Research, 2016. **49**(3): p. 503-510.
21. Berg, F., et al., *AFM-Based Quantification of Conformational Changes in DNA Caused by Reactive Oxygen Species.* The Journal of Physical Chemistry B, 2015. **119**(1): p. 25-32.
22. Pang, D., A.R. Thierry, and A. Dritschilo, *DNA studies using atomic force microscopy: capabilities for measurement of short DNA fragments.* Frontiers in Molecular Biosciences, 2015. **2**: p. 1.
23. Pyne, A., et al., *Single-Molecule Reconstruction of Oligonucleotide Secondary Structure by Atomic Force Microscopy.* Small, 2014. **10**(16): p. 3257-3261.
24. Lou, Z., et al., *Molecular-level insights of early-stage prion protein aggregation on mica and gold surface determined by AFM imaging and molecular simulation.* Colloids and Surfaces B: Biointerfaces, 2015. **135**: p. 371-378.

25. Ge, G., et al., *MAC mode atomic force microscopy studies of living samples, ranging from cells to fresh tissue.* Ultramicroscopy, 2007. **107**(4): p. 299-307.

26. Bianchi, M.E., *Interaction of a protein from rat liver nuclei with cruciform DNA.* EMBO J., 1988. **7**: p. 843-849.

27. Pinson, V., M. Takahashi, and J. Rouviere-Yaniv, *Differential binding of the Escherichia coli HU, homodimeric forms and heterodimeric form to linear, gapped and cruciform DNA.* J. Mol. Biol., 1999. **287**: p. 485-497.

28. Brazda, V., et al., *Cruciform structures are a common DNA feature important for regulating biological processes.* BMC Mol. Biol., 2011. **12**: p. 33.

29. Bayer, T.S., et al., *Arginine-rich motifs present multiple interfaces for specific binding by RNA.* RNA, 2005. **11**: p. 1848-1857.

30. Weiss, M.A. and N. Narayana, *RNA recognition by arginine-rich peptide motifs.* Biopolymers, 1998. **48**: p. 167-180.

31. Hentze, M.W. and T. Preiss, *The REM phase of gene regulation.* Trends Biochem. Sci., 2010. **35**: p. 423-426.

32. Castello, A., M.W. Hentze, and T. Preiss, *Metabolic enzymes enjoying new partnerships as RNA-binding proteins.* Trends Endocrinol. Metab., 2015. **26**: p. 746-757.

33. Nagy, E. and W.F. Rigby, *Glyceraldehyde-3-phosphate dehydrogenase selectively binds AU-rich RNA in the NAD(+)-binding region (Rossmann fold).* J. Biol. Chem., 1995: p. 2755-63.

34. Singh, R. and M.R. Green, *Sequence-specific binding of transfer RNA by glyceraldehyde-3-phosphate dehydrogenase.* Science, 1993: p. 365-8.

35. Nagy, E., et al., *Identification of the NAD(+)-binding fold of glyceraldehyde-3-phosphate dehydrogenase as a novel RNA-binding domain.* Biochem. Biophys. Res. Commun., 2000: p. 253-60.

36. Anderson, P.J., et al., *One pathway can incorporate either adenine or dimethylbenzimidazole as an alpha-axial ligand of B_{12} cofactors in Salmonella enterica.* J. Bacteriol., 2008. **190**: p. 1160-1171.

37. Escalante-Semerena, J.C., *Conversion of cobinamide into adenosylcobamide in bacteria and archaea.* J. Bacteriol., 2007. **189**: p. 4555-4560.

38. Hentchel, K.L. and J.C. Escalante-Semerena, *Acylation of biomolecules in prokaryotes: a widespread strategy for the control of biological function and metabolic stress.* Microbiol. Mol. Biol. Rev., 2015. **79**: p. 321-346.

39. Pellegrini, O., et al., *Overproduction and improved strategies to purify the threenative forms of nuclease-free HU protein.* Biochimie, 2000. **82**: p. 693-704.

40. Elbing, K. and R. Brent, *Media preparation and bacteriological tools*, in *Curr. Prot. Molec. Biol.*, F.M. Ausubel, et al., Editors. 2002, Wiley Interscience: New York. p. Unit 1.1.3.

41. Laemmli, U.K., *Cleavage of Structural Proteins during Assembly of Head of Bacteriophage-T4.* Nature, 1970. **227**(5259): p. 680-685.

42. Blommel, P.G. and B.G. Fox, *A combined approach to improving large-scale production of tobacco etch virus protease.* Protein Expr. Purif., 2007. **55**: p. 53-68.

43. Nečas, D. and P. Klapetek, *Gwyddion: an open-source software for SPM data analysis*, in *Open Physics*. 2012. p. 181.

44. Petr, K., et al., *Gwyscan: a library to support non-equidistant scanning probe microscope measurements.* Measurement Science and Technology, 2017. **28**(3): p. 034015.

45. Datsenko, K.A. and B.L. Wanner, *One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products.* Proc. Natl. Acad. Sci. USA, 2000. **97**: p. 6640-6645.

46. Berkowitz, D., et al., *Procedure for identifying nonsense mutations.* J. Bacteriol., 1968. **96**: p. 215-220.

47. Balch, W.E. and R.S. Wolfe, *New approach to the cultivation of methanogenic bacteria: 2-mercaptoethanesulfonic acid (HS-CoM)-dependent growth of Methanobacterium ruminantium in a pressurized atmosphere.* Appl. Environ. Microbiol., 1976. **32**: p. 781-791.

48. Rigaud, J.L., B. Pitard, and D. Levy, *Reconstitution of membrane proteins into liposomes: application to energy-transducing membrane proteins.* Biochim. Biophys. Acta, 1995. **1231**: p. 223-246.

Table 5.1. Strains and plasmids used in this study.		
Strain	**Relevant genotype**	**Reference/source[a]**
S. enterica serovar Typhimurium LT2 strains		
JE7088	*metE2702 ara-9*	Laboratory strain
Derivatives of JE7088		
JE11686	*cobB*$^{+}$ *cobT*$^{+}$/ pBAD30	Tucker 2010
JE12939	*ΔcobB1375 ΔcobT1380*	Tucker 2010
JE12972	*ΔcobB1375 ΔcobT1380*/ pBAD30	Tucker 2010
JE12974	*ΔcobB1375 ΔcobT1380*/ pCOBB19	Tucker 2010
JE12975	*ΔcobB1375 ΔcobT1380*/ pCOBB24	Tucker 2010
JE24046	*ΔcobB1375 ΔcobT1380*/ pCOBB45	This work
JE24047	*ΔcobB1375 ΔcobT1380*/ pCOBB46	This work
JE12998	*ΔcobB1375 ΔcobT1380*/ pCOBB47	This work
***E. coli* strains**		
E. coli C41 (λDE3)	*pka12*::*kan*$^{+}$ *ompT hsdS* ($r_B m_B$) *gal λ (DE3)*	Laboratory collection

Table 5.2. Primers and DNA probes used in this study.		
Primer Name	**Primer Sequence 5' → 3'**	
EMSA primers		
5' *tacA* EMSA primer (5' 6-FAM label)	ACGACGACTTGATGTATAC	
3' *tacA* EMSA primer	ACCCCTTGTATAGCATTTGC	
ssDNA, dsDNA, gap-2 DNA, cruciform DNA primers		
Primer A (5' 6-FAM label)	CCCTATAACCCCTGCATTGAATTCCAGTCTGATAA	Bianchi, 1988 #26705;
Primer B	TAGTCGTGATAGGTGCAGGGGTTATAGGG	Bianchi, 1988 #26705;
Primer C	ACAGTAGCTCTTATTCGAGCT CGCGCCCTATCACGACTA	Bianchi, 1988 #26705;
Primer D	TTATCAGACTGGAATTCAAGCGCGAGCTCGAATAAGAGCTACTGT	Bianchi, 1988 #26705;
Primer E	TTATCAGACTGGAATTCAATGCAGGGGTTATAGGG	
Primer F	TACGGATCGCAGC	
Primer G	GGTTAGGGAAGTTGG	Pinson 1999
Primer H (3' 6-FAM label)	CCAACTTCCCTAACCCAGCTGCGATCCGTA	

Table 5.3. Plasmids used in this study			
Plasmid	**Genotype**	**Description**	**Source**
pBAD30	*araC*$^+$ *bla*$^+$	P_{araBAD} expression vector	Guzman 1995
pTEV6	*bla*$^+$	N-terminal rTEV-cleavable MBP-His fusion overexpression vector	Rocco 2008
pCOBB19	*cobB*$_L^+$ *bla*$^+$	*cobB1372* allele (CobB$^{M37A\ M38A}$) in pBAD30	Tucker 2010
pCOBB24	*cobB*$_s^+$ *bla*$^+$	*cobB1373* allele (CobB^{M1A}) in pBAD30	Tucker 2010
pCOBB45	*cobB*$_L^{R11A\ R26A}$ *bla*$^+$	CobB$^{M37A\ M38A\ R11A\ R26A}$ in pBAD30	This work
pCOBB46	*cobB*$_L^{R11A\ R17A\ R26A}$ *bla*$^+$	CobB$^{M37A\ M38A\ R11A\ R17A\ R26A}$ in pBAD30	This work
pCOBB47	*cobB*$_L^{R11A\ R17A\ R22A\ R26A}$ *bla*$^+$	CobB$^{M37A\ M38A\ R11A\ R17A\ R22A\ R26A}$ in pBAD30	This work
pCOBB72	*cobB*$^+$ *bla*$^+$	CobB ORF in pTEV6	Tucker 2010
pHUPB5	*hupB*$^+$ *bla*$^+$	*hupB* cloned into T7-7 overexpression vector	Laboratory plasmid
pBAD24	*araC*$^+$ *bla*$^+$	P_{araBAD} expression vector	Guzman 1995

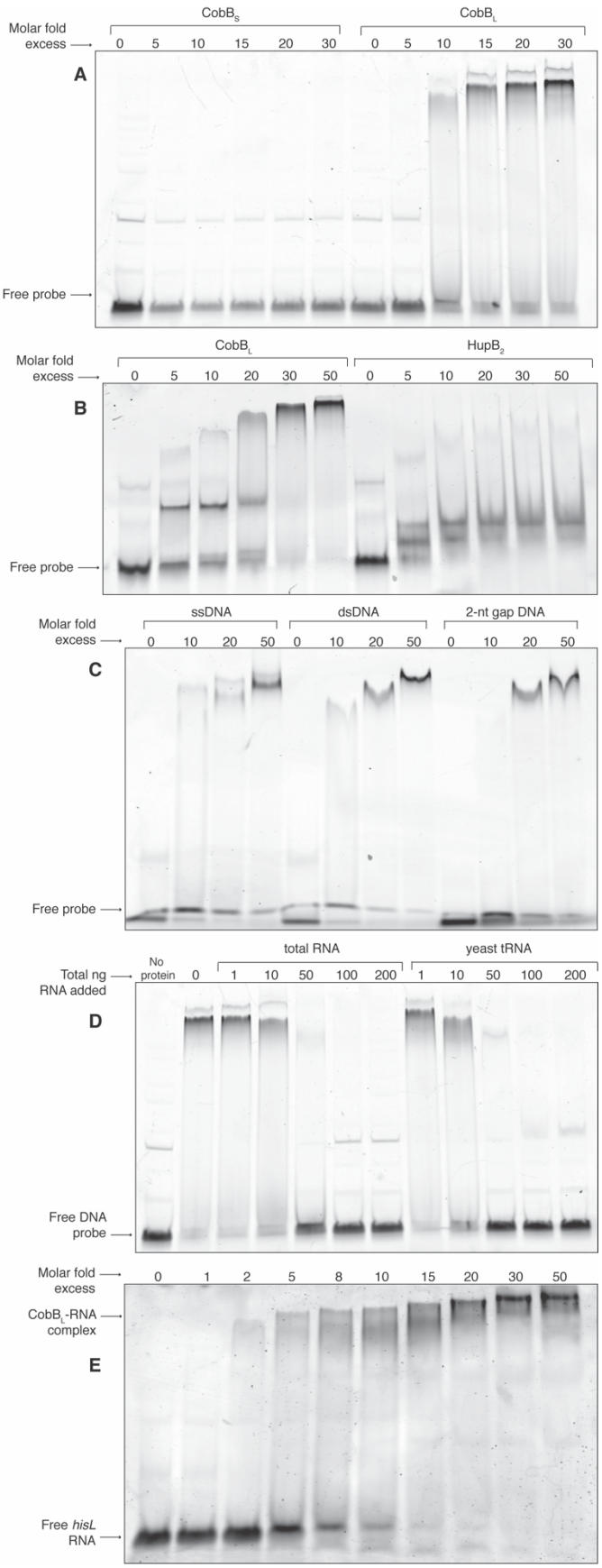

Figure legend on next page.

Figure 5.1. Electrophoretic mobility shift assays. **A**. *$CobB_L$ binds to dsDNA, $CobB_S$ does not*. The ability of the CobB isoforms to bind to DNA was assessed via electrophoretic mobility shift assay using a 6-FAM 5'-labeled dsDNA probe. Labeled promoter DNA for the *S. enterica* gene *tacA* (*stm3652*, 75 bp, 1.01 pmol) was incubated with increasing concentrations of $CobB_S$ (left) or $CobB_L$ (right) (5.05, 10.1, 15.1, 20.2, 30.3 pmol protein to DNA) and resolved by non-denaturing PAGE. As a control labeled DNA was incubated alone (0-fold excess protein). **B**. *$CobB_L$ binds to ssDNA, gapped DNA and cruciform DNA.* EMSA was performed with either 6-FAM labeled single-stranded DNA (35-bp, 2.16 pmol), 6-FAM-labeled annealed double-stranded oligonucleotides of linear dsDNA (35-bp, 2.16pmol) or dsDNA with a 2-nt gap (30-bp, 2.52pmol). $CobB_L$ was added in molar fold-excess of protein to DNA (for ssDNA and 35-nt dsDNA, $CobB_L$ was added at 21.6, 43.2, and 108.2 pmol protein, and for 2-nt gap DNA $CobB_L$ was added at 25.2, 50.5, and 126.2 pmol protein to DNA). **C.** *$CobB_L$ binds to cruciform DNA.* EMSA was performed with 6-FAM-labeled cruciform DNA (151-bp, 0.5 pmol) and either $CobB_L$ or $HU\beta_2$ protein at 0, 2.5, 5, 10, 15, and 25 pmol protein to DNA. $HU\beta_2$ was included as a positive control for cruciform binding. **D.** *RNA competes with DNA for $CobB_L$ binding.* To each reaction mixture, *tacA* promoter DNA (75-bp, 1.01 pmol, 50 ng) and $CobB_L$ protein (20.2 pmol) was added except in the *tacA* DNA only control. A sweep of unlabeled total RNA from *S. enterica* or unlabeled yeast tRNA was added (1, 10, 50, 100, 200 ng). **E.** *$CobB_L$ binds to hisL RNA.* EMSA was conducted where unlabeled *E. coli hisL* RNA (manufactured from IDT Technologies, 51-bp, 5.9 pmol) was incubated with either 0, 5.9, 11.8, 29.7, 47.5, 59.4, 89.1, 118.8, 179.2, 297.0 pmol of $CobB_L$ protein to RNA.

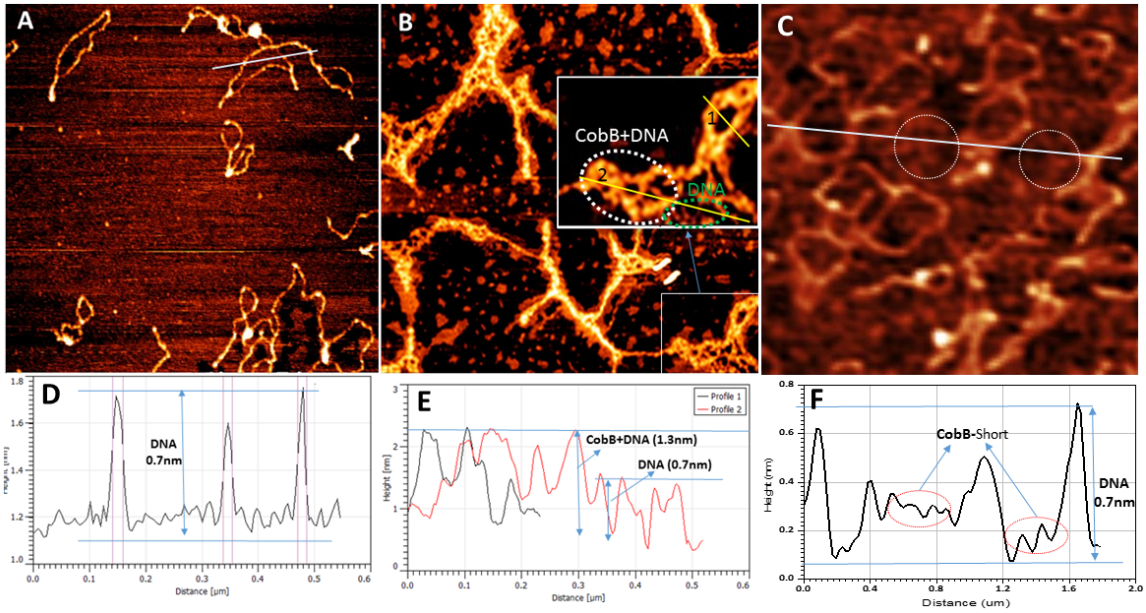

Figure 5.2. AFM images of DNA (A), CobB-DNA complex under (B), and CobB (short)-DNA complex (C) under TopMAC mode (size: 1.8um X 1.8um). Insert in B shows a zoomed in area as highlighted by the white box. (D) Measured height profile along the white line in (A) of bare DNA molecule. (E) Two height profiles along the yellow line labeled 1 and 2 of the DNA-CobB complex. (F) Measured height profile along the white line in (C) of CobB (short)-DNA complex.

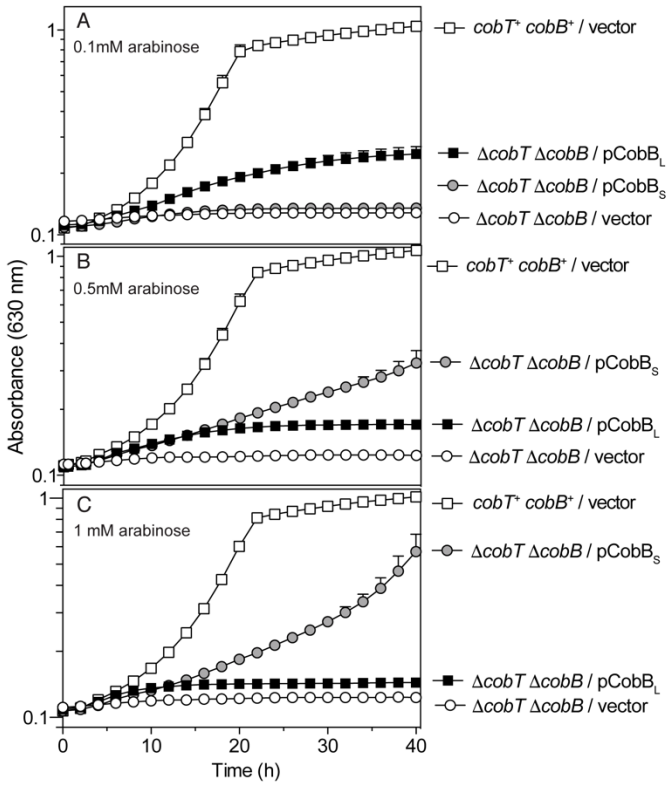

Figure 5.3. Growth analysis when synthesis of CoB_{12} is required and ethanolamine is the sole carbon and energy source. Genes encoding either $CobB_L$ or $CobB_S$ in arabinose-inducible complementation vectors and were electroporated into a Δ*cobT* Δ*cobB* background as indicated by the following symbols: Δ*cobT* Δ*cobB* / vector, white circles; Δ*cobT* Δ*cobB* / $pCobB_L$, black squares; Δ*cobT* Δ*cobB* / $pCobB_S$, grey circles; *cobT*$^+$ *cobB*$^+$ / vector, white squares. Cells were grown on NCE minimal medium supplemented with $(CN)_2$Cbi (dicyanocobinamide), DMB (5,6-dimethylbenzimidazole), and ethanolamine (90mM). Growth behavior of *S. enterica* was assessed for the ability of strains with CobB isoforms *in trans* to grow and synthesize CoB_{12}. **A**. *Only $CobB_L$ conferred growth on ethanolamine with $(CN)_2$Cbi + DMB under low L-(+)-arabinose induction (0.1mM).* **B** and **C**. *Only $CobB_S$ conferred growth ethanolamine with $(CN)_2$Cbi + DMB under high arabinose induction (0.5mM or 1mM).* All growth curves were conducted three times in technical triplicate of biological triplicates using a microtiter plate reader (biotek instruments). Error bars represent standard deviations (SD) of replicates.

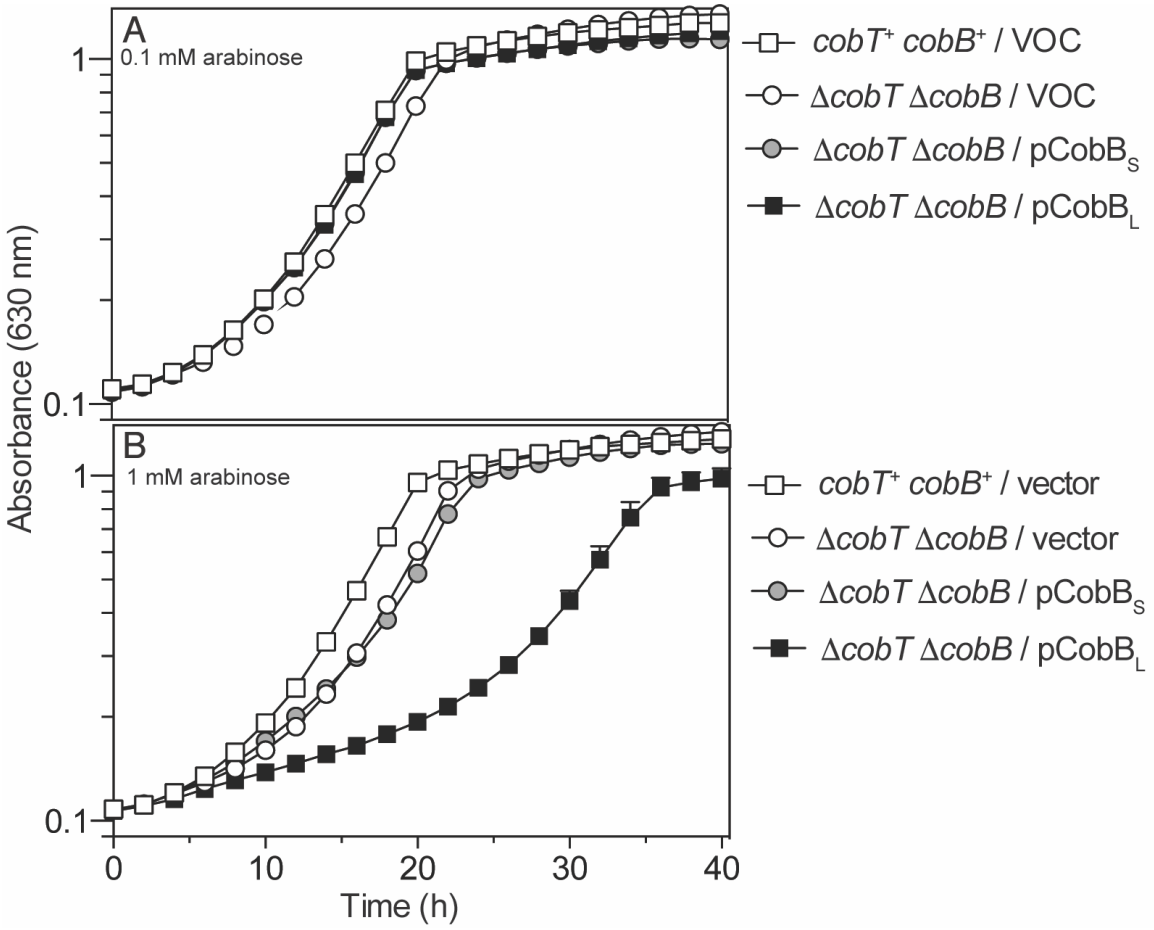

Figure 5.4. Growth analysis when CoB_{12} is supplied and ethanolamine is the sole carbon and energy source. Growth analysis was conducted of strains of *S. enterica* indicated by the following symbols: Δ*cobT* Δ*cobB* / vector, white circles; Δ*cobT* Δ*cobB* / $pCobB_L$, black squares; Δ*cobT* Δ*cobB* / $pCobB_S$, grey circles; *cobT*$^+$ *cobB*$^+$ / vector, white squares. Cells were grown on NCE minimal medium supplemented with CoB_{12} (150μM) and ethanolamine (90mM) was the sole carbon and energy source. A. *No phenotypes are observed during low L-(+)-arabinose induction (0.1mM).* **B**. *High arabinose induction (1mM)* of *$CobB_L$ causes a growth phenotype when CoB_{12} is supplied during growth on ethanolamine.* All growth curves were conducted three times in technical triplicate of biological triplicates using a microtiter plate reader (biotek instruments). Error bars represent standard deviations (SD) of replicates.

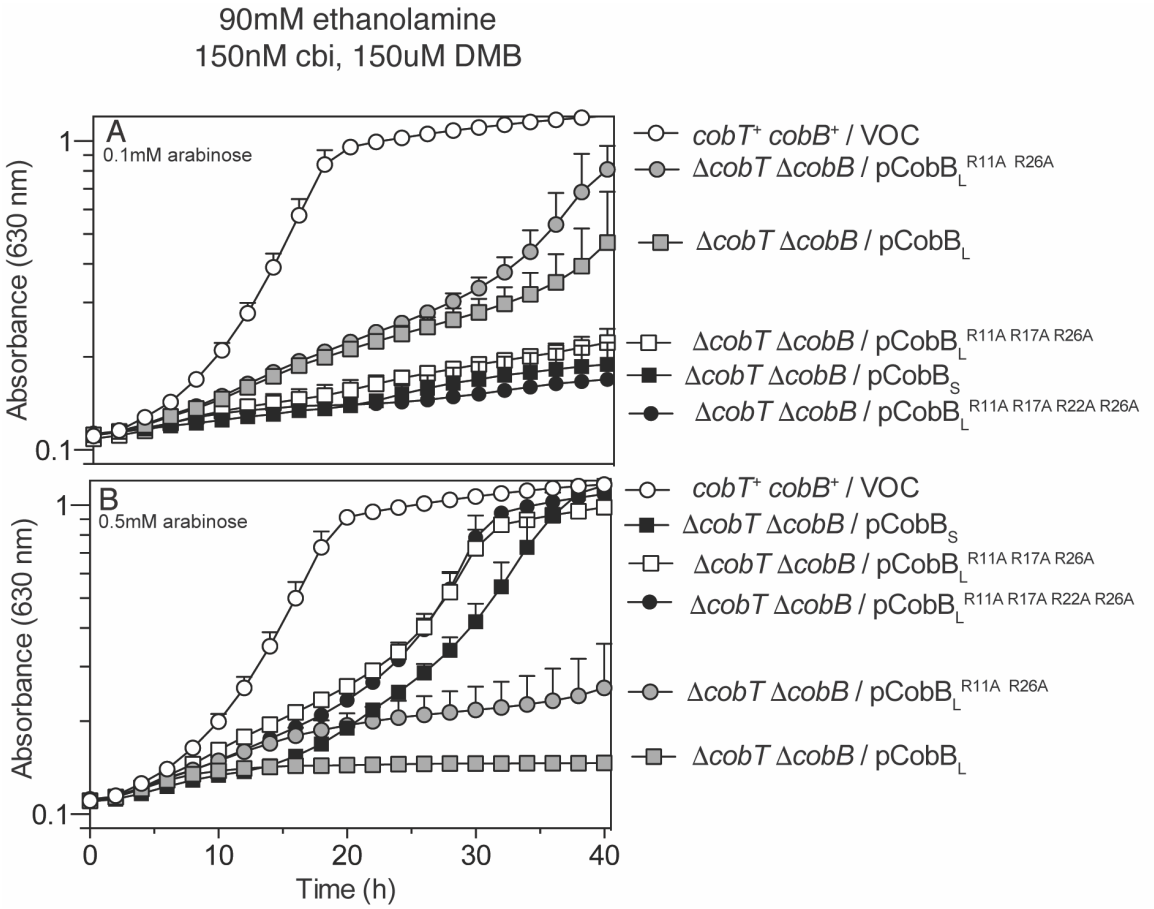

Figure 5.5. Growth analysis of arginine variants of $CobB_L$ when synthesis of CoB_{12} is required during growth with ethanolamine. Cells were grown as figure 3 but with *N*-terminal arginine variants of CobBL. Arg to Ala $CobB_L$ variants mimic $CobB_S$ only growth as compared to cells with only $CobB_L$

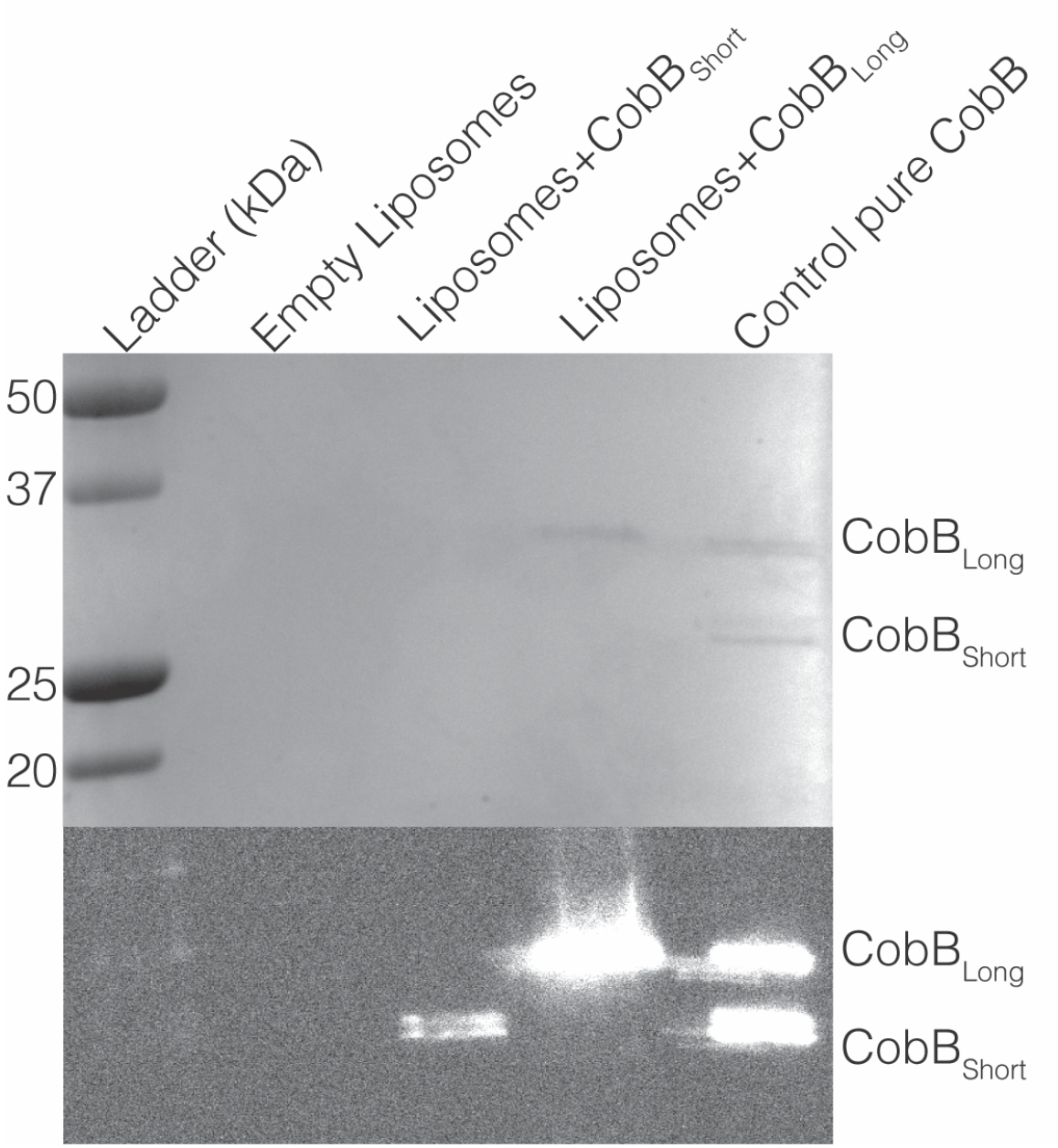

Figure 5.6. Western blot analysis of CobB-Liposome flotation assay. CobB isoform proteins were incubated with prepared liposomes then subjected to ultracentrifugation. Liposomes were isolated from a Histodenz gradient and assayed for the presence of CobB protein using anti-CobB antibodies.

CHAPTER 6

NUCLEOID-ASSOCIATED PROTEIN HU IS N-TERMINALLY ACETYLATED BY THE N-ALPHA ACETYLTRANSFERASE NATA IN S. ENTERICA[5]

[5]Parks A.R. and Escalante-Semerena J.C. To be submitted to *PNAS*.

6.1 ABSTRACT

Here we report that the *Salmonella enterica* NatA protein N-terminal acetyltransferase acetylates the N-terminal methionine of the nucleoid-associated proteins HU. Our findings are supported by in vitro analysis of acetylation of the HU proteins and lysine-deficient variants, as well as in vivo analysis of the effect of acetylation on HU-mediated transcriptional regulation of a known target of HU, the *hilA* promoter. We also found that *Se*NatA is unable to acetylate initiating methionines of HU that have been oxidized to methionine sulfoxide, and that reduction of these methionine sulfoxide residues restores acetylation of HU proteins by *Se*NatA. This demonstrates that the *Se*HU proteins are *bonafide* substrates for the methionine sulfoxide reductases MsrA and MsrB. Finally, our data suggest that the *Bacillus subtilis* acetyltransferase, YfmK, is not a protein-lysine acetyltransferase, but a functional homologue of *Se*NatA, and that *Bs*YfmK acetylates the *B. subtilis* HU protein, HBsu, on its N-terminal methionine and not on various lysyl residues.

6.2 INTRODUCTION

N-terminal acylation, or acylation of the free amino moiety of the leading amino acid in a peptide chain is a common occurrence in eukaryotic cells and serves a wide variety of purposes, such as targeting proteins for degradation [1-4], subcellular localization [5-8], or protein folding [9-11] and protein-protein

interactions [12-17]. In prokaryotes, however, the extent and purpose of N-terminal acylation of proteins has just started to be realized. Recently, several studies have taken an acetylomics approach at addressing the identification of *N*-terminally acylated proteins and acylation abundance in prokaryotes [18, 19].

Here we took a targeted approach to advancing our understanding of the physiological impact of *N*-terminal acetylation of proteins in *S. enterica*. For this purpose, we used the recently discovered protein *N*-terminal acetyltransferase (NatA) of this bacterium. We recently reported that NatA (hereafter *Se*NatA) modulates the activity of the CobB sirtuin deacylase long isoform (hereafter $CobB_L$) [20]. Our search for other protein substrates for *Se*NatA identified the homodimers of HU, $HU\alpha_2$ and $HU\beta_2$ as substrates of *Se*NatA.

In *S. enterica*, HU proteins are small nucleoid-associated proteins (NAPS) that form three different types of oligomers, *i.e.*, $HU\alpha_2$ and $HU\beta_2$ homodimers, or the $HU\alpha\beta$ heterodimer. These relatively non-specific DNA binding proteins are present in the cells in large amounts and influence the expression of a large number of genes through nucleoid compaction and organization [21]. The abundance of the different HU oligomers changes as a function of the growth phase of the cell [22], with Lon protease involved in the degradation of the $HU\beta_2$ to allow dimeric form exchange to take place [23].

It was recently published that the GCN5 acetyltransferase YfmK could acetylate HBsu in *Bacillus subtilis* [24]. Acetyltransferases possessing the GCN5-related *N*-

acetyltransferase domain (GNAT domain) are known to have low sequence identity to each other, making elucidation of homologues or their physiological roles difficult. Case in point, an alignment of *Se*NatA and *Bs*YfmK shows only a 21.1% percent identity between the two acetyltransferases, but a large range of amino acid conservation (Figure 6.1A).

Here we propose that *Se*NatA and *Bs*YfmK are functional homologues and that *Bs*YfmK is in fact a protein *N*α acetyltransferase, that *Se*NatA and *Bs*YfmK *N*-terminally acetylate HU in *S. enterica* and *B. subtilis*, respectively, and that the acetylation of the *N* terminus of HU alters HU-regulated gene expression.

6.3 RESULTS

***Se*NatA acetylates all three forms of HU.** During screening experiments for potential substrates for putative GCN5-related *N*-acetyltransferases (*i.e.*, GNATS) we found that the *S. enterica* NatA protein acetylated the homodimeric HUβ_2 form of this bacterium (hereafter *Se*HUβ_2). Further analysis of the HU isoforms led us to discover that *Se*NatA was also able to acetylate homodimeric form *Se*HUα (Figure 6.2). We also found that *Se*NatA acetylated the *Bacillus subtilis* homologue of HU, HBsu (Figure 6.2, lane 9). We further questioned the ability of *Bacillus subtilis* YfmK (hereafter *Bs*YfmK) to acetylate the *Se*HU isoforms because of the high sequence similarity between *Se*HU proteins and *Bs*HBsu (Fig. 6.1B). As shown in Figure 6.3, *Bs*YfmK is able to acetylate HBsu, *Se*HUα_2, *Se*HUβ_2 proteins. During our initial screens, we found that acetylation signal seen of *Se*HU$\alpha\beta$ by *Se*NatA and *Bs*YfmK was very weak to non-existent from experiment

to experiment. Initially we believed that the heterodimer of HU was not acetylatable by SeNatA, however further investigation into the differences in purification of the *Se*HU$\alpha\beta$ protein complared to the homodimers or HBsu was inspected. A special note is made to the method of purification of the HU proteins; the *Se*HU$\alpha\beta$ protein was purified following the Pellegrini et. al. protocol [25] using 20 liters of our lab parent strain of *S. enterica* LT2 harboring no plasmids. This protocol relies on the fact that the HU proteins are among the most abundant proteins in the cell and purifies and separates out the three biological isoforms of *Se*HU (for full details refer to material and methods section). To acquire homogeneous *Se*HUα_2, *Se*HUβ_2, or HBsu protein, a *hupA hupB* deletion strain of *S. enterica* harboring a chromosomally encoded T7 polymerase, was complemented with a T7-7 overexpression vector encoding for either, *hupA*, *hupB* or *hbsu*. The results of our investigation are described below.

Several variations to *Se*HU protein N-termini exist *in vivo*

To confirm the location of modification of homodimers of *Se*HU proteins by *Se*NatA, we needed an alternative protease to use instead of trypsin, which cleaves C-terminal to lysyl or arginyl residues, in order to resolve peptides of the arginine and lysine rich *N*-termini of the HU proteins. Endoproteinase Asp-N cleaves peptide bonds *N*-terminal to aspartic acid residues, and in the case of HU digestion yielded peptides long enough for resolution of chemical modifications to be seen via LC-MS/MS [26, 27] (Fig. 6.4). Preliminarily, we digested the natively purified, without plasmid, *Se*HU$\alpha\beta$ protein by Asp-N digestion and LC-MS/MS to

get a baseline of N-terminal peptide coverage. We discovered several variations to the *N*-termini of the native *Se*HU proteins (Fig. 6.5, 6.6). First, a population of the HU proteins in the heterodimer are missing their initiator N-terminal methionine (iMet). Second, a portion of the HU proteins possess the iMet residue, but the methionine has been oxidized to methionine sulfoxide. Finally, a portion of the population possess the iMet without having an oxidized methionine. These results are surprising due to the amino acid sequence of the N-terminus of HU proteins not being an optimal sequence for iMet cleavage by the *S. enterica* methionine aminopeptidase enzyme, MAP[28] as well as no prior evidence that HU proteins undergo iMet excision.

We next subjected our plasmid-overexpressed but also natively purified homodimeric SeHU proteins to Asp-N digestion and LC-MS/MS analysis and discovered that, based on the restriction sites chosen to clone into the plasmid encoding the overexpression of the *hupA* or *hupB* genes for protein purification, an extra N-terminal methionine residue is added to the resulting protein (Fig. 6.8 for SeHUα_2 LC-MS/MS peaks, Fig. 6.9 for SeHUβ_2 LC-MS/MS peaks, corresponding plasmids are pHUPA3 and pHUPB5). In order to verify that *Se*NatA is able to acetylate the native N-termini of *Se*HU proteins, we performed a deletion mutagenesis of the extra methionine on the pT7-7 plasmids encoding *hupA* or *hupB* and repurified them to homogeneity. Refer to figure 6.7 for a diagram of the multiple cloning site for pT7-7. The purified proteins were re-tested for *Se*NatA mediated acetylation, and, just as observed for the *Se*HU$\alpha\beta$ native proteins,

inconsistent acetylation of homodimers with single iMet residues was observed (Fig. 6.10, lanes 4, 5, and 6).

***Se*NatA cannot acetylate *N*-terminal methionine sulfoxide residues.**

We speculated as which population of N-termini are unable to be acetylated in the SeHU proteins. We first decided to ask if SeNatA is unable to acetylate N-terminal methionines (Met) that have been oxidized to methionine sulfoxide (Met-SO) by radical oxygen species (ROS), as a subpopulation of *Se*HU proteins was seen to have Met-SO N-termini by LC-MS/MS. In *S. enterica*, peptide-bound Met-SO is reduced to Met through the enzymatic activity of the Msr methionine sulfoxide reductases. When Met is oxidized to Met-SO, two enantiomers are generated, (*S*)-Met-SO and (*R*)-Met-SO (Fig 6.11). To see if reduction of the iMet-SO residue led to increased acetylation by SeNatA, we purified the *Salmonella* MsrA and MsrB proteins and pre-incubated the Msr proteins with the *Se*HU proteins and DTT, then conducted the radiolabeled acetylation assay with *Se*NatA. As shown in figure 6.12, MsrA and MsrB were able to reduce *Se*HU iMet-SO to iMet, which allowed increased radiolabel transfer signal to be seen for the *Se*HU proteins. This suggests that *Se*NatA is unable to acetylate *N*-termini with iMet-SO residues. This also demonstrates that MsrA and MsrB can use *Se*HU proteins as substrates.

***Se*NatA and *Bs*YfmK acetylate lysine-null variants of *Se*HU and HBsu proteins *in vitro*.** Carabetta *et. al.* recently published that *Bs*YfmK acetylated seven lysyl groups of *B. subtilis* HBsu. Supporting Information figure S16 of the

Carabetta paper shows that mass spectrometry analysis did not unambiguously establish whether Lys3 or the *N*-terminus were acetylated based on protease peptide sequence coverage. Based on our knowledge of *Se*NatA function as an *N*α-acetyltransferase (NAT) [20], we hypothesized that the *N*-terminus of HBsu was the true site of enzymatic acetylation by *Bs*YfmK (and *Se*NatA), and that the lysyl-residues were chemically acetylated during the protein purification process due to the ability of acetyl-CoA to nonenzymatically react with lysyl residues that have a reduced pK value [29-31]. To further confirm that the lysines of *Se*HUβ_2 were not being enzymatically acetylated by SeNatA, we generated a completely lysine null variant of *Se*HUβ_2, where all nine lysyl residues were changed to arginine. These lysine null variants of *Se*HUα_2, *Se*HUβ_2, and *Bs*HBsu were subjected to *in vitro* acetylation using homogenous *Se*NatA or *Bs*YfmK proteins. As shown in figure 6.13, the lysine null variants of all HU proteins were still acetylated by both acetyltransferases.

***N*-terminally acetylated *Se*HU proteins cannot be deacetylated by the *S. enterica* CobB sirtuin deacylase.** A common feature of *Na* acetylation is that is not reversible. To support our conclusion that *Se*NatA and *Bs*YfmK acetylated the *N*-terminus of *Se*HU proteins, we conducted *in vitro* deacetylation assays using radiolabeled, acetylated HU (^{*Ac}HU) proteins as substrates for the *S. enterica* CobB short isoform [32]. ^{*Ac}HUβ, ^{*Ac}HUα, and ^{*Ac}HU$\alpha\beta$ proteins acetylated by *Se*NatA, and ^{*Ac}HBsu protein acetylated by *Bs*YfmK were incubated with $CobB_s$ in the presence or absence of NAD^+. To ensure that the CobB proteins were functional,

CobB was incubated with radiolabeled Acs (Acs^{Ac*}), a *bonafide* substrate of CobB, in the presence or absence of NAD^+. As seen in figures 6.14 and 6.15, $CobB_S$ deacetylated Acs^{Ac*} (lane 3), but did not deacetylate $^{*Ac}HU\beta_2$, $^{*Ac}HU\alpha_2$, $^{*Ac}HU\alpha\beta$ (panel A), or ^{*Ac}HBsu.

***N*-terminal acetylation of *Se*HU proteins alters *Se*HU-dependent gene regulation.** After establishing that *Se*NatA acetylated *Se*HUβ_2, *Se*HUα_2, and HU$\alpha\beta$, we sought to determine the effect of acetylation on *Se*HU-dependent regulation of gene expression *in vitro* and *in vivo*. To answer this question *in vitro*, we performed electrophoretic mobility shift assays (EMSAs) to quantify changes in DNA-binding of acetylated HU proteins to *bona fide* DNA substrates. *Se*HUα_2 protein was pre-incubated with SeNatA, or *Se*HUα_2 protein that had been pre-acetylated by *Se*NatA with AcCoA, were incubated with 5' 6-FAM-labeled *hilA* promoter DNA probe. The *hilA* promoter was chosen as a DNA probe for *Se*HUα_2 protein binding because it has been published that *Se*HU modulates the expression of the *hilA* gene in *S. enterica* [33]. As seen in figure 6.16, *Se*HUα_2 protein pre-incubated with *Se*NatA and AcCoA do not bind as robustly at 10- or 15-fold excess of protein to probe as the *Se*HUα_2 protein pre-incubated with *Se*NatA only. This suggests that *Se*HUα_2 protein is being acetylated by SeNatA, and that this acetylation affects the ability of *Se*HUα_2 to bind to the *hilA* promoter DNA.

In vivo evidence. It has been established that *Se*HU aids in transcriptional activation of master transcriptional regulator for invasion genes on SPI-1 in *S. enterica,* HilA [33]. To test whether NatA alters HU-mediated *hilA* expression in *S. enterica in vivo*, we monitored the expression of a chromosomal *hilA*::*lacZ* transcriptional fusion as a function of the ectopic expression of *Se*NatA. As expected, in a *hupB*$^+$ *hupA*$^+$ strain grown under conditions that triggered the expression of SPI-1 (*i.e.*, high salt, low oxygen), *Se*HU positively affected *hilA* expression [33, 34]. However, when *Se*NatA was ectopically expressed, the beta-galactosidase units produced by the *hilA-lacZ* fusion decreased significantly (Figure 6.17). When the *hupA* gene was deleted, *hilA* expression decreased to levels seen in the *hupA*$^+$ *hupB*$^+$ strain expressing *Se*NatA, and when *Se*NatA was overexpressed in the *hupA* deletion, *hilA* expression level further decreases, suggesting a decrease in activation of *hilA* by *Se*HUβ_2 in the presence of *Se*NatA. When *hupB* is deleted, the *hilA* expression drops to levels that mimic the the *hupA* deletion or the *hupA*$^+$ *hupB*$^+$ strain expressing *Se*NatA. However, when *Se*NatA was overexpressed in a *hupB* deletion where *Se*HUα_2 is the only HU able to induce *hilA* expression, we unexpectedly saw an increase in the expression of *hilA*. This result is unexpected, especially since figure 6.16 shows that acetylated *Se*HUα_2 protein displays a decreased ability to bind to the *hilA* promoter *in vitro*. Finally, *hilA* expression in a *hupA hupB* double deletion was not different than a double deletion with *Se*NatA overexpressed, suggesting that *Se*NatA cannot alter *hilA* expression in the absence of the *hup* genes under the condition tested (Figure 6.17).

6.4 DISCUSSION

The finding that SeHU proteins undergo multiple N-terminal modifications, including iMet excision, methionine oxidation, and N-terminal acetylation is novel and sheds light on the importance of N-terminal modification in modulation of proteins. At first the finding that SeHU proteins have iMet cleavage was surprising due to in vitro analysis of the MAP enzyme yielding results that suggest that proteins with a P2' residue of asparagine are unfavorable substrates for iMet cleavage. However, further analysis into N-terminomics in bacteria revealed that biochemical characterization of MAP activity does not perfectly reflect in vivo iMet cleavage, where bioinformatic analysis across several species of bacteria shows that cleavage of iMet with Asn as the penultimate residue is a common occurrence and physiologically relevant [35]. This is also consistent with the yeast MAP, yMetAP1, that is active against Met-Asn *N*-termini [36].

We also found that those SeHU proteins that retain their iMet are susceptible to N-terminal methionine oxidation, converting methionine to methionine sulfoxide. We demonstrate that the HU proteins are bonafide substrates for the methionine sulfoxide repair enzymes MsrA and MsrB in *S. enterica*. Perhaps Msr repair of iMet oxidation mediates NatA acetylation, suggesting a role for reduced iMet N-termini and/or acetylation in cellular response to oxidative stress. From our previous study of acetylation of $CobB_L$ by *Se*NatA, we found that *Se*NatA can accommodate up to two extra residues on the N-terminus of $CobB_L$ and still recognize it as an

acetylatable substrate [20]. However, we are unsure if *Se*NatA is able to acetylate SeHU lacking the iMet residue, in this case with N-termini that begin with Asn. Further analysis of SeNatA substrate recognition is required.

In contrast to Carabetta et. al. that published that BsYfmK is a protein lysine acetyltransferase and acetylates HBsu on multiple lysyl residues [24], we found that BsYfmK is able to acetylate the *S. enterica* HU proteins even when no lysines are present. This suggests that BsYfmK is actually a protein N-α acetyltransferase and the previous findings were in vitro artifacts due to the lability of acetyl-CoA at pH 8 (The pH used in the previous study), or that BsYfmK is able to perform both lysine and N-terminal acetylation. Further work needs to be conducted to confirm the flexibility of amine substrates for BsYfmK.

6.5 MATERIALS AND METHODS

Bacterial strains and growth conditions. All strains constructed were derivatives either *Salmonella enterica enterica* sv Typhimurium LT2 or *Salmonella enterica enterica* sv Typhimurium ATCC 14028s strains (*S. enterica* throughout the paper) (Table 6.1). Genes were deleted using described protocols [37]. Bacterial cultures were grown in Luria Bertani broth (LB) at 37ºC with shaking unless stated otherwise.

Strain construction. The HU deletion strains were constructed as follows. Pfu Ultra II DNA polymerase (Stratagene) was used to amplify flanking regions of the plasmid pKD3 [37] fused to 36 to 39 bp of internal to the beginning or end of the *hupA* or *hupB* genes. PCRs were resolved on a 1% (w/v) agarose gel and checked for amplification by post-staining with ethidium bromide (0.5 μg /ml) for 10 min. Polymerase chain reaction (PCR) fragments were cleaned with the Wizard SV gel and PCR cleanup kit (Promega), and ~200 ng of product was electroporated into *S. enterica* strain JE6692 (*metE205 ara-9* / pKD46) using a 0.2-cm electroporation cuvette (MidiSci) and a microPulser electroporator (Bio-Rad Laboratories) on Ec2 setting. Cells were incubated at 37ºC with shaking, plated on lysogeny broth (LB) supplemented with 12.5 μg/ml of chloramphenicol, and incubated at 37ºC overnight. Insertion of *cat*$^{+}$ was confirmed by PCR, then the marker and deletion were moved by P22-mediated transduction into strain JE6583 as described elsewhere [38]. Colonies were streaked to isolation several times under antibiotic selection and phage sensitivity was tested. The deletion of the gene was confirmed by Sanger sequencing using primers flanking the gene of interest.

Plasmid construction for complementation and overexpression. Plasmids used in this study are listed in Table 6.2. Plasmids were constructed as described previously [20]. Lysine residues for the lysine null *Se*HUβ$_2$ were changed by site-directed mutagenesis of hupB on plasmid pHUPB5. Primers used for this purpose were designed using Stratagene's QuikChange Primer Design software package (https://www.agilent.com/store/primerDesignProgram.jsp). PCR was conducted using Pfu Ultra II DNA polymerase (Agilent) with primers listed in Table 6.3.

Modification to this polymerase protocol included an extension temperature of 68ºC and extension time of 2.5 min per kb. Mutations were confirmed by Sanger sequencing. The *Se*HUα_2 lysine null construct was ordered through GenScript and cloned into pT7-7 using NdeI (5') and XbaI (3') cut sites.

Purification of *Se*NatA and *Bs*YfmK, *Se*HUα_2, *Se*HUβ_2, *Se*HU$\alpha\beta$ proteins.

***Se*NatA and *Bs*YfmK.** Plasmid pYIA6 and pYfmK1 were transformed into *E. coli* C41(λDE3) Δ*pat* (strain JE9314). Overnight cultures of *E. coli* harboring either pYIA6 or pYFMK1 were subcultured 1:100 into 1 liter each of Terrific Broth (TB; 12g/L tryptone, 24g/L yeast extract, 4mL/L glycerol, 2mM $MgCl_2$ buffered with 0.17M KH_2PO_4 and 0.72M K_2HPO_4) containing 100μg/ml ampicillin. Cells were growing shaking at 37°C to an OD600 nm of 0.5, induced with 0.5mM IPTG (isopropyl b-d-1-thiogalactopyranoside), then allowed to grow shaking at 37°C for 18 hours. Filtered crude extracts were applied to a 1ml HisBind FF column (GE Healthcare) pre-equilibrated with 50mM HEPES pH 7.5, 500mM NaCl, 20mM imidazole, 5% glycerol, 1mM TCEP (tris(2-carboxyethyl)phosphine hydrochloride) using an ÄKTA fast protein liquid chromatography (FPLC) system (GE Healthcare). The column was washed with Buffer A and proteins of interest were eluted with a linear NaCl gradient. Fractions containing each protein (as determined by SDS-PAGE) were pooled and dialyzed in Buffer A with 1mM DTT for 3 hours at room temperature with 10 mg : 1 mg ratio of MBP-6xHis-NatA or MBP-6xHis-YfmK to 6xHis-rTEV protease. Cleaved proteins were dialyzed twice more at 4°C in Buffer A then reapplied to a 1ml HisBind FF column to separate cleaved proteins from MBP-6xHis and rTEV protease. Cleaved proteins were

collected in the flow through, analyzed by SDS-PAGE, pooled and dialyzed in storage buffer of 20mM HEPES pH 7.5, 150mM NaCl, 1mM TCEP, 20% glycerol.

Native HU heterodimer purification. The heterodimer of HU was natively purified according to Pellegrini et. al. [25] as follows: overnight cultures *Salmonella* strain JE6583 grown in lysogeny broth (LB) were subcultured 1:100 into 20 liters of Terrific Broth (TB; 12g/L tryptone, 24g/L yeast extract, 4mL/L glycerol, 2mM MgCl2 buffered with 0.17M KH2PO4 and 0.72M K2HPO4) containing 100ug/ml ampicillin. Cells were grown shaking at 37C until late-logarthmic phase when the HU heterodimer is most abundant compared to homodimers. Cells were pelleted at 6,000 x g for 15 min in a Beckman Coulter Avanti J-2 XPI centrifuge with a JLA-8.1000 rotor. The resulting 150 g cell pellet was resuspended in 900 mL of Buffer A (Tris-HCl pH 7.8 (20 mM), NaCl (1.7 M), ethylenediaminetetraacetic acid (EDTA, 1 mM), 2-mercaptoethanol (1 mM)) and was lysed by passage through a TS-Series 0.75kW French press cell disruptor (Constant Systems) at 1200 bar and 20,000 psi. Cell debris was then pelleted for 30 min at 6,000 x g in the centrifuge listed above and cleared supernatant was saved. Polyethylene glycol (PEG, 4000 dalton) was added to the cell-free extract at a final concentration of 15% (w/v) and was mixed slowly for 2 hours at 4°C. Cell-free extract was centrifuged again for 30 min at 12,000 x g. Supernatant after centrifugation was dialyzed over the span of 24 h three times in 2 liters each dialysis of Buffer B (Tris-HCl pH 7.8 (20 mM), NaCl (200 mM), EDTA (1 mM)). Dialyzed cell-free extract (CFE) was centrifuged for 20 min at 12,000 x g, then applied to a 20 mL HiPrep Heparin FF 16/10 attached to

an ATKA fast protein liquid chromatography (FPLC) system (GE Healthcare), 150-200 mL of CFE applied per run (3 runs total). A linear elution gradient from 200 mM NaCl to 1M NaCl in buffer B was applied; the HU heterodimer eluted at approximately 0.65M NaCl. The HiPrep Heparin FF 16/10 column was cleaned in between runs with 80 mL of 0.1M NaOH.

Eluted HU fractions were pooled and dialyzed twice in 1 L of Tris-HCl pH 7.8 (20 mM), NaCl (50 mM), EDTA (1 mM), then then HU isoforms were separated by running dialyzed protein over 5 mL of P11 phosphocellulose resin via gravity. HU isoforms were eluted from resin with the application of a linear gradient of 50 mM to 1M NaCl, 10 mL fractions collected in 5% increasing increments of NaCl. The heterodimer HU eluted at 40-45% of 1M NaCl, or 450-500 mM NaCl.

To ensure HU isoforms were separated, fractions were resolved by 7.5% Triton-X100 acid-urea (TAU) gels, an acidic gel system able to separate the HU isoforms based on pI. The gel recipe for TAU is as follows: 25% glacial acetic acid, 2.5M urea, 1% Triton-X100, 0.0005% (5 μg/mL) riboflavin, 7.5% final concentration of a 40% Bisacrylamide, 150 μl of 10% APS, per 8 ml gel volume and 5μl of 100% TEMED per 8 ml gel volume. Gels were solidified by exposure to UV for 30 min. Samples of fractions from the phosphocellulose column were suspended in TAU sample buffer (5% acetic acid, 50% glycerol (v/v), and 0.6% methyl green) and were resolved by TAU electrophoresis with 5% acetic acid pH 3.6 running buffer at 200V for 45min at room temperature with 5 μl of 10mg/ml cytochrome C as a running marker. The anode and cathode leads were reversed in order to ensure electrophoresis to the negative cathode.

Fractions containing the heterodimer of HU as verified by TAU electrophoresis were pooled and mixed by a slow addition of ammonium sulfate to a final concentration of 2.7M. The ammonium sulfate-protein mixture was then applied to a 5 mL phenyl sepharose FF column on the FPLC and an inverse linear gradient from 2.7M ammonium sulfate to 0M ammonium sulfate was performed to elute heterodimer HU. Fractions with elution peaks corresponding to the FPLC chromatogram were dialyzed in Storage buffer (50 mM HEPES pH 7.5, 50mM NaCl, 10% glycerol) to verify presence of HU by 15% SDS-PAGE. Once fractions were confirmed, they were concentrated and flash frozen in liquid nitrogen in storage buffer and stored at -80°C.

Plasmid derived *Se*HUα_2, *Se*HUβ_2, and lysine variants purification. The homodimers and lysine variants of the homodimers were purified as described above, except as follows: Overnight cultures of *Salmonella* strain JE13152 containing an arabinose inducible chromosomally encoded T7 RNA polymerase and either plasmid pHBSU1, pHUPB5, pHUPB14, pHUPB15, pHUPA3, pHUPA6, or pHUPA7 were grown in LB were sub-cultured 1:100 into 1.5L of TB supplemented with 5μM L(+)-histidine and grown shaking at 37°C to OD600 of 0.7-0.8. T7 RNA polymerase synthesis was induced with the addition of 5mM arabinose and the plasmid encoding the HU protein was induced with 0.5 mM IPTG. Proteins were induced overnight and cells were harvested the next morning by centrifugation at 6,000 x *g* for 15 min. Purification protocol is the same as outlined above under **Native HU heterodimer** and as described by reference [25].

Purification of SeMsrA. MsrA cloned into pTEV18 rTEV-cleavable N-terminal His_6-fusion was electroporated into C43 *E. coli* cells and plated onto LB agar containing ampicillin. The next day, a single colony was inoculated into 10ml of LB containing 100μg/ml of ampicillin and grown overnight shaking at 37ºC. The next morning 1 liter of LB was inoculated with the 10ml culture and grown to OD600 nm of 0.6. The plasmid was induced with 1 mM of IPTG and grown shaking at 37ºC overnight. Cells were centrifuged as described above and the resulting cell pellet was resuspended in 50mM Tris-HCl, pH 8, 500mM NaCl, 20mM imidazole, and 20mM DTT (dithiothreitol), phenylmethanesulfonyl fluoride (PMSF, 1mM), lysozyme (1 mg/ml), and DNase I (1μg/ml). The resuspended cells were sonicated 3x for 60 seconds total, 2 seconds on 2 seconds off at 60% amplitude on a Qsonica sonicator. The lysate was centrifuged at 30,000 x g for 30 minutes in a Beckman Coulter Avanti J-25I centrifuge equipped with a JA-25.50 rotor centrifuge. The resulting supernatant containing soluble His_6-MsrA was purified by gravity column on a 1ml HisPur Ni-NTA resin. The resin was washed with 10ml of 50mM Tris-HCl, pH 8, 500mM NaCl, 20mM imidazole, and 20mM DTT, then 7ml of the previous buffer mixed with 4% (v/v) of 50mM Tris-HCl, pH 8, 500mM NaCl, 500mM imidazole, and 20mM DTT. The His_6-MsrA protein was eluted with 1ml of 50mM Tris-HCl, pH 8, 500mM NaCl, 500mM imidazole, and 20mM DTT. His_6-MsrA protein was dialyzed thrice at 4ºC in a buffer containing 20% glycerol, 100mM NaCl, 50mM Tris-HCl pH 8, and 20mM DTT and was flash frozen at -80ºC.

Purification of SeMsrB. MsrB cloned into pTEV18 with a rTEV-cleavable N-terminal His_6-fusion to make pMsrB1. pMsrB1 was electroporated into C43 *E. coli* cells and plated onto LB agar containing ampicillin. The next day, a single colony was inoculated into 10ml of LB containing 100μg/ml of ampicillin and grown overnight shaking at 37ºC. The next morning 1 liter of LB was inoculated with the 10ml culture and grown to OD600 nm of 0.6. The plasmid was induced with 1 mM of IPTG and grown for 4 more hours. Cells were centrifuged as described above and the resulting cell pellet was resuspended in 50mM Tris-HCl pH 8 and 20mM DTT (dithiothreitol), phenylmethanesulfonyl fluoride (PMSF, 1mM), lysozyme (1 mg/ml), and DNase I (1μg/ml). The resuspended cells were sonicated 3x for 60 seconds total, 2 seconds on 2 seconds off at 60% amplitude on a Qsonica sonicator. The lysate was centrifuged at 30,000 x g for 30 minutes in a Beckman Coulter Avanti J-25I centrifuge equipped with a JA-25.50 rotor centrifuge. The overexpressed MsrB protein was in the insoluble pellet, so the supernatant was removed and the insoluble pellet was resuspended in 50mM Tris-HCl pH 8, 20mM imidazole, and 6M urea. A gravity column NTA purification was performed using a 1ml slurry of HisPur Ni-NTA resin. The insoluble cell lysate was applied to the resin, followed by washing the resin with 10mls of 50mM Tris-HCl pH 8, 20mM imidazole, and 6M urea. A second wash was added with 7ml wash of 50mM Tris-HCl pH 8, 40mM imidazole, and 6M urea. The His_6-MsrB protein was eluted with 1ml of 50mM Tris-HCl pH 8, 500mM imidazole, and 6M urea. Eluted His_6-MsrB protein was slowly refolded by dialysis at 4ºC and step-wise decreased urea concentrations in 50mM Tris-HCl pH 8, starting with 4M urea, then 3M, 1.5M, 0.5M,

and zero molar urea. Each buffer was one liter volume and the protein was dialyzed for 3 hours in each buffer in the cold. The resulting refolded His_6-MsrB protein was filtered through a 0.45un filter and flash frozen at -80ºC in 50mM Tris-HCl pH 8 and 20% glycerol.

TLC of Msr activity. Msr activity assay reactions were set up in 50μl volumes containing 10μg of MsrA and/or MsrB proteins incubated with 10ul of 100mM DTT, 2.5ul of 1M HEPES pH 7, and 5ul of 100mM methionine sulfoxide (Met-SO). Reactions were incubated at 37C for 2 hours, then 10ul of each reaction was spotted in 1ul increments onto a Baker flex silica gel IB2-F TLC plate with half centimeter spacing and 1 cm lanes. Meanwhile, a TLC chamber was equilibrated with N-butanol:acetic acid:water at a 60:12:25 ratio for several hours. The dried TLC plate was run in the chamber until the mobile phase reached close to the top of the plate. The plate was dried and sprayed with ninhydrin spray containing 0.3g of ninhydrin in 100ml of N-butanol and 3ml of acetic acid. The plate was left to dry for less than 10 minutes at 70ºC to develop.

Asp-N digestion and LC-MS/MS analysis. HU proteins were resolved on a 7.5% TAU gel as described above and stained/destained with Coomassie and acetic acid to visualize proteins. Proteins were excised from the gel and in-gel digestion of the HU proteins was conducted according to the Promega Asp-N, sequencing grade In-Gel digestion protocol, cat# V162A.

***In vitro* acetylation assays.** Homogeneous HU proteins (5 μM) were incubated with with [1-^{14}C]-AcCoA (40 μM) in HEPES buffer (50 mM, pH 7) containing *tris*-(2-carboxyethyl-ohosphine (TCEP) hydrochloride (1 mM) with or without *Se*NatA protein (3 μM) for 1 h at 37°C in a total volume of 25 μl. Reactions were stopped by the addition of 5μL of SDS-loading dye (glycerol (60%, v/v), Tris-HCl [0.3 M, pH 6.8], 12 mM ethylenediaminotetracetic acid (EDTA), sodium dodecyl sulfate (SDS, 12% w/v) plus 2-mercaptoethanol (0.87 mM), bromophenol blue (0.05%, w/v) and samples were run by SDS-PAGE at 200 V for 45 min on 20% (w/v) polyacrylamide gels with Tris-HCl pH 8.8 as resolving buffer in the gel and Tris-HCl pH 6.8 as stacking layer buffer in the gel. Gels were stained with Coomassie dye [Coomassie Brilliant Blue R (1g/L), isopropanol (25%, v/v), glacial acetic acid (10%)] and exposed for two days to BAS storage phosphor screems (GE Healthcare). Radiolabel transfer was visualized by imaging on the phosphor imaging setting using a Typhoon trio Plus variable mode imager (GE Healthcare).

***In vitro* deacetylation assays.** Deacetylation assays were conducted by first setting up *in vitro* acetylation assays as described above. Acetylation reactions were then buffer exchanged with an Amicon Ultra 0.5-mL centrifugal filter Ultracel-3K (3 kDa molecular mass cut off) to remove excess AcCoA. Proteins bound to the Amicon filter were washed with 400uL of HEPES buffer (50mM, pH 7). Eluted HU proteins or control Acs protein were brought to a concentration of 3 μM in HEPES buffer (50mM, pH 7). Each eluted HU protein or control Acs protein were added to a reaction mixture at a final concentration of 3 μM in a reaction mixture

containing dithiothreitol (DTT, 2mM), NAD^+ (2mM), $CobB_S$ pure protein (2 μM) and HEPES buffer (50 mM, pH 7). Negative controls include a reaction without $CobB_S$ protein added. Reactions incubated at 37°C for one hour, then were terminated by the addition of SDS loading dye and resolved and visualized as described under *acetylation assays*.

Miller (β-galactosidase) assays. β-galactosidase enzyme activities were determined as described previously[39]. Two biological replicates of each strain were grown overnight in LB plus ampicillin (100ug/ml). In the morning cells were subcultured 1:100 (v/v) into 4ml of LB containing ampicillin, in 10x75mm borosilicate glass tubes with tight fitting lids. L-(+)-arabinose (0.5mM) was added to induce complementation plasmids. Cultures were placed in a 37°C incubator and left static until cells had reached OD600 nm of 0.5-0.6.

Electrophoretic mobility shift assays (EMSAs). 5-6-FAM labeled DNA oligonucleotides were purchased from Integrated DNA Technologies (IDT, Coralville, IA). dsDNA probes were generated by PCR (polymerase chain reaction) using PFU Ultra II polymerase (Agilent) as according to manufacturer protocol. DNA probes were cleaned of excess labeled nucleotides using a Promega PCR and Gel Clean up Kit. 50ng of *hilA* promoter DNA was used per reaction in a total reaction volume of 25 μl. EMSA buffer contained a final concentation of 10mM Tris-HCl (pH 7.6), 15mM NaCl, 15% (v/v) glycerol, and 0.1mM EDTA (ethylenediaminetetraacetic acid). Increasing concentrations of *Se*HUα2 protein

were added in excess to DNA probe as labeled in figure X. Reactions were incubated at room temperature for 10 minutes, then 5 μl of 50% glycerol was added to each reaction and 20 μl of reactions were resolved by native polyacrylamide gel electrophoresis on non-denaturing Criterion *tris*(hydroxymethyl)aminomethane (Tris)-HCl buffer (375 mM, pH 8.6) 7.5% (w/v) polyacrylamide gel (BioRad) gels in 0.5X TBE (tris-borate-EDTA) buffer at 120 volts. An empty lane was loaded with 2 μl of 10X DNA loading dye (containing 50% glycerol, bromophenol blue, and xylene cyanol) to use as a visual tracking aid. Gels were electrophoresed at 4°C until the bromophenol blue indicator reached the bottom of the gels. Gels were imaged using a Typhoon Trio^{+} variable imager (GE Healthcare) at the wavelength 488 nm, and quantified with ImageQuant v5.2 software.

6.6 ACKNOWLEDGEMENTS

We would like to thank Chau-wen Chou at the Proteomics and Mass Spectrometry (PAMS) core facility at the University of Georgia for her LC/MS-MS analysis. We would also like to thank Stephane Benoit for all of his insight and assistance with Msr proteins.

6.7 REFERENCES

1. Hwang, C.S., A. Shemorry, and A. Varshavsky, *N-terminal acetylation of cellular proteins creates specific degradation signals.* Science, 2010. **327**(5968): p. 973-7.

2. Shemorry, A., C.S. Hwang, and A. Varshavsky, *Control of protein quality and stoichiometries by N-terminal acetylation and the N-end rule pathway.* Mol Cell, 2013. **50**(4): p. 540-51.

3. Xu, F., et al., *Two N-terminal acetyltransferases antagonistically regulate the stability of a nod-like receptor in Arabidopsis.* Plant Cell, 2015. **27**(5): p. 1547-62.

4. Sheikh, T.I., et al., *MeCP2_E1 N-terminal modifications affect its degradation rate and are disrupted by the Ala2Val Rett mutation.* Hum Mol Genet, 2017. **26**(21): p. 4132-4141.

5. Behnia, R., et al., *The yeast orthologue of GRASP65 forms a complex with a coiled-coil protein that contributes to ER to Golgi traffic.* J Cell Biol, 2007. **176**(3): p. 255-61.

6. Behnia, R., et al., *Targeting of the Arf-like GTPase Arl3p to the Golgi requires N-terminal acetylation and the membrane protein Sys1p.* Nat Cell Biol, 2004. **6**(5): p. 405-13.

7. Forte, G.M., M.R. Pool, and C.J. Stirling, *N-terminal acetylation inhibits protein targeting to the endoplasmic reticulum.* PLoS Biol, 2011. **9**(5): p. e1001073.

8. Setty, S.R., et al., *Golgi targeting of ARF-like GTPase Arl3p requires its Nalpha-acetylation and the integral membrane protein Sys1p.* Nat Cell Biol, 2004. **6**(5): p. 414-9.

9. Trexler, A.J. and E. Rhoades, *N-Terminal acetylation is critical for forming alpha-helical oligomer of alpha-synuclein.* Protein Sci, 2012. **21**(5): p. 601-5.

10. Bartels, T., J.G. Choi, and D.J. Selkoe, *alpha-Synuclein occurs physiologically as a helically folded tetramer that resists aggregation.* Nature, 2011. **477**(7362): p. 107-10.

11. Holmes, W.M., et al., *Loss of amino-terminal acetylation suppresses a prion phenotype by modulating global protein folding.* Nat Commun, 2014. **5**: p. 4383.

12. Singer, J.M. and J.M. Shaw, *Mdm20 protein functions with Nat3 protein to acetylate Tpm1 protein and regulate tropomyosin-actin interactions in budding yeast.* Proc Natl Acad Sci U S A, 2003. **100**(13): p. 7644-9.

13. Yang, D., et al., *Nalpha-acetylated Sir3 stabilizes the conformation of a nucleosome-binding loop in the BAH domain.* Nat Struct Mol Biol, 2013. **20**(9): p. 1116-8.

14. Monda, J.K., et al., *Structural conservation of distinctive N-terminal acetylation-dependent interactions across a family of mammalian NEDD8 ligation enzymes.* Structure, 2013. **21**(1): p. 42-53.

15. Scott, D.C., et al., *N-terminal acetylation acts as an avidity enhancer within an interconnected multiprotein complex.* Science, 2011. **334**(6056): p. 674-8.

16. Coulton, A.T., et al., *The recruitment of acetylated and unacetylated tropomyosin to distinct actin polymers permits the discrete regulation of specific myosins in fission yeast.* J Cell Sci, 2010. **123**(Pt 19): p. 3235-43.

17. Arnaudo, N., et al., *The N-terminal acetylation of Sir3 stabilizes its binding to the nucleosome core particle.* Nat Struct Mol Biol, 2013. **20**(9): p. 1119-21.

18. Kentache, T., et al., *Proteomic characterization of Nalpha- and Nepsilon-acetylation in Acinetobacter baumannii.* J Proteomics, 2016. **144**: p. 148-58.

19. Ouidir, T., et al., *Characterization of N-terminal protein modifications in Pseudomonas aeruginosa PA14.* J Proteomics, 2015. **114**: p. 214-25.

20. Parks, A.R. and J.C. Escalante-Semerena, *Modulation of the bacterial CobB sirtuin deacylase activity by N-terminal acetylation.* Proc Natl Acad Sci U S A, 2020. **117**(27): p. 15895-15901.

21. Oberto, J., et al., *The HU regulon is composed of genes responding to anaerobiosis, acid stress, high osmolarity and SOS induction.* PLoS One, 2009. **4**(2): p. e4367.

22. Claret, L. and J. Rouviere-Yaniv, *Variation in HU composition during growth of Escherichia coli: the heterodimer is required for long term survival.* J Mol Biol, 1997. **273**(1): p. 93-104.

23. Bonnefoy, E., A. Almeida, and J. Rouviere-Yaniv, *Lon-dependent regulation of the DNA binding protein HU in Escherichia coli.* Proc Natl Acad Sci U S A, 1989. **86**(20): p. 7691-5.

24. Carabetta, V.J., et al., *YfmK is an N(epsilon)-lysine acetyltransferase that directly acetylates the histone-like protein HBsu in Bacillus subtilis.* Proc Natl Acad Sci U S A, 2019. **116**(9): p. 3752-3757.

25. Pellegrini, O., et al., *Overproduction and improved strategies to purify the threenative forms of nuclease-free HU protein.* Biochimie, 2000. **82**(8): p. 693-704.

26. Ingrosso, D., et al., *Specificity of endoproteinase Asp-N (Pseudomonas fragi): cleavage at glutamyl residues in two proteins.* Biochem Biophys Res Commun, 1989. **162**(3): p. 1528-34.

27. Tetaz, T., et al., *Relaxed specificity of endoproteinase Asp-N: this enzyme cleaves at peptide bonds N-terminal to glutamate as well as aspartate and cysteic acid residues.* Biochem Int, 1990. **22**(3): p. 561-6.

28. Solbiati, J., et al., *Processing of the N termini of nascent polypeptide chains requires deformylation prior to methionine removal.* J Mol Biol, 1999. **290**(3): p. 607-14.

29. Favrot, L., J.S. Blanchard, and O. Vergnolle, *Bacterial GCN5-Related N-Acetyltransferases: From Resistance to Regulation.* Biochemistry, 2016. **55**(7): p. 989-1002.

30. VanDrisse, C.M. and J.C. Escalante-Semerena, *Protein Acetylation in Bacteria.* Annu Rev Microbiol, 2019. **73**: p. 111-132.

31. Hentchel, K.L. and J.C. Escalante-Semerena, *Acylation of Biomolecules in Prokaryotes: a Widespread Strategy for the Control of Biological Function and Metabolic Stress.* Microbiol Mol Biol Rev, 2015. **79**(3): p. 321-46.

32. Tucker, A.C. and J.C. Escalante-Semerena, *Biologically active isoforms of CobB sirtuin deacetylase in Salmonella enterica and Erwinia amylovora.* J Bacteriol, 2010. **192**(23): p. 6200-8.

33. Schechter, L.M., et al., *The small nucleoid-binding proteins H-NS, HU, and Fis affect hilA expression in Salmonella enterica serovar Typhimurium.* Infect Immun, 2003. **71**(9): p. 5432-5.

34. Ellermeier, J.R. and J.M. Slauch, *Fur regulates expression of the Salmonella pathogenicity island 1 type III secretion system through HilD.* J Bacteriol, 2008. **190**(2): p. 476-86.

35. Bonissone, S., et al., *N-terminal protein processing: a comparative proteogenomic analysis.* Mol Cell Proteomics, 2013. **12**(1): p. 14-28.

36. Walker, K.W. and R.A. Bradshaw, *Yeast methionine aminopeptidase I. Alteration of substrate specificity by site-directed mutagenesis.* J Biol Chem, 1999. **274**(19): p. 13403-9.

37. Datsenko, K.A. and B.L. Wanner, *One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products.* Proc. Natl. Acad. Sci. USA, 2000. **97**: p. 6640-6645.

38. Davis, R.W., D. Botstein, and J.R. Roth, *A manual for genetic engineering: advanced bacterial genetics.* 1980, Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

39. Miller, F.D. and C.L. Hershberger, *A quantitative beta-galactosidase alpha-complementation assay for fusion proteins containing human insulin B-chain peptides.* Gene, 1984. **29**(1-2): p. 247-50.

40. Tabor, S. and C.C. Richardson, *A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes.* Proc Natl Acad Sci U S A, 1985. **82**(4): p. 1074-8.

Table 6.1 Strains used in this study.		
Strain	**Relevant genotype**	**Reference/source[a]**
***S. enterica enterica* serovar Typhimurium LT2 strains**		
JE6583	*metE205 ara-9*	Laboratory strain
Derivatives of JE6583		
JE9637 **(**SB300A#1)	SB300A#1 *S. enterica* aadA::araCPBADT7-1	J. Bact (2002) Vol. 184, p6056-6059
JE13152	aadA::araCPBADT7-1 ΔhupA102 ΔhupB104 lon-71 zaj-1034::tn10 yfiQ::mudJ	
***S. enterica* serovar Typhimurium ATCC 14028s strains**		ATCC
JE7658		
Derivatives of JE7658		
JE18227 (JS575)	*att*λ::pDX1::*hilA*'-*lac*$^+$	Ellermeier & Slauch 2008
JE25581	*att*λ::pDX1::*hilA*'-*lac*$^+$ *hupA101*::*kan*$^+$ / pCV1	This work
JE25582	attλ::pDX1::*hilA*'-*lac*$^+$ hupA*101*::*kan*$^+$ / pNatA10	This work
JE25583	attλ::pDX1::*hilA*'-*lac*$^+$ *hupB103*::cat$^+$ / pCV1	This work
JE25584	attλ::pDX1::*hilA*'-*lac*$^+$ *hupB103*::cat$^+$ / pNatA10	This work
JE25588	attλ::pDX1::*hilA*'-*lac*$^+$ *hupB103*::cat$^+$ hupA*101*::*kan*$^+$ / pCV1	This work
JE25589	attλ::pDX1::*hilA*'-*lac*$^+$ *hupB103*::cat$^+$ hupA*101*::*kan*$^+$ / pNatA10	This work
***E. coli* strains**		
E. coli C41 (IDE3)	*pka12*::*kan*$^+$ *ompT hsdS* (r_BmB) *gal* l (DE3)	Laboratory collection
E. coli C43 (IDE3)	F-ompT gal hsdSB (rB-mB-) dcm lon C41 (IDE3)	Laboratory collection

Table 6.2 Plasmids used in this study			
Plasmid	**Genotype**	**Description**	**Source**
pCV1	*araC*$^{+}$ *bla*$^{+}$	P_{araBAD} expression vector	VanDrisse 2016
pNatA10	*natA*$^{+}$ *bla*$^{+}$	seNatA cloned into pCV1	Parks 2020
pTEV6	*bla*$^{+}$	N-terminal rTEV-cleavable MBP-His fusion overexpression vector	Rocco 2008
pYfmk1	*yfmK*$^{+}$ *bla*$^{+}$	*yfmK* cloned into pTEV6	This work
pNatA6	*natA*$^{+}$ *bla*$^{+}$	*natA* cloned into pTEV6	Parks 2020
pTEV18	*bla*$^{+}$	*N*-termnal rTEV-cleavable MBP-His fusion overexpression vector	VanDrisse 2016
pMsrA1	*msrA*$^{+}$ *bla*$^{+}$	*Stm4408 (msrA)* cloned into pTEV18	This work
pMsrB1	*msrB*$^{+}$ *bla*$^{+}$	*Stm*1291 (*yeaA* also called *msrB*) cloned into pTEV18	This work
pT7-7	*bla*$^{+}$	Overexpression vector with T7 promoter	Tabor&Richardson 1985 [40]
pSAPKO-WT	*Kan*$^{+}$	Derivative of pET28b with BspQI cut sites and T7 promoter for overexpression of tagless vectors	Gift from Momany Lab
pHBSU1	*hbs*$^{+}$ *kan*$^{+}$	*hbs* cloned into pSAPKO-WT	This work
pHUPB5	*hupB*$^{+}$ *bla*$^{+}$	*hupB* cloned into pT7-7 (Contains extra start ATG)	This work
pHUPB14	*hupB*$^{+}$ *bla*$^{+}$	Encodes $HupB^{K3R,K9R,K18R,K37R,K53R,K67R,K75R,K83R,K86R}$ in pT7-7	This work
pHUPB15	*hupB*$^{+}$ *bla*$^{+}$	Encodes *hupB* with deleted extra ATG in pT7-7. Mutagenized pHUPB5.	This work
pHUPA3	*hupA*$^{+}$ *bla*$^{+}$	Encodes *hupA* in pT7-7	This work
pHUPA6	*hupA*$^{+}$ *bla*$^{+}$	Encodes $HupA^{K3R,K13R,K18R,K22R,K37R,K51R,K67R,K83R,K86R,K90R}$ in pT7-7	This work
pHUPA7	*hupA*$^{+}$ *bla*$^{+}$	Encodes *hupB* with deleted extra ATG in pT7-7. Mutagenized pHUPA3.	This work

Table 6.3. Primers used in this study.	
Primer Name	**Primer Sequence 5' → 3'**
5' pYiaC10 BspQI	NNGCTCTTCNTTCATGATTCGCAAATCCCAGAGTGAAGAC
3' pYiaC10 BspQI	NNGCTCTTCNTTATTACGGCGTTTGATCCGCCTGCCAAC
5' msrB_F_BSPQ	NNGCTCTTCNTTCATGAGCACGTTTAAAGTGAG
msrB_R_BSPQ	NNGCTCTTCNTTATCAGCCTTTCAGTTGAT
Yfmk_F_BspQI	NNGCTCTTCNTTCatgGCTTCAATAGACAGG
YfmK_R_BspQI	NNGCTCTTCNTTATCAGTTGCGAAGAATCAG
Hbsu_F_sapkowT	NNGCTCTTCNATGatgAACAAAACAGAACT
Hbsu_R_sapkowT	NNGCTCTTCNTTAAGTTGCCGGAAAAtaa
del_M1_HupA1	ctttaagaaggagatatacatatgaacaagactcaactgattgatgta
del_M1_hupA2	tacatcaatcagttgagtcttgttcatatgtatatctccttcttaaag
DelM1_HupB1	tttaactttaagaaggagatatacatatgaataaatctcaactgatcgaaaaa attgc
DelM1_HupB2	gcaatttttcgatcagttgagatttattcatatgtatatctccttcttaaagttaaa
SeHupB_K3R_P1T77	tttcgatcagttgagatctattcaccatatgtatatctccttcttaaagt
SeHupB_K3R_P2T77	actttaagaaggagatatacatatggtgaatagatctcaactgatcgaaa
SeHupB_K9R_P1	cagcccctgcagcaattctttcgatcagttgagat
SeHupB_K9R_P2	atctcaactgatcgaaagaattgctgcaggggctg
QC1_K18R_HupB	caggggctgatatctctagggctgcggctg
QC2_K18R_HupB	cagccgcagccctagagatatcagcccctgc
QC1_K37R_HupB	tcatccccttctctcagagattcggtaacagaagca
QC2_K37R_HupB	tgcttctgttaccgaatctctgagagaaggggatga
QC1_K75R_HupB	ctcggcactctggcagcggcgatggt
QC2_K75R_HupB	accatcgccgctgccagagtgccgag
QC1_ K83R_86R_HupB	taccgcgtctctcagcgctctacctgcacggaaactcg-
QC2_ K83R_86R_HupB	cgagtttccgtgcaggtagagcgctgagagacgcggta
hilA_R1_FAM	taaaatgtggcatgataatagt
hilA_F1	ctattgcaatgaggcca
yfmk_F_KpnI-pTEV6	NNNGGTACCatgGCTTCAATAGACAGG
yfmk_R_NotI-pTEV6	NNNGCGGCCGCtcaGTTGCGAAGAATCAG

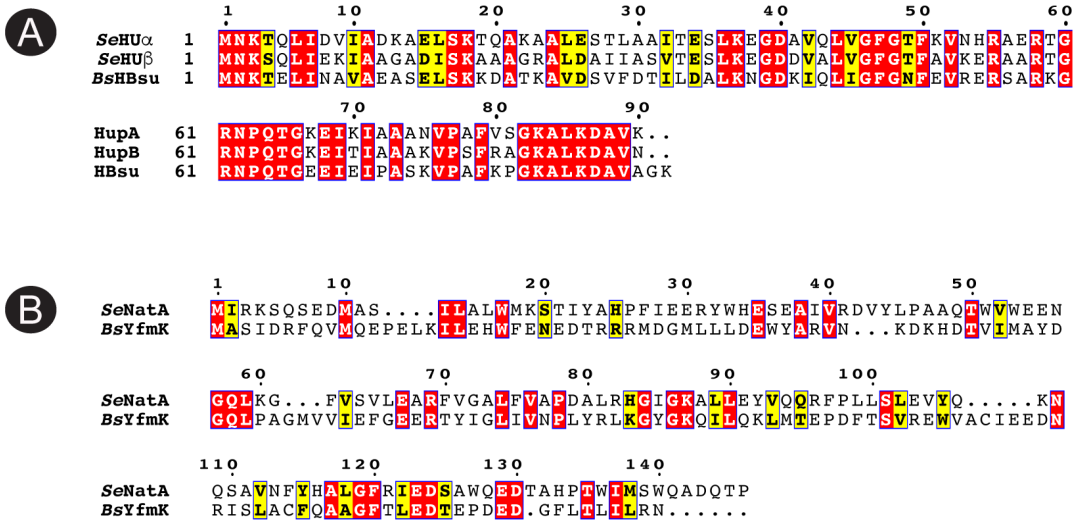

Figure 6.1. A. *Se*NatA and *Bs*YfmK similarity and identity level. B. *Se*HUα, *Se*HUβ, and HBsu amino acid sequence alignment. Similar residues are written in black bold characters and boxed in yellow. Identical residues are boxed in red. Figure generated in ESPript 3.0.

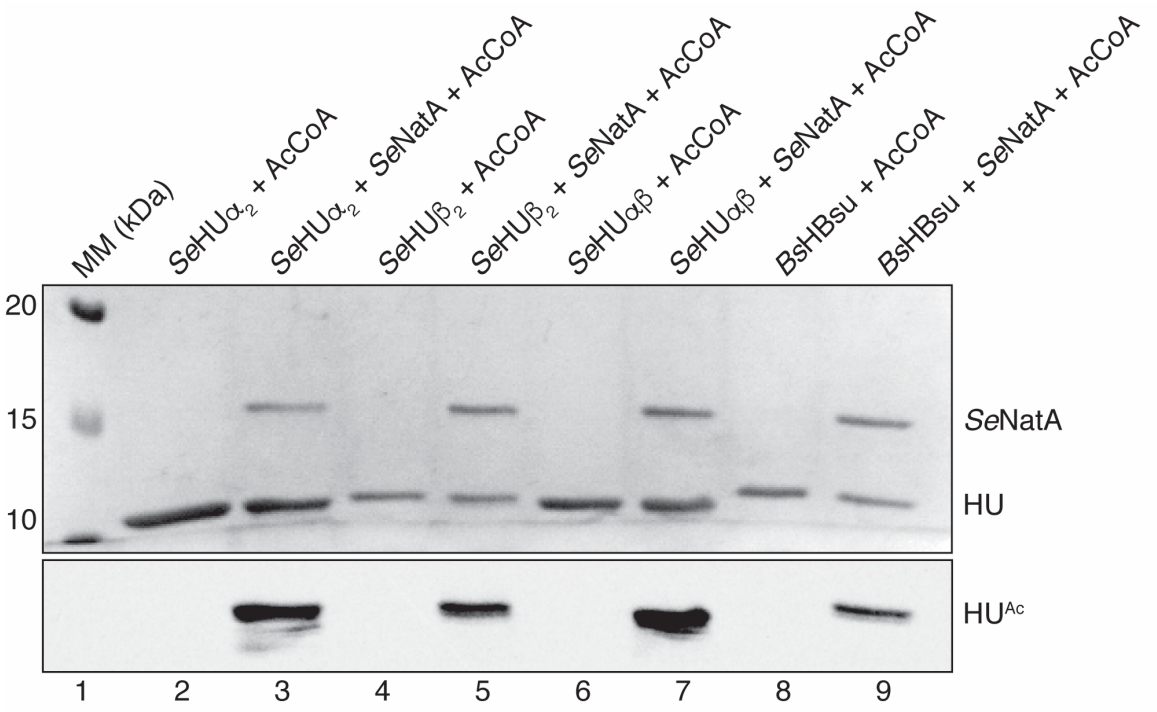

Figure 6.2. *Se*NatA acetylates *Se*HUβ2, *Se*HUα2, *Se*HUαβ2, and *Bs*HBsu *in vitro*. The ability of *Se*NatA to acetylate HU proteins was assessed via incubation of 40 μM [acetyl-1-^{14}C]-AcCoA with *Se*NatA (2 μM) and either *Se*HUβ2, *Se*HUα2, *Se*HUαβ, and *Bs*HBsu proteins (5 μM each) for one hour at 37°C. Controls included reaction mixtures containing each of the HU proteins, [acetyl-1-^{14}C]-AcCoA, but no acetyltransferase. Proteins were resolved by SDS-PAGE and visualized by Coomassie Blue R staining (top image) using Precision Plus protein standard (Bio-Rad Laboratories) as the molecular mass (MM) marker. The distribution of radiolabel signal was visualized by phosphor imaging (bottom image).

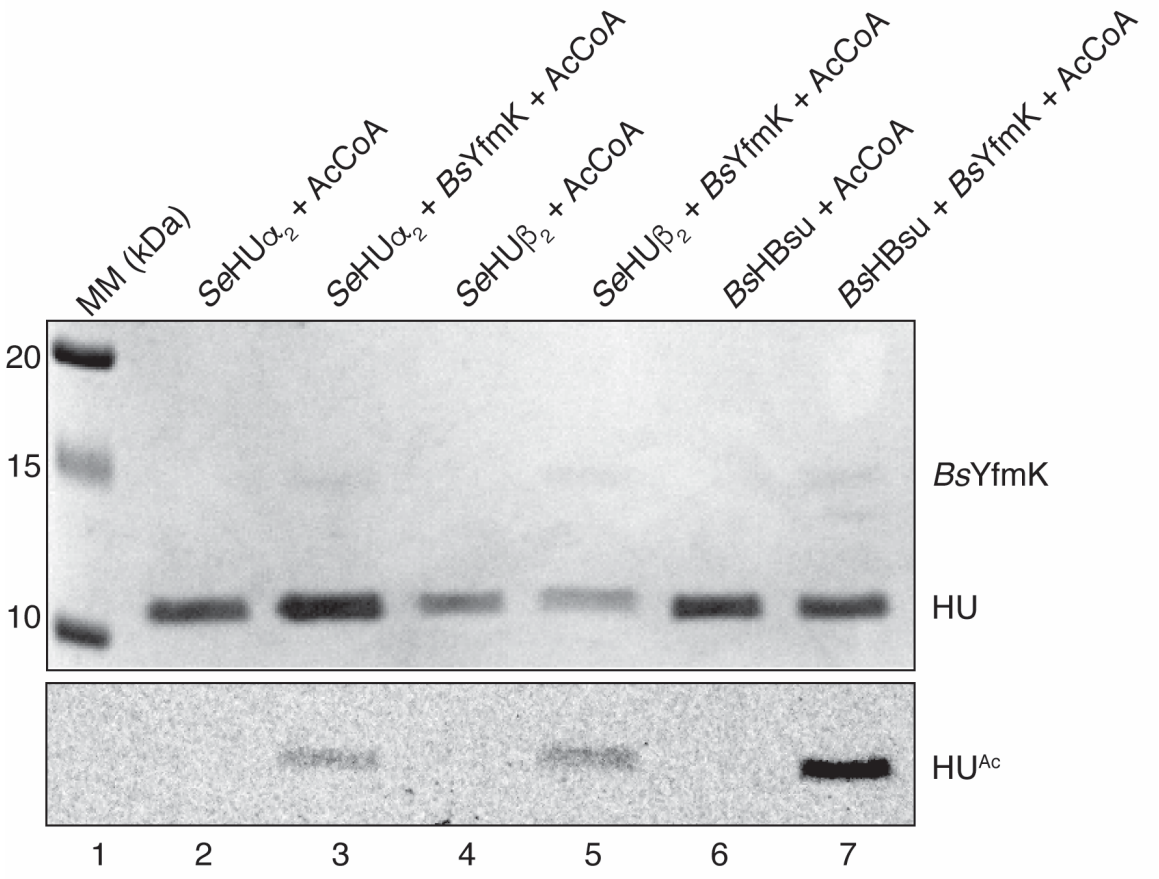

Figure 6.3. *Bs*YfmK can acetylate *Se*HU and *Bs*HBsu proteins. The ability of *Bs*YfmK to acetylate HU proteins was assessed via incubation of 40 μM [acetyl-1-^{14}C]-AcCoA with *Bs*YfmK (2 μM) and either *Se*HUβ$_2$, *Se*HUα$_2$, or *Bs*HBsu proteins (5 μM each) for one hour at 37°C. Controls included reaction mixtures containing each of the HU proteins, [acetyl-1-^{14}C]-AcCoA, but no acetyltransferase. Proteins were resolved by SDS-PAGE and visualized by Coomassie Blue R staining (top image) using Precision Plus protein standard (Bio-Rad Laboratories) as the molecular mass (MM) marker. The distribution of radiolabel signal was visualized by phosphor imaging (bottom image).

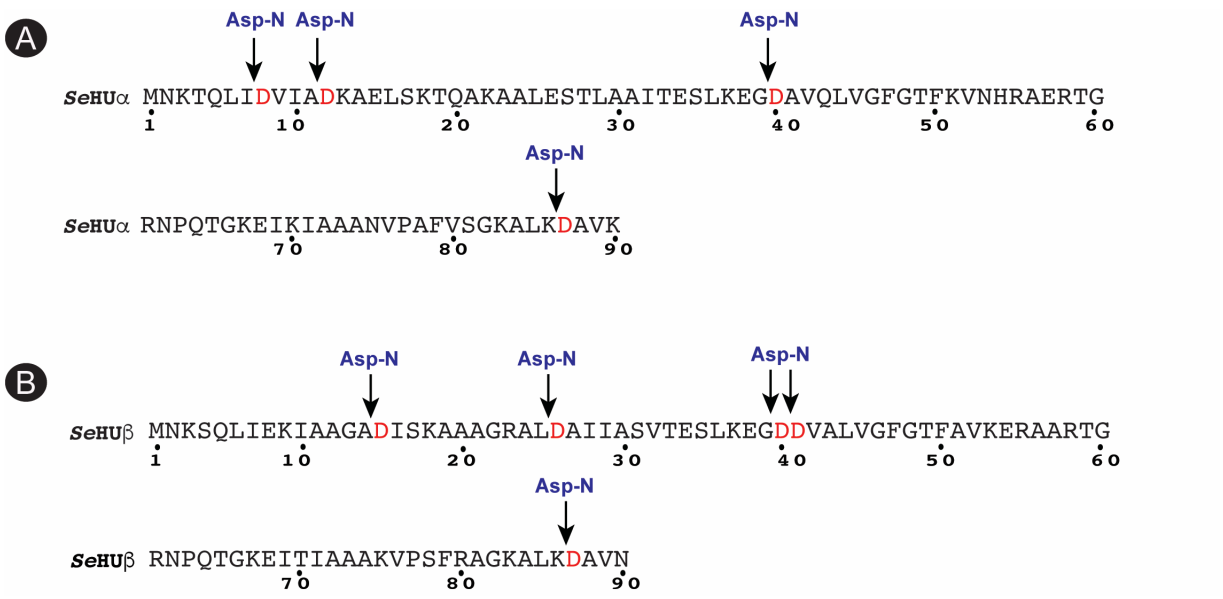

Figure 6.4. Asp-N endoproteinase digestion mapping of HU proteins. The predicted digestion sites of A: SeHUα or panel B: SeHUβ by Asp-N protease, which cleaves N-terminal to aspartic acid residues, as labeled in red.

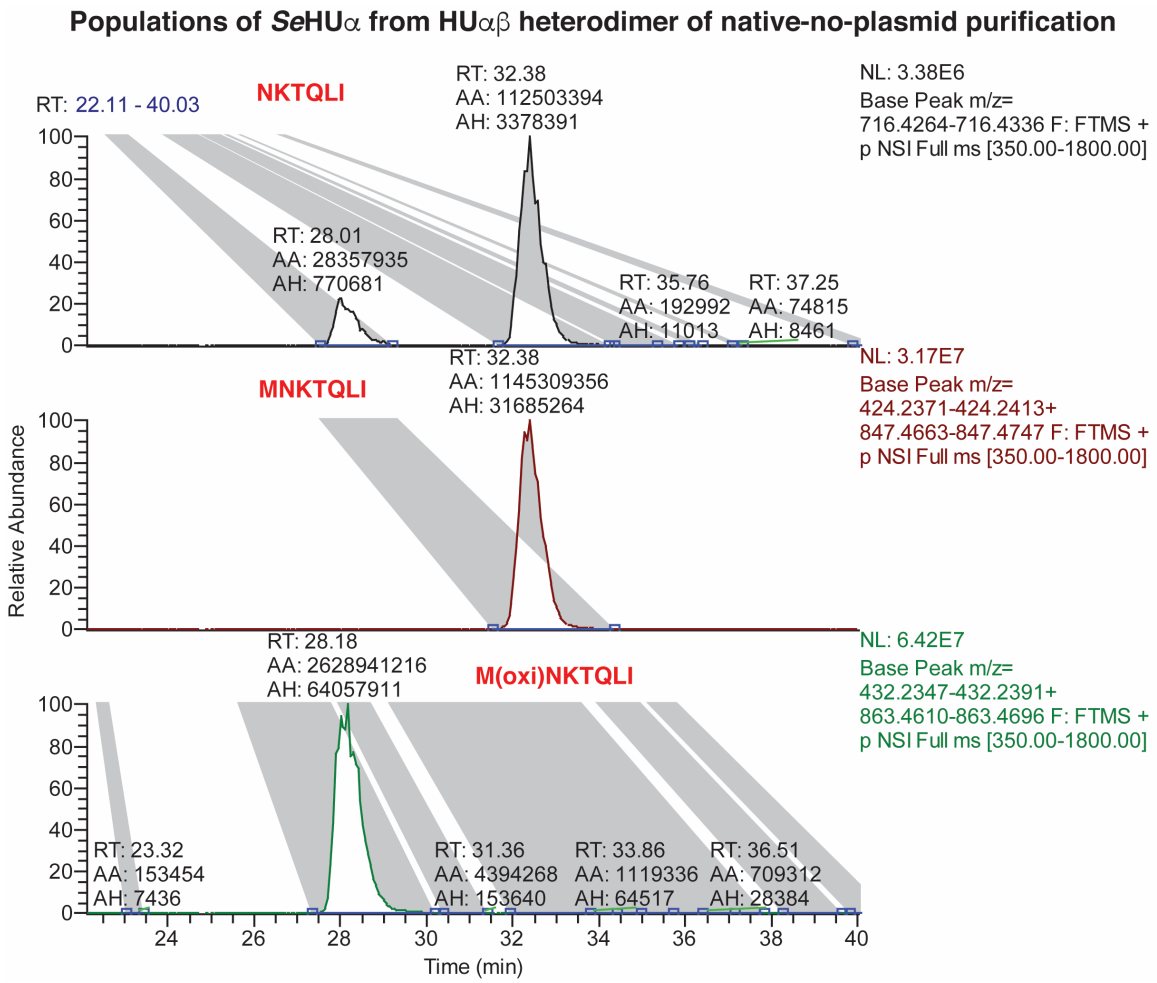

Figure 6.5. LC-MS/MS peaks of the N-terminus of SeHUα separated from HU heterodimer. Native-no-plasmid purification of HU heterodimer from *S. enterica* was resolved on a 7.5% TAU gel to separate the a and b subunits. The HUα protein was excised from the gel, digested with Asp-N protease and analyzed by LC-MS/MS.

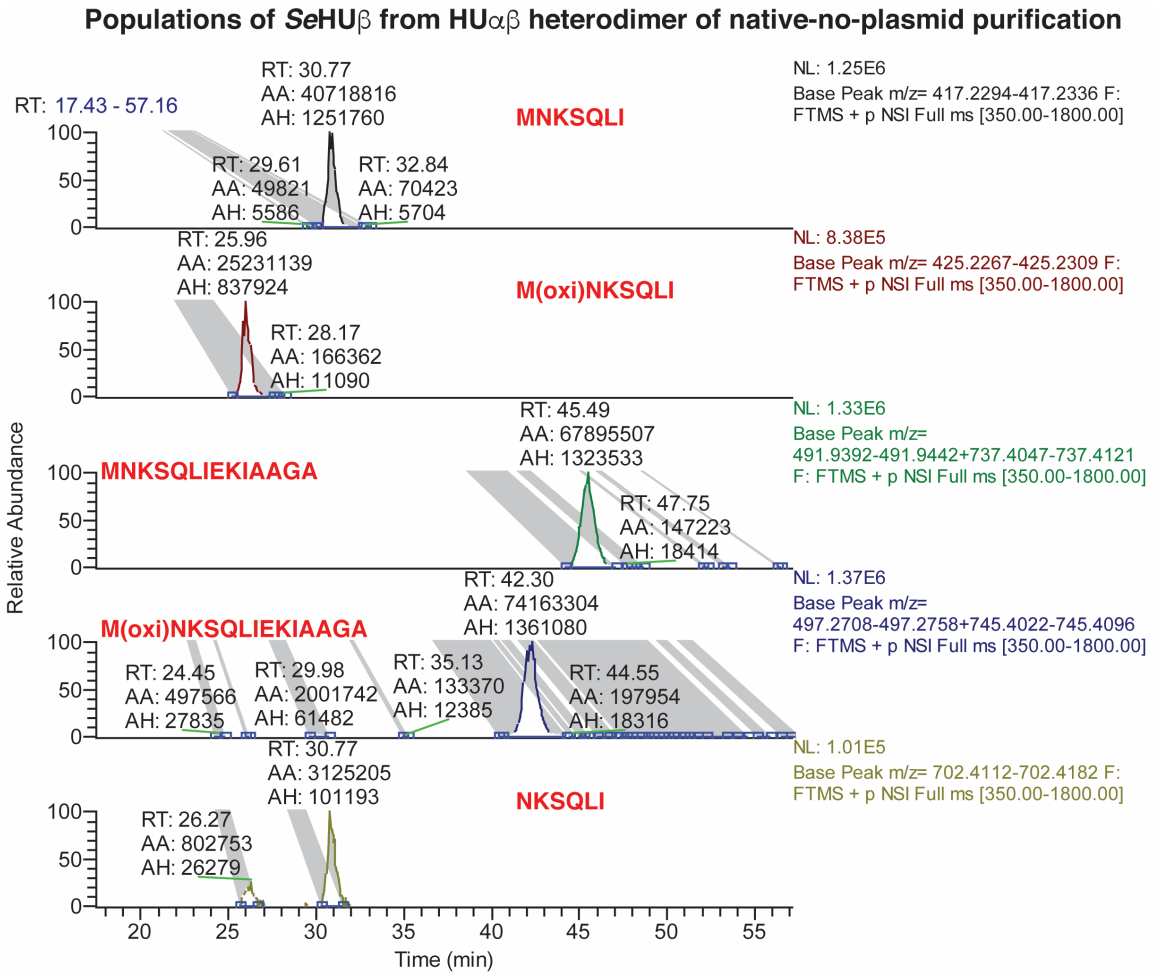

Figure 6.6. LC-MS/MS peaks of the N-terminus of SeHUβ separated from HU heterodimer. Native-no-plasmid purification of HU heterodimer from *S. enterica* was resolved on a 7.5% TAU gel to separate the a and b subunits. The HUβ protein was excised from the gel, digested with Asp-N protease and analyzed by LC-MS/MS.

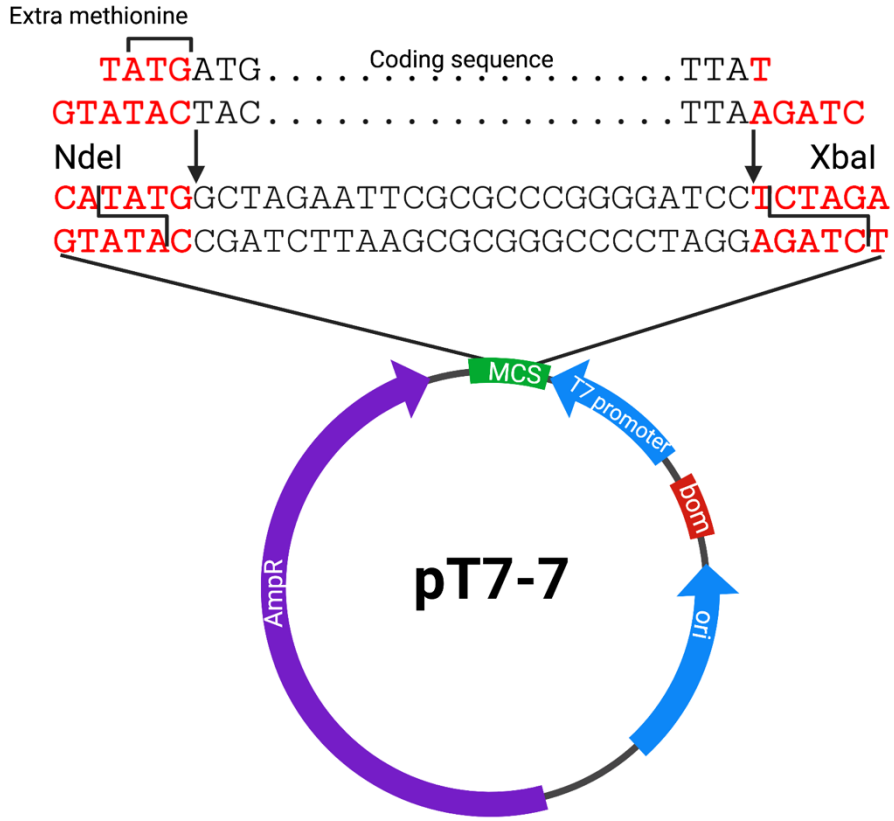

Figure 6.7. Cloning of HU proteins into pT7-7 that resulted in extra N-terminal methionine. Traditional cloning into pT7-7 protein overexpression and purification vector using restriction endonucleases and cut sites of NdeI for the 5' excision and overhang and XbaI for the 3' excision and overhang. Using NdeI resulted in an additional ATG before the coding ATG for the HU genes.

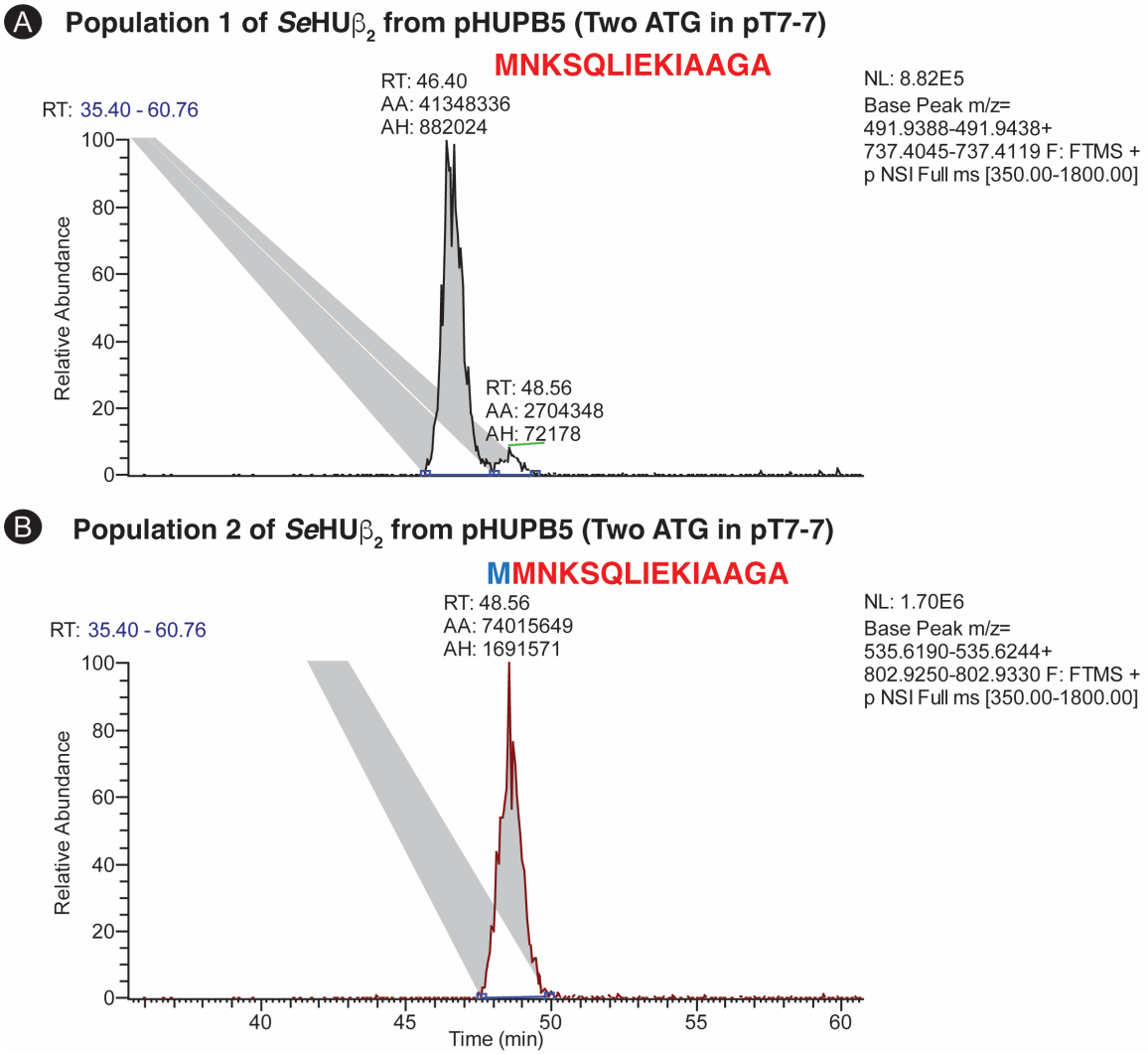

Figure 6.8. LC-MS/MS peaks of the N-terminus of SeHUβ2 purified from pHUPB5. Plasmid pHUPB5 contains the coding sequence for the *S. enterica* gene *hupB*, that was cloned into plasmid pT7-7 using restriction enzymes 5'NdeI and 3' XbaI.

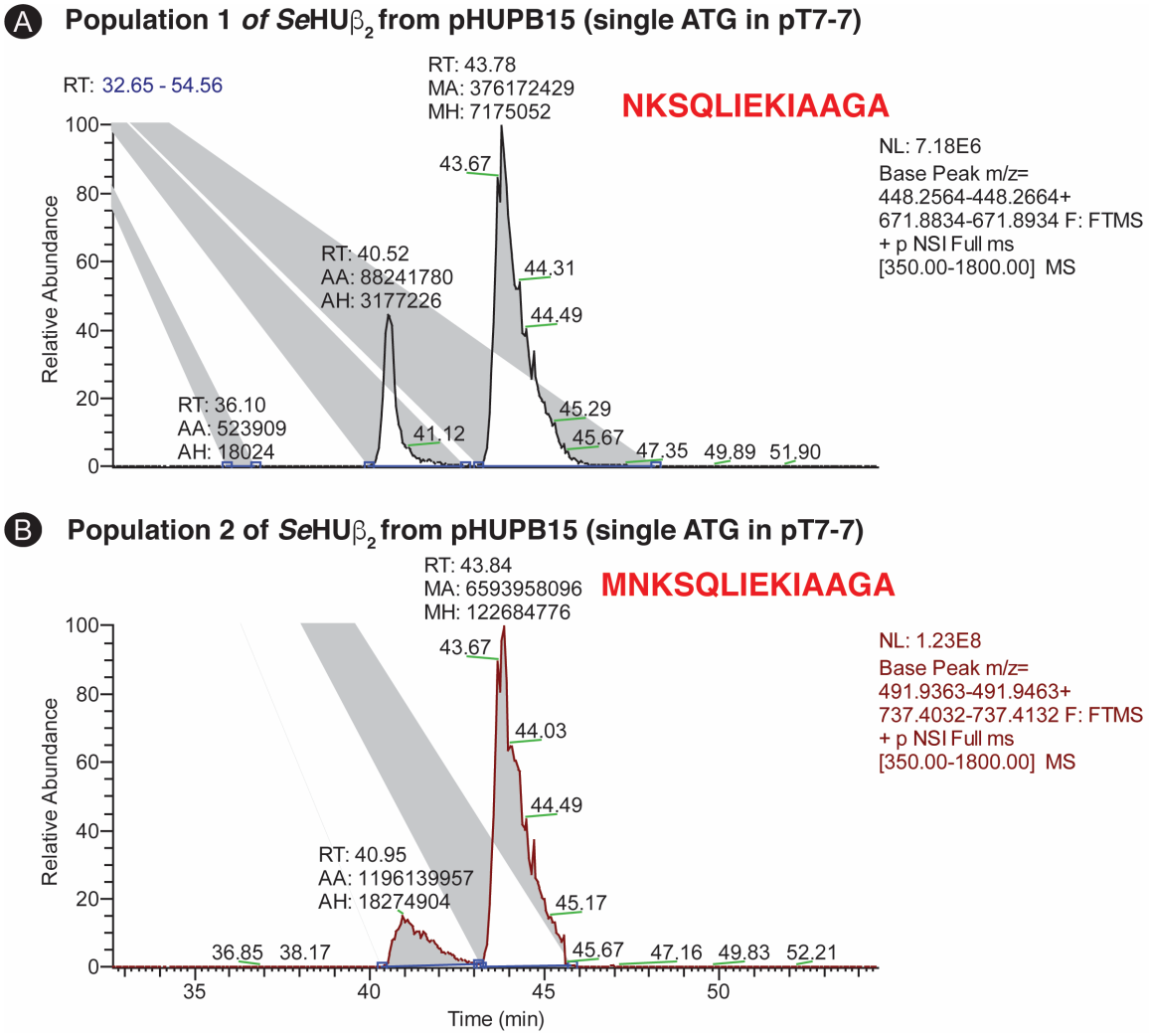

Figure 6.9. LC-MS/MS of the N-terminus of SeHUβ2 purified from pHUPB15. Plasmid pHUPB5 contains the coding sequence for the *S. enterica* gene *hupB*, that mutated by quick-change primer mutagenesis to remove the extra ATG at the beginning of the coding sequence, generating pHUPB15.

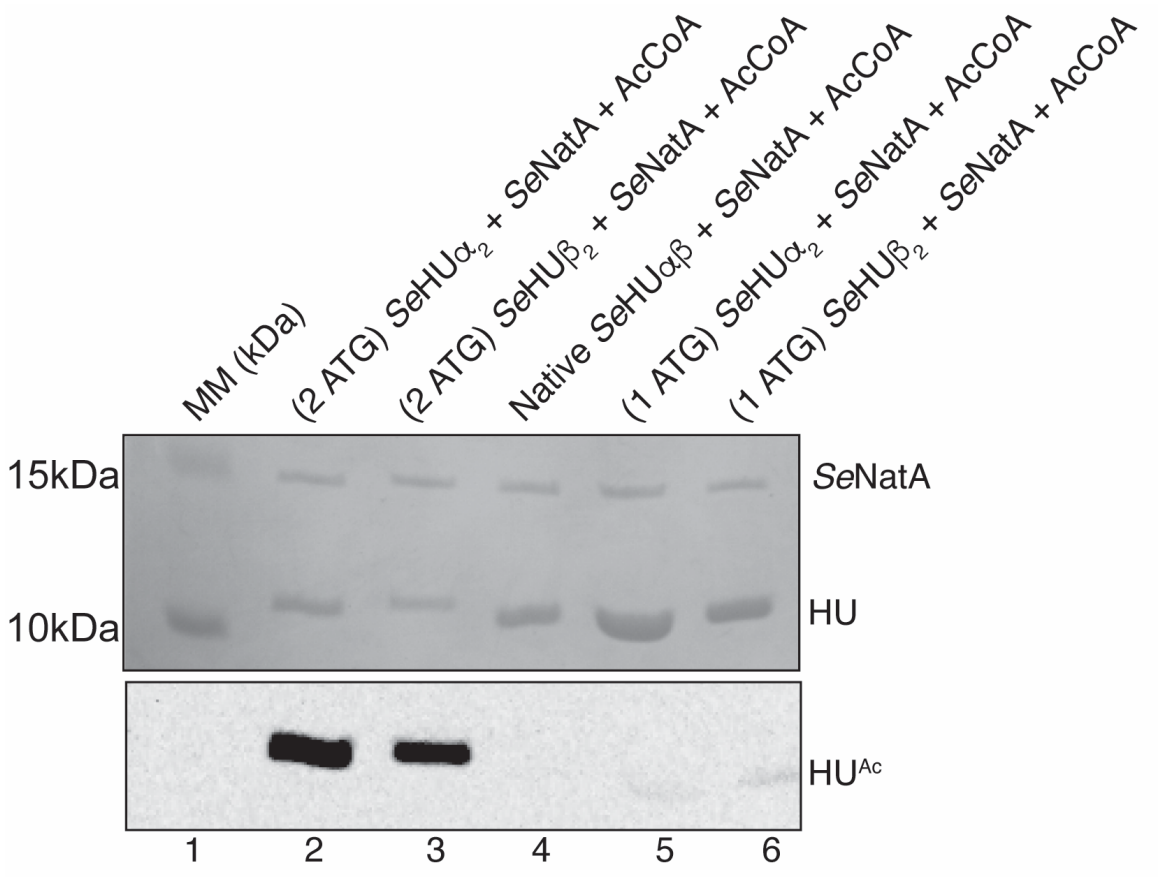

Figure 6.10. Differences in acetylation of HU proteins containing extra N-terminal Met verse a single Met. The ability of *Se*NatA to acetylate HU proteins with an extra N-terminal methionine compared to those purified with a single N-terminal methionine was assessed via incubation of 40 μM [acetyl-1-^{14}C]-AcCoA with *Se*NatA (2 μM) and either *Se*HUβ$_2$, *Se*HUα$_2$, or *Se*HUαβ proteins (5 μM each) for one hour at 37°C. Proteins were resolved by SDS-PAGE and visualized by Coomassie Blue R staining (top image) using Precision Plus protein standard (Bio-Rad Laboratories) as the molecular mass (MM) marker. The distribution of radiolabel signal was visualized by phosphor imaging (bottom image).

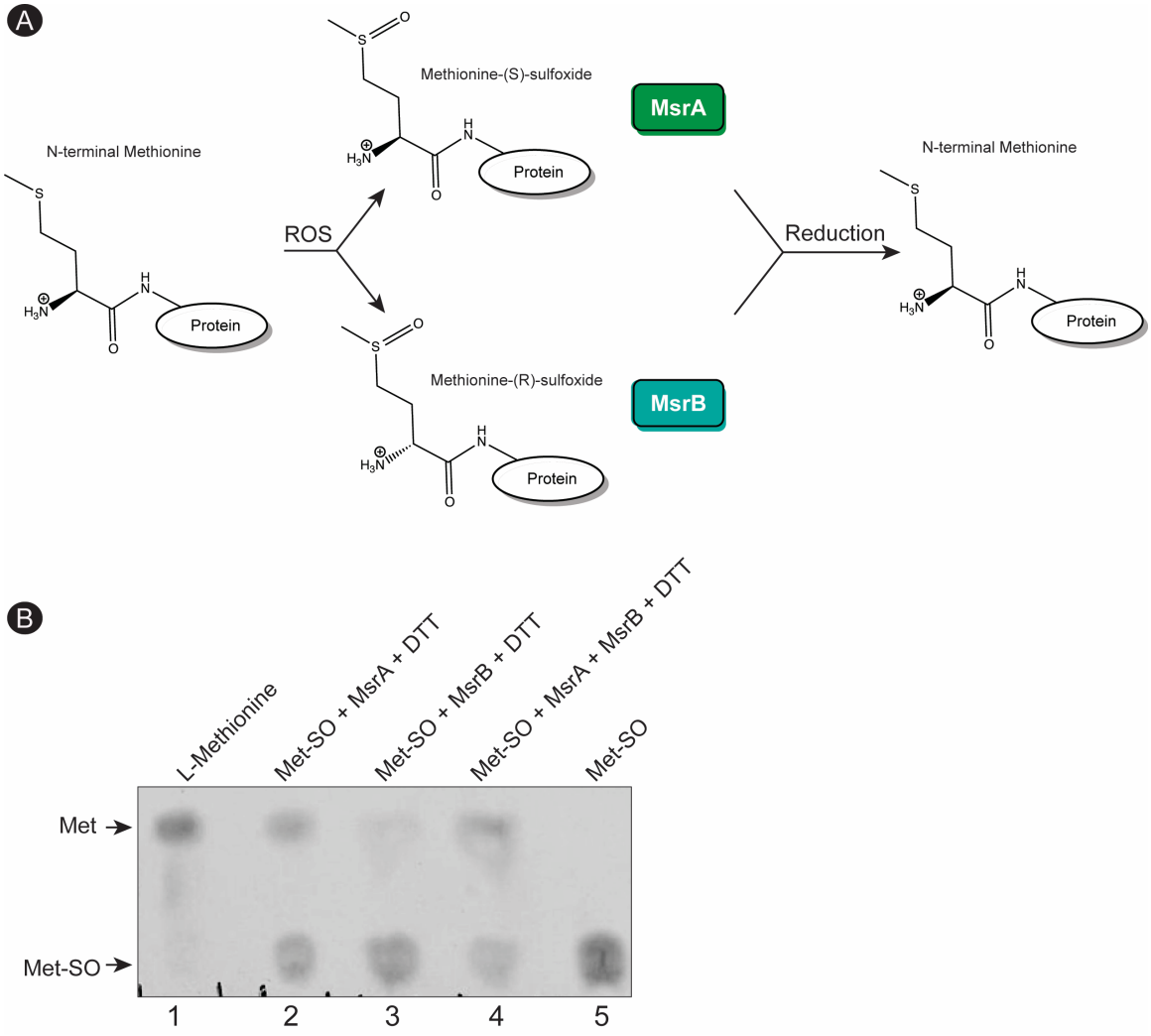

Figure 6.11. Methionine sulfoxide reductases repair stereospecific oxidized methionines of proteins. A. Schematic of how Msr enzymes repair methionine sulfoxide, but repair by Msr is diastereoselective. B. TLC verification of enzymatic activity of *Se*Msr proteins purified to homogeneity.

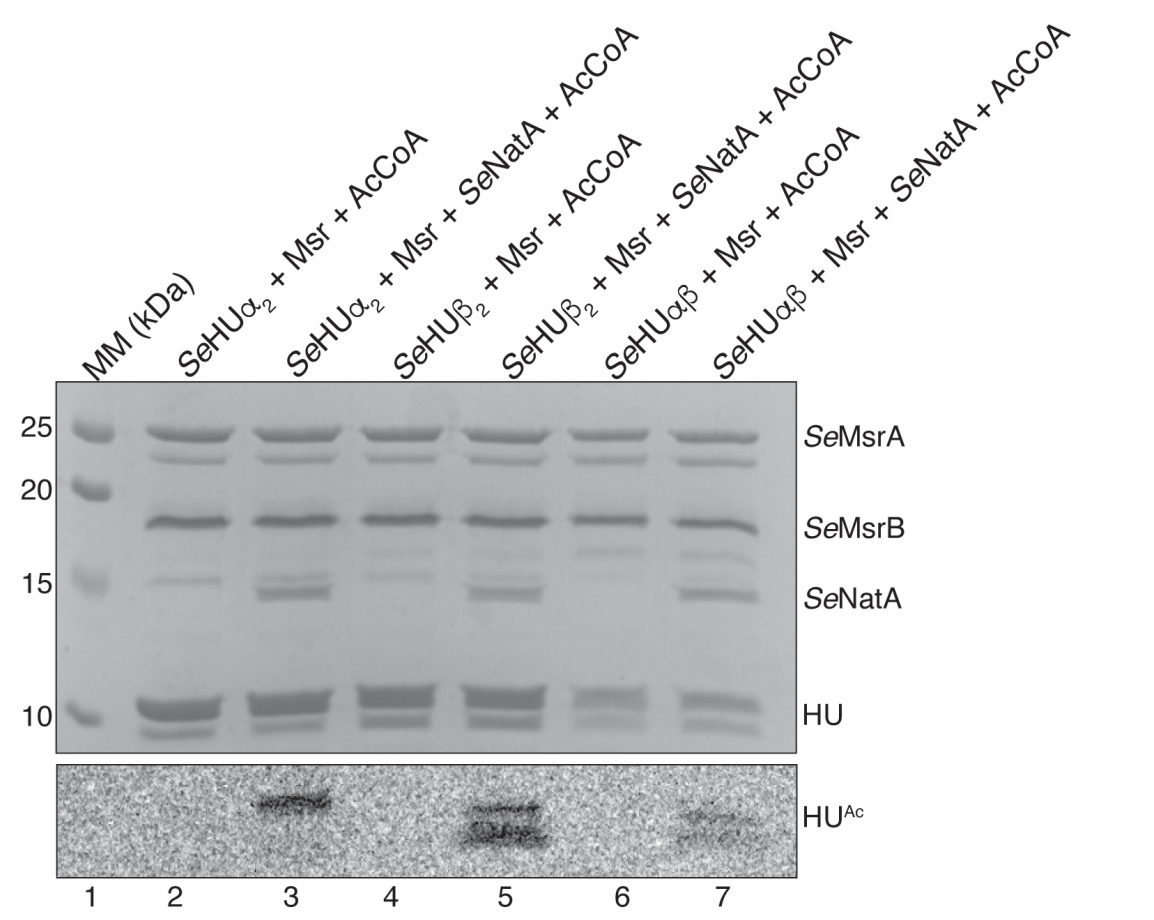

Figure 6.12. Repair of Met-SO by Msr proteins increase *Se*NatA-mediated acetylation. Radiolabeled acetylation assays were conducted via incubation of either single iMet *Se*HUβ_2, *Se*HUα_2, or *Se*HU$\alpha\beta$ proteins (5 μM each) for one hour at 37°C with 5 μM each of MsrA and MsrB proteins with 20mM DTT. After an hour, 40 μM [acetyl-1-^{14}C]-AcCoA and *Se*NatA (2 μM) was added and reactions incubated an additional hour. Reactions were resolved by SDS-PAGE and imaged as described for other acetylation assays.

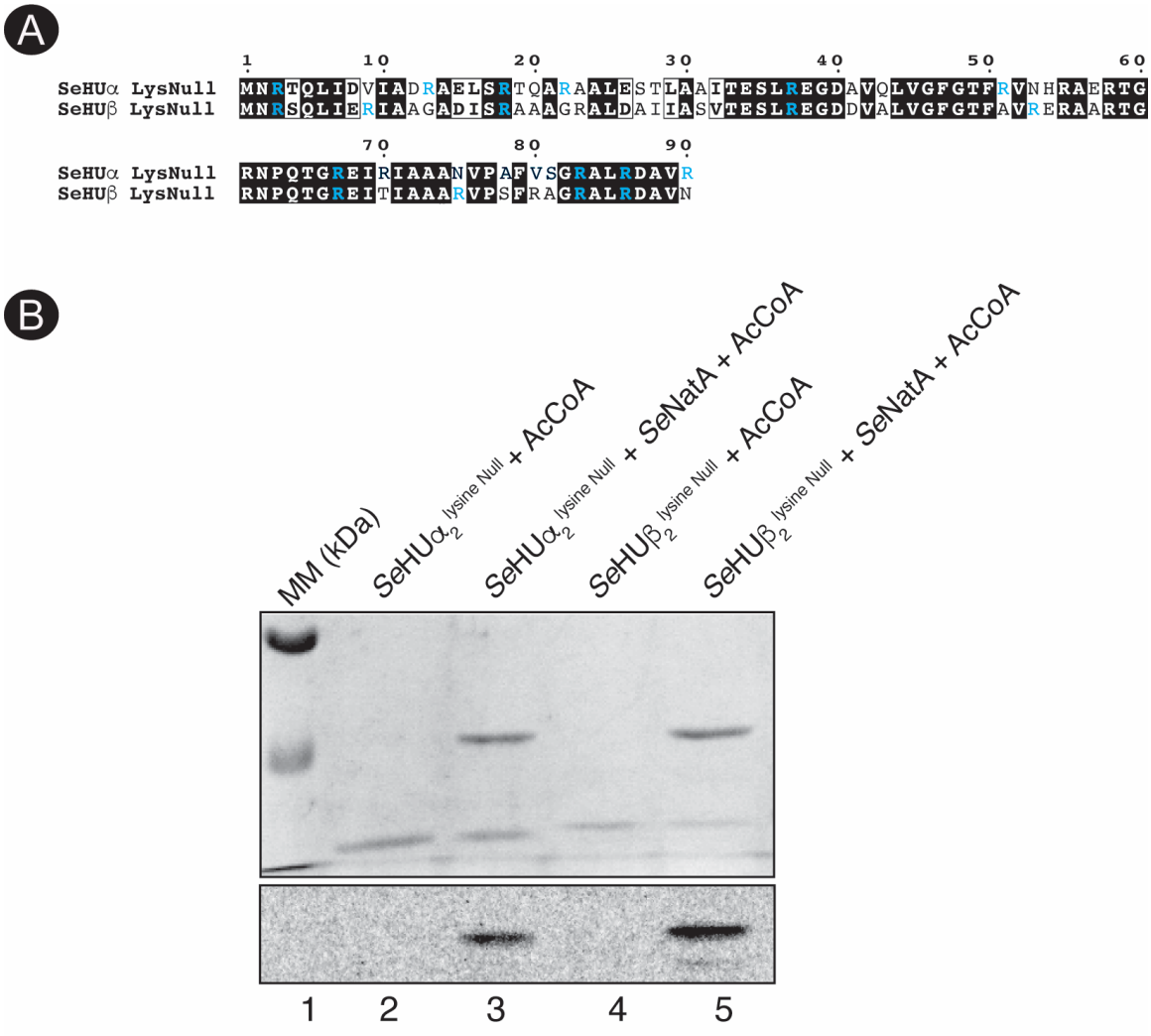

Figure 6.13. *Se*NatA does not acetylate lysyl residues of *Se*HUα$_2$ or *Se*HUβ$_2$. Panel A shows all residues that were lysine to arginine changes. B. To verify the N-terminus as the site of acetylation, lysine null variants of *Se*HUα$_2$ (HUα$_2^{K3R,K13R,K18R,K22R,K37R,K51R,K67R,K83R,K86R,K90R}$) and *Se*HUβ$_2$ (HUβ$_2^{K3R,K9R,K18R,K37R,K53R,K67R,K75R,K83R,K86R}$) were purified and tested as substrates of *Se*NatA and [acetyl-1-^{14}C]-AcCoA for acetylation. This experiment was conducted as described in the legend to figure 1 where lysine null proteins were incubated with [acetyl-1-^{14}C]-AcCoA in the presence or absence of *Se*NatA.

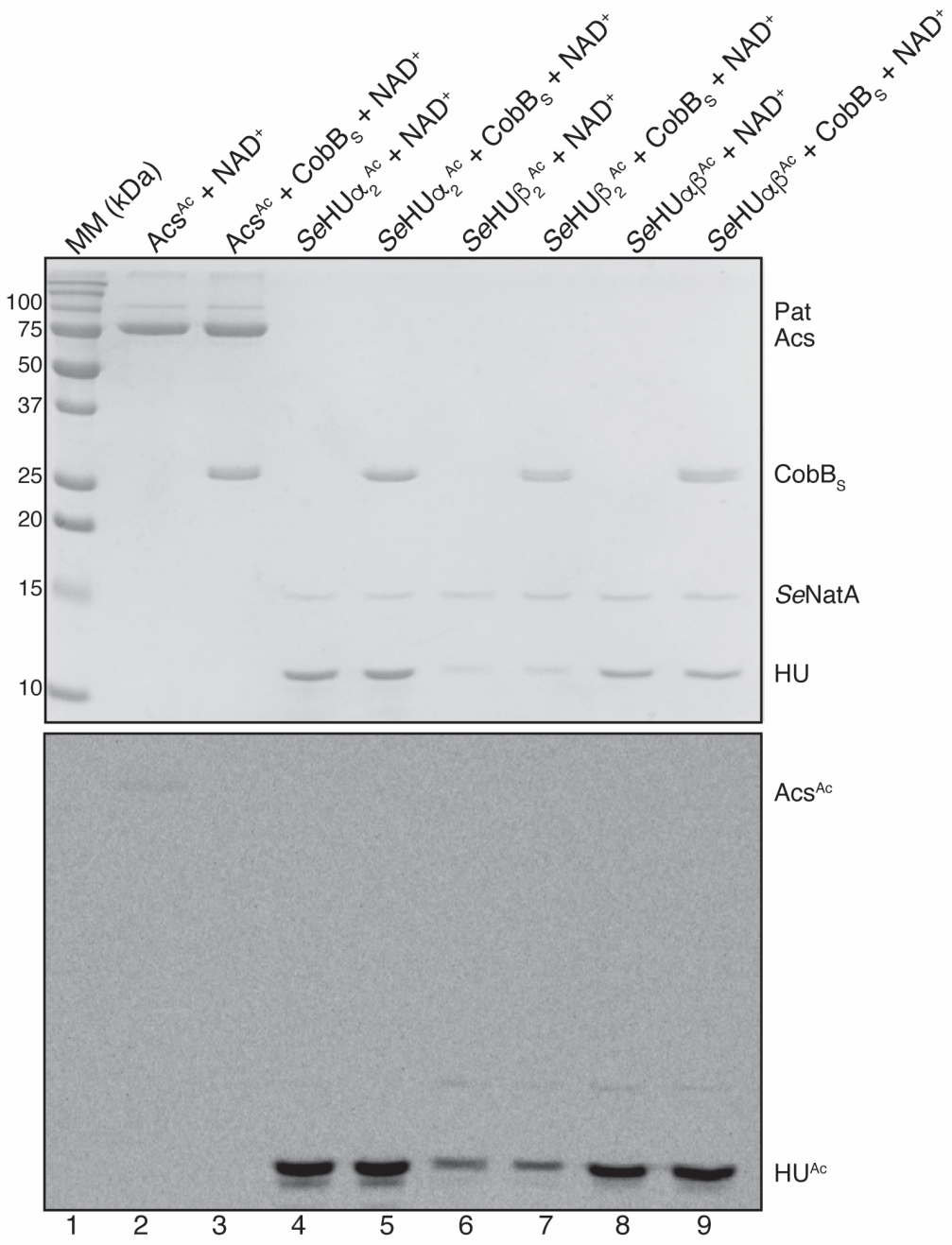

Figure 6.14. *Se*NatA-mediated acetylation of HU cannot be deacetylated by the CobB sirtuin deacylase. [Acetyl-1-^{14}C]-labeled HU proteins were incubated with NAD$^+$ with CobB$_S$ (lanes 5, 7, and 9,) and without CobB$_S$ (lanes 4, 6, and 8,). Positive controls of CobB deacetylating [acetyl-1-^{14}C]-Acs (lanes 2 and 3) was included. Samples were resolved by SDS-PAGE and radiolabel transfer was visualized by phosphor imaging analysis.

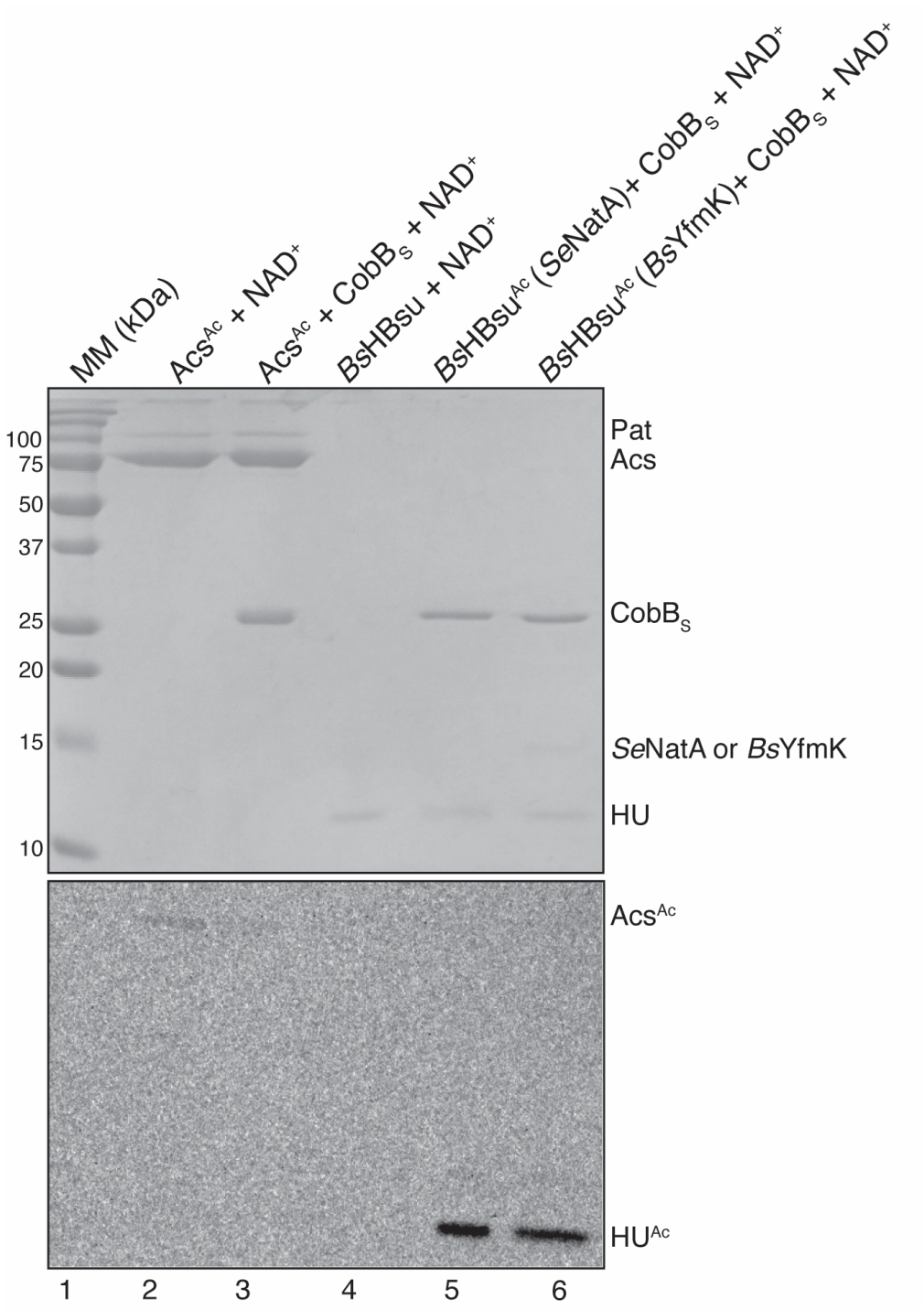

Figure 6.15. HBsu acetylated by either *Se*NatA or *Bs*YfmK cannot be deacetylated by the CobB sirtuin deacylase. [Acetyl-1-^{14}C]-labeled HBsu protein was incubated with NAD^+ with $CobB_S$ (lanes 5 and 6). Positive controls of CobB deacetylating [acetyl-1-^{14}C]-Acs (lanes 2 and 3) was included. Samples were resolved by SDS-PAGE and radiolabel transfer was visualized by phosphor imaging analysis.

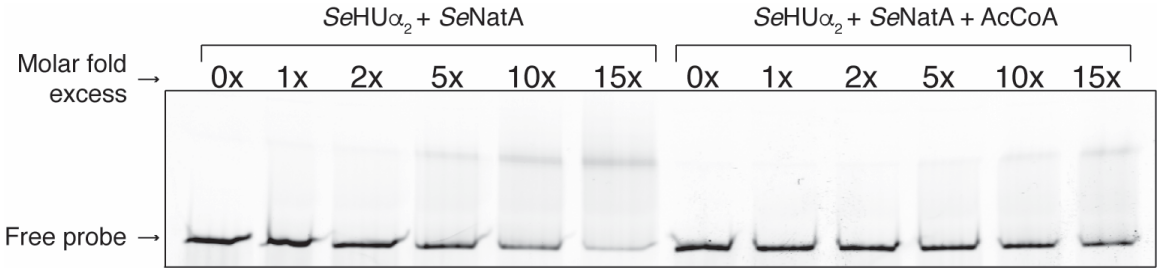

Figure 6.16. N-terminal acetylation of HUα2 affects its DNA binding ability *in vitro*. HUα2 was either pre-acetylated with *Se*NatA and cold acetyl-CoA or incubated with *Se*NatA and no AcCoA in HEPES buffer. $HU\alpha_2^{Ac}$ or HUα2 proteins were added at increasing fold excess of protein to dsDNA hilA probe and incubated at room temperature for 10 min. Reactions were resolved on a 7.5% TBE gel and imaged using a Typhoon Trio$^+$ imager at wavelength 488 nm.

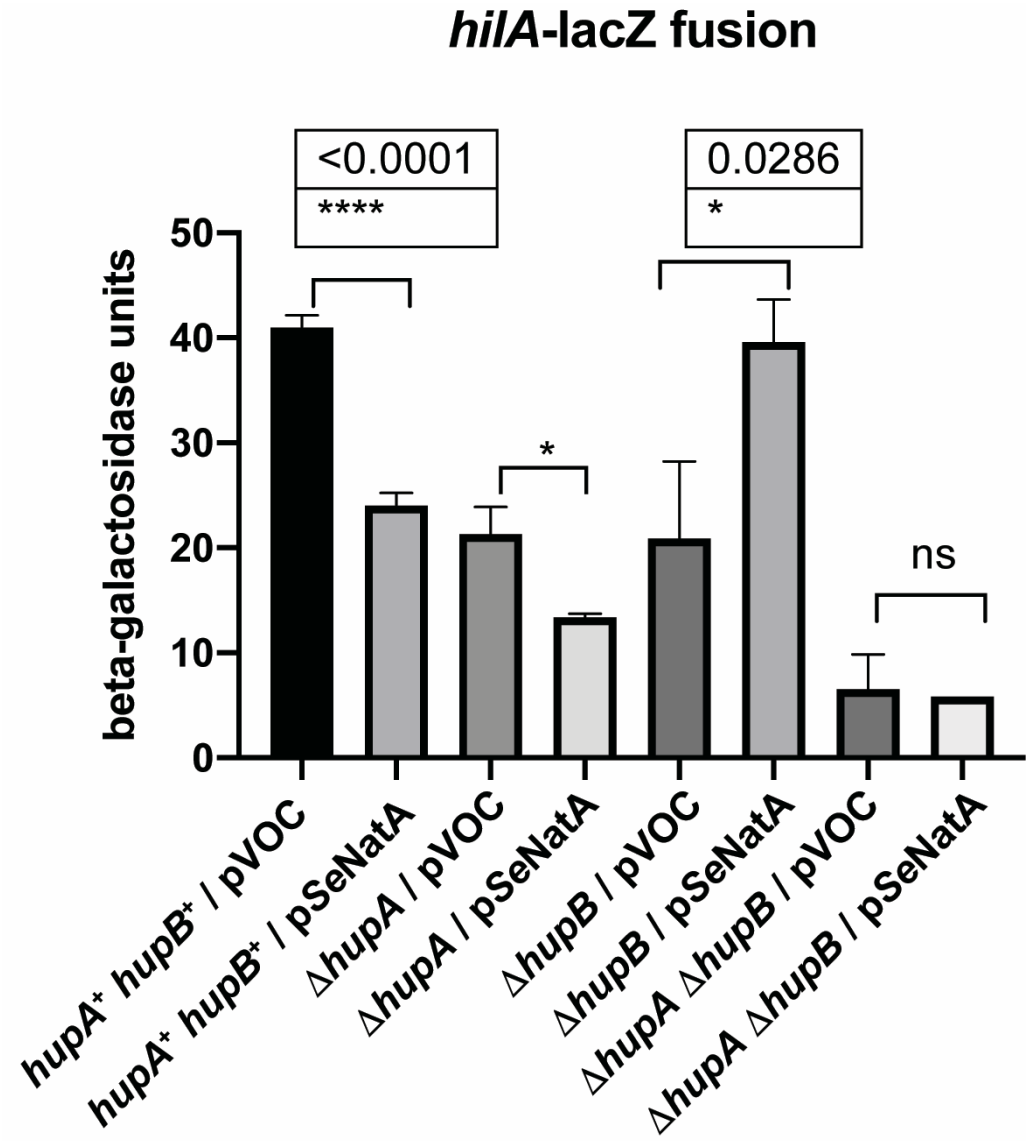

Figure 6.17. SeNatA affects *hilA* gene expression in a HU-dependent manner. β-galactosidase assays (Miller assays) were used to assess the activity of the hilA gene in the presence of overexpressed SeNatA as compared to the absence of SeNatA. Cells contain a chromosomally encoded lacZ fused to the promoter region of the *hilA* gene. Cells were grown under SPI-1 inducing conditions of low oxygen, 37°C, stationary in LB medium containing L-(+)-arabinose (0.5mM) to mid-log and were assayed for β-galactosidase units per strain. Strains were grown in biological duplicate and the experiment was conducted three independent times.

CHAPTER 7

CONCLUSIONS AND FUTURE DIRECTIONS

7.1 CONCLUSIONS

Acetylation has shown to be an important regulatory mechanism in all domains of life and provides the ability to rapidly respond to external stimuli through the chemical modification of proteins or small molecules that already exist in the cell. This body of work expands our knowledge of several different acetylation/deacetylation systems in the bacterium *Salmonella enterica*.

In chapter 3, we found the first instance of an acetyltransferase (Name TacT) that can acetylate a small molecule (amino charged tRNA) and the lysine of a protein (the cognate antitoxin, TacA). This shows the flexibility of this acetyltranferase to perform N-ε and N-α acetylation, but not so flexible that it is not specific for its two substrates. We also found that acetylation of the antitoxin TacA is reversible by the protein lysine deacetylase CobB, and we predict that this allows entrance and exit of a persistence state in infection.

In chapter 4, another novel discovery was made where an uncharacterized GCN5-family acetyltransferase, named YiaC, is able to N-terminally acetylate protein substrates. We rename this enzyme to NatA and show that NatA is able to acetylate the sirtuin deacetylase $CobB_{Long}$. This is one of the first, if not the first, instance of acetylation of a sirtuin in bacteria. We found that acetylation of $CobB_L$

affects its enzymatic activity in vitro and in vivo, which modulates the flux of acetate and acetyl-CoA in *S. enterica*.

In chapter 5, we try to understand the physiological relevance of *S. enterica* possessing two biologically active isoforms of the sirtuin deacetylase CobB. The long isoform of CobB has an arginine-rich N-terminal extension. We tested whether the N-terminus of CobBL is interacting with the positively charged phosphate backbones of nucleic acids, as well as if it can interact with the negative charge of lipids. A phenotype between the CobB isoforms was found during growth utilization of the carbon source ethanolamine that specifically involves the N-terminal arginines on $CobB_L$, however the reason for the differences in growth observed between strains encoding different isoforms needs further investigation, as well as what substrate is the physiologically relevant substrate to interact with $CobB_L$'s N-terminal extension.

In chapter 6 a novel substrate for NatA is identified as the nucleoid-associated proteins HU. We provide in-depth analysis of the in vivo N-termini of the HU proteins, as well as what conditions of the N-termini are required for NatA-mediated acetylation. In doing so we found the HU proteins are a novel substrate for the methionine sulfoxide reductases MsrA and MsrB, and suggest that the HU protein N-termini may need to be reduced in vivo, especially in order for NatA-mediated acetylation to occur. We also found that the previously studied *Bacillus subtilis* acetyltransferase YfmK is a functional homologue of NatA, and provide in vitro evidence that YfmK is in fact a protein N-terminal acetyltransferase.

The appendices of this dissertation outline several manuscripts that are in preparation for a side project that decided to come to fruition at the end of my tenture. I hope you can get a taste for those rudimentary manuscripts as the appendices attached to this dissertation. We are excited to characterize a novel transcriptional regulator that, at this time, is named PagR. Our data suggest that PagR regulates several iron-acquisition related genes, as well as several genes involved in metabolism and with flux of amino acids.

7.2 FUTURE DIRECTIONS

Chapter 2 of this dissertation is a review of protein N-terminal acetylation in bacteria. That review plays a large role in the future directions of this body of work, as the field of protein N-terminal acetylation is just getting started, and is even less studied in prokaryotes than in eukaryotes. In researching N-terminal acetylation I was made aware of how intricately connected other protein N-terminal modifications are to N-terminal acetylation and perhaps to protein synthesis itself. Hopefully new tools and discoveries will be made soon in this field of study that can help expand the work of chapters 4 and 6. In the meantime, a lot of studies can be done to understand *Se*NatA and *Bs*YfmK as well as other potential protein substrates may have been found already in our lab.

For the future of chapter 5, I generated a model for how $CobB_{Long}$ may be interacting at the cytosolic face of the lipid bilayer. Perhaps $CobB_{L}$ is working with the O-acetyl-ADP-ribose deacetylase named YmdB, which deacetylates the O-Acetyl-ADP-ribose generated from CobB, and to the best of our knowledge, only

CobB. YmdB is translationally fused to the cardiolipin synthase ClsC, which catalyzes the synthesis of cardiolipin from phosphatidylglycerol and phosphatidylethanolamine. I hypothesize that perhaps the phenotypic differences seen of the CobB isoforms may be related to ethanolamine incorporation into cardiolipin. Perhaps a connection can be made by assessing the acetate paradigm of reversible lysine acetylation, acetyl-CoA synthetase Acs, Protein-lysine acetyltransferase Pat, and sirtuin deacetylase CobB's involvement in YmdB and ClsC in *S. enterica*.

APPENDIX A

IN SALMONELLA ENTERICA, PAGR REGULATES ITS OWN EXPRESSION AND THE EXPRESSION OF A FIVE-GENE OPERON THAT ENCODES AN ALTERNATIVE TRANSKETOLASE[7]

[7]Parks A.R., Byrne J.A., and Escalante-Semerena J.C. To be submitted.

ABSTRACT

The enteropathogen *Salmonella enterica* subsp. *enterica* sv Typhimurium str. LT2 (hereafter *S. enterica*) utilizes a cluster of genes encoded within the pathogenicity island 2 (SPI-2) of its genome to proliferate inside macrophages. The expression of SPI-2 is controlled by a complex network of transcriptional regulators and environmental cues, of which now include a recently characterized DNA binding protein named PagR. During low phosphate and low magnesium conditions, such as conditions experienced inside macrophages, PagR is involved in the signal cascade of SPI-2 induction by upregulating the transcription of *slyA*, a known activator of SPI-2 expression during survival in macrophages. Here we report that PagR acts as a transcriptional repressor for the divergent polycistronic operon that encodes the two subunits of the third transketolase (TktCD) of *S. enterica*. TktCD contributes to the non-redox rearrangements of phosphorylated sugars of the pentose phosphate pathway, which may play a role in providing important building block intermediates to *S. enterica* during intracellular proliferation. We have found that de-repression of the *tktC and tktD* genes encoding the subunits of TktCD negatively impacted the activity of other thiamine-pyrophosphate (TPP)-dependent enzymes such as 2-ketoglutarate decarboxylase (KGD) and pyruvate dehydrogenase (PDH) during growth when *tktCD* expression was de-repressed. PagR regulation of SPI-2 and *tktCD* links central carbon metabolic requirements to virulence.

INTRODUCTION

Transketolases are well-characterized enzymes that catalyze the non-oxidative transfer of a 2-carbon dihydroxyethyl moiety from a ketose to an aldose using thiamine pyrophosphate as a carrier. Transketolase-catalyzed reactions are critical in processes such as the non-oxidative branch of the pentose phosphate pathway (non oxPPP) {Stincone, 2015 #31401} and the Calvin-Benson-Basshan (CBB) cycle {Raines, 2003 #31406}. Up until recently, the genome of the enteropathogenic bacterium *Salmonella enterica subsp. enterica* serovar Typhymirium str. LT2 (hereafter *S. enterica*) was thought to encode only two transketolases. During the analysis of the genomic context of the *pagR* gene, we noticed that *pagR* was transcribed from a set of genes encoding a putative sugar transporter (STM2342, STM2343, STM2344) and two genes encoding the subunits of a new transketolase (*tktC* = STM2340 and *tktD* = STM2341) that we refer to as TktCD to reflect the fact that the enzyme is a two-subunit complex (Fig. 1). TktCD was recently characterized as an isozyme of transketolase A and transketolase B {Shaw, 2018 #31383}. The housekeeping transketolases TktA and TktB are thiamine-pyrophosphate (TPP)-dependent enzymes involved in the non-oxidative branch of the pentose phosphate pathway (non-oxPPP) where they perform reversible, rate-limiting steps of interconversion between ketose and aldose sugars through the transfer of two-carbon moieties. The reversibility of reactions catalyzed by transketolase enzymes involved in the non-oxPPP allows the production of intermediate compounds to meet metabolic demands of the cell through the diversion of metabolites to other pathways such as replenishment of

glycolysis, or to provide building blocks for the biosynthesis of aromatic amino acids, nucleotides or lipopolysaccharides {Stincone, 2015 #31401}.

MATERIALS AND METHODS

Bacterial strains. All strains were derivatives of *Salmonella enterica* subsp. *enterica* serovar Typhimurium str. LT2 (*S. enterica*). Gene deletions were constructed using the protocol that employs l Red recombinase to engineer in-frame deletions as described elsewhere {Datsenko, 2000 #2595}.

Strain construction. All strains constructed were *metE205 ara-9* derivatives of *S. enterica*. Primers used in this study were synthesized by Integrated DNA Technologies, Inc. (IDT, Coralville, IA), and are listed in Table S2 in the Supplemental Material. The *pagR*::*kan*$^+$ and *tktC-tktD*::*cat*$^+$ strains were engineered following the Wannder and Dansko Datsenko (2000) protocol as follows: PFU Ultra II Fusion DNA polymerase (Stratagene) was used to amplify flanking regions of plasmids pkD3 (*cat*$^+$ marker) or pkD4 (*kan*$^+$ marker) using primers designed with 36 base pairs of overlapping region to the beginning or the end of *pagR* or *tktC-tktD* genes. Polymerase chain reaction (PCR) amplifications were visualized by post-staining with ethidium bromide (0.5 μg /ml) for 10 min. Products were cleaning with the Wizard SV gel and PCR clean up kit (Promega), and ~200ng of PCR product was electroporated into *S. enterica* strain JE6692 (*metE205 ara-9* / pKD46 *bla*$^+$) using a 0.2-cm electroporation cuvette (MidiSci) and a microPulser electroporator (Bio-Rad Laboratories) on Ec2 setting. Cells were

recovered by incubation at 37 ºC with shaking and plated on lysogeny broth (LB) agar supplemented with either 12.5 μg/ml of chloramphenicol for the *cat*$^{+}$ marker or 25 μg/ml for the *kan*$^{+}$ marker. Correct antibiotic insertions were PCR verified then moved by P22-mediated transduction into strain JE6583 as described elsewhere {Davis, 1980 #11417}. Colonies were streaked to isolation and select colonies were Sanger sequenced to verify correct insertion location and sequence using primers flanking genes of interest.

Plasmid construction for complementation of function and gene overexpression. All plasmids used here are listed in Table S3. Primers used in this study were synthesized by Integrated DNA technologies, Inc. (IDT; Coralville, IA), and are listed in Table S2. The high-efficiency cloning method of BspQI digestion described elsewhere {VanDrisse, 2016 #25211} was used to clone *pagR*, *tktC-tktD*, and *sucA* into plasmids pCV1, pCV3, or pTEV19. All genes cloned were PCR amplified from *S. enterica* strain JE6583 genomic DNA using PFU Ultra II Fusion DNA polymerase (Stratagene). PCR products were visualized by DNA gel electrophoresis on 1% (w/v) agarose gels that were post-stained with ethidium bromide (0.5 mg/ml). PCR products were cleaned using the Promega Wizard SV gel and PCR clean up kit, and cloning was conducted as described previously (VanDrisse 2016). After transformation into *E. coli* DH5a cells and colony PCR screening of correctly ligated plasmids, correct plasmids were isolated using the Wizard Plus SV miniprep kit (Promega). Sanger DNA sequencing was performed by Eton Bioscience to rule out mutations and confirm insertions.

***In vivo* growth analyses including culture media and chemicals.** All bacterial strains and plasmids used in this study are listed in Table S1 and S3. Growth studies were conducted as follows: starter bacterial cultures were grown from a single colony in nutrient broth (NB, Difco) with overnight shaking at 180 rpm at 37 ºC. In the morning, cells were sub-cultured into no-carbon essential (NCE) minimal medium {Berkowitz, 1968 #11328} containing MgSO4 (1mM), Wolfe's trace minerals {Balch, 1976 #11539}, ampicillin (100 μg/ml), L-(+)-arabinose (concentration designated in figure legends) and either sodium succinate (30 mM), *myo*-inositol (55 mM), or sodium pyruvate (20 mM) as carbon and energy sources. Cultures were inoculated at 1% (v/v), 2 μl of culture was used to inoculate 198 μl of fresh medium and pipetted into individual wells of a 96-well polystyrene plate (Falcon). Microtiter plays were incubated at 37 ºC shaking continuously inside a PowerWave microtiter plate reader (Bio-Tek Instruments). Density of cells was monitored hourly at 630 nm and data were analyzed using Prism 8 (GraphPad).

RNA isolation. Strains JE22070 (*pagR*$^{+}$ / vector), JE21566 (*pagR*::*kan*$^{+}$ / vector), JE21577 (*pagR*::*kan*$^{+}$ / pPagR), were grown overnight in triplicate in nutrient broth (2 ml; NB, Difco) with shaking at 37 ºC. After incubation, strains were diluted 1:100 into 5 ml of fresh LB rich medium supplemented with L(+)-arabinose (100 μM). Cultures were grown shaking at 37 ºC to an optical density (600 nm) of 0.5, then 5 ml of each sample were quickly centrifuged in 15 mL Falcon tubes at 4000 x

Kushner

, supernatant was removed, and pellets were flash-frozen in liquid nitrogen and kept on dry ice. RNA was isolated following the RNAsnap™ protocol [1]. Pellets were re-suspended in 150 μl of boil solution (ethylenediaminetretraacetic acid (EDTA, 18 mM), SDS (0.025%, w/v) formamide (95%, v/v; RNA grade), 2-mercaptoethanol (1% v/v) in RNase-free water) and were vortexed vigorously to break up the cell pellet. Pellets were incubated at 95 ºC for 7 min and centrifuged at 16,000 x *g* for 5 min at room temperature; 100 μl of supernatant was transferred to a fresh tube. A sodium acetate/ethanol RNA precipitation was then conducted by the addition of 400 μl of RNase-free water, 50 μl of sodium acetate (3 M, pH 5.2; final concentration of 0.3 M), and finally 1650 μl of ice-cold absolute ethanol (100%), with mixing briefly before the addition of the next reagent. The mixture was incubated at -80 ºC for one hour, centrifuged at 16,000 x *g* for 30 min at 4 ºC, and ethanol was decanted. Ethanol (300 μl of cold 70% v/v) was added and pellets were centrifuged at 8,000 x g for 5 min at 4ºC in an Eppendorf 5415D centrifuge. Ethanol was removed and pellets were allowed to dry. RNA pellets were re-suspended in RNase-free water. Subsequent RNase-free DNase I treatment was conducted using the Ambion Turbo DNA-free kit according to manufacturer's instructions (ThermoFisher Scientific). After DNA cleavage, a final sodium acetate/ethanol precipitation was performed as described above, except using 360 μl of water, 50 μl of 3 M sodium acetate, and 1500 μl of cold 100% ethanol. After a one-hour incubation at -80 ºC, RNA was centrifuged at 16,000 x *g* for 30 min, then washed with 300 μl of cold 70% ethanol (v/v). RNA decanted of ethanol, and

pellets, were dried for 20 min at room temperature. The final resuspension of RNA was resuspended in 100 μl of water. A small sample of each preparation was used to quantify using the RNA Broad Range (BR) Assay kit by Qubit on a Qubit 4 fluometer. A small amount of each preparation was also tested for quality and integrity using the Qubit RNA IQ Assay on a Qubit 4 fluometer. Primers for qPCR were designed using Primer 3.

cDNA synthesis and quantitiative real time polymcerase chain reaction (RT-qPCR). Total RNA (972 ng) from each sample was used for the synthesis of cDNA using the iScript™ cDNA synthesis Kit from Bio-Rad Laboratories according to manufacturer's protocol. Each cDNA reaction was then diluted to 5 ng/μl and used as template for PCR. For real-time PCR, 20 μl reactions were prepared with 10 μl of 2X FastSYBR Green master mix (Applied Biosystems), 500 nM of each gene-specific primer (1μl of 10 μM primer stock), and 10 ng of cDNA (2 μl of 5 ng/μl cDNA). The real-time PCR reaction was performed using a 7500 Fast real-time PCR system (Applied Biosystems). The threshold cycle values of *rpoB* and *gyrB* were checked first to ensure that both genes were optimal for use as reference genes for these strains under the conditions chosen for RT-qPCR. Cycle threshold (C_T) data were normalized to the *rpoB* gene {Rocha, 2015 #26344}. These normalized values (ΔC_T) were transformed using $2(e\text{-}\Delta Ct)/10^{-6}$ [2], and were reported as arbitrary gene expression units (EU), or the gene expression ratio of the mutant strains/the parent strain (JE22070 *pagR*$^{+}$). Mean EU values were used to calculate the standard error of the mean (SEM) using Prism8 from three

biological replicates that were each tested in technical triplicate. Differences in EU between mutant strains and JE22070, were compared using Welch's t-test with the GraphPad Prism8 software.

Operon PCR. As described above, total RNA was isolated from strain JE21107 and was used to generate cDNA, and genomic DNA isolated from strain JE21107 (as described elsewhere REF) was used to in PCR reactions containing Green GoTaq master mix to amplify overlapping genes within the STM2340-STM2344 operon.

Purification of PagR. PagR protein was purified to homogeneity from plasmid pPagR-8 encoding a PagR protein with a maltose binding protein-hexahistidine (MBP-H6) fusaed to its *N* terminus. The tag was removed after incubation with recombinant tobacco etch virus (rTEV) protease since the plasmid used to produce MBP-H_6-PagR (pTEV19) contained a rTEV protease cleavage site {VanDrisse, 2016 #25211}. A sample (10 ml) of an overnight culture of *E. coli* C41 (λDE3) / pPagR-8 strain was used to inoculate one liter of lysogeny broth (LB) containing 100 μg/ml of ampicillin. Cells were grown shaking at 125 rpm at 37 °C until the culture reached an optical density (OD) at 600 nm of 0.7. Expression of genes of interest encoded by the plasmids was induced by the addition of IPTG (0.25 mM) followed by ~12 h of overnight incubation at 37 °C. The next morning cells were harvested by centrifugation at 6,000 x *g* for 15 min using a refrigerated Beckman-Coulter Avanti J-20-XPI centrifuge equipped with a JLA 8.1 rotor. Cell pellets were

resuspended in 20 ml of buffer A [(4-(2-hydroxyethyl-1-piperazineethanesulfonic acid (HEPES) buffer (50 mM, pH 7.5 at 4 ºC) containing NaCl (0.5 M), glycerol (20% v/v), and imidazole (20 mM)] and were sonicated thrice for 30-s intervals, and during each interval, sonication was on for 2 s and off for 2 s, at 60% amplitude. The resulting whole-cell lysates were centrifuged for 30 min at 40,000 x g and the supernatants were filtered with a 0.45-μ filter to remove large particulates. Each filtered lysate was applied onto a 1-ml nitrilotriacetic acid (NTA) affinity chromatography column pre-equilibrated with buffer A. Fractions were collected by gravity at 4 °C. The purification was performed as follows: After all the lysate was applied, the column was washed with 10 column volumes (CV, *i.e.,* 10 mL) of buffer A, 7 CV (*i.e.,* 7 mL) of buffer A containing 4% elution buffer B [HEPES buffer (50 mM, pH 7.5 at 4 ºC) containing NaCl (0.5 M), glycerol (20% v/v), and imidazole (0.5M)], and finally, MBP-H_6-PagR was eluted in two fractions, first with 1 CV (*i.e.,* 1mL) of 100% elution buffer B, and the second fraction being 4 CV (*i.e.,* 4mL) of 100% elution buffer B. Both elution fractions were pooled and MBP-H_6-PagR was cleaved with rTEV protease at a concentration of 1 mg rTEV:100 mg of PagR protein while dialyzing at room temperature in HEPES buffer (50 mM, pH 7.5 at 4 ºC) containing NaCl (0.5M), glycerol (10% v/v), and dithiothreitol (DTT, 1 mM). Cleaved protein was dialyzed twice more at 4 °C in buffer A. Cleaved STM2345 protein was loaded again onto a 2-ml NTA column to remove MBP-His and rTEV protease, both of which were fused to a hexahistidine tag. Tag-less PagR protein did not interact with the NTA resin, and was collected in the flow-

through fraction. To further remove MBP from PagR, flow-through fractions from the second NTA purification were run over a 1-mL amylose resin to remove contaminating MBP. Pure PagR was dialyzed overnight HEPES buffer (50 mM, pH7.5 at 4ºC) containing NaCl (150 mM) and glycerol (20% v/v) at 4˚C. Dialyzed PagR was stored at -80˚C in 15-μl aliquots that were flash frozen with liquid nitrogen. Protein concentration was determined by Nanodrop using the molar absorptivity and molecular mass of PagR.

Electrophoretic mobility shift assays Electrophoretic mobility shift assays (EMSAs) were performed to quantify PagR binding to DNA. EMSAs were performed as follows: PagR protein purified to >95% homogeneity was incubated at 0, 0.15, 0.3, 0.6, 1.2 and 2.4 pmol of protein with a 5'-6-FAM-labeled DNA probe (150 base pair promoter region of the STM2340-5 operon; 50 ng [*i.e.,* 0.3 pmol] was added). EMSA buffer [HEPES buffer (50 mM, pH 7.5 at 4ºC) containing NaCl (150 mM), and glycerol (10% v/v)] was added to the reaction mixture (total volume = 25 μl) and DNA and protein were incubated at room temperature for 40 min. During incubation, a 7.5% Tris-boric acid-EDTA (TBE) polyacrylamide gel was pre-developed at 100 V for 40 min in 0.5X TBE buffer at 4˚C. After incubation, 5 μl of glycerol (50% v/v) was added to the reaction mixtures, and 20 μl of each reaction mixture was resolved by the polyacrylamide gel. A lane of xylene cyanol and bromophenol blue dye was added as a tracking indicator, and the gel was run until

bromophenol blue reached the bottom of the gel. The gel was imaged using a Typhoon Trio imager at 525 nm with the 488 (Blue) filter.

RESULTS

A *S. enterica* Δ*pagR* strain struggles to grow on succinate or *myo*-inositol as a sole carbon and energy source. We probed for areas of *S. enterica* metabolism that might be regulated by PagR (STM2345). To this end, we tested the ability of a *S. enterica* Δ*pagR* strain to catabolize various carbon sources. During this screen we observed two notable growth phenotypes. First, a Δ*pagR* strain (JE21566) showed a delay in the onset of exponential growth and a slower growth rate relative to the *pagR*$^{+}$ strain when succinate was the sole carbon and energy source (Fig. A.2, gray triangles *vs* red squares). As expected, we corrected the phenotype by ectopically expressing the *pagR*$^{+}$ allele (Fig. A.2, green squares). Surprisingly, we also corrected the growth phenotype by deleting the neighboring genes encoding the two subunits of the TktCD trasketolase (Fig. A.2, magenta diamonds). The latter was a notable result because the correction of the phenotype was as efficient as the ectopic expression of *pagR*$^{+}$. Also unexpected was the observation that the ectopic expression of the *tktCD*$^{+}$ genes delayed exponential growth mimicking the growth behavior of a Δ*pagR* strain (Fig. A.2, black squares).

Second, we observed a severe growth phenotype of the Δ*pagR* strain when *myo*-inositol was the sole source of carbon and energy provided (Fig. A.3, grey triangles *vs* red squares). The lack of growth with *myo*-inositol was solely due to the absence of PagR, as seen by the results of *pagR*$^{+}$ complementation (Fig. A.3, green

squares *vs* grey triangles). Here again, the deletion of the *tktCD* genes in the Δ*pagR* strain restored growth at a wild-type rate, but the delayed onset of exponential growth (~30 h) was similar to the one observed with succinate (Fig. A.3, magenta diamonds). Regardless, the Δ*pagR* Δ*tktCD* strain grew fast, and reached the same cell density as the *pagR*$^+$ strain (Fig. A.3, red squares *vs* magenta diamonds). Notably, ectopic expression of *pagR*$^+$ reduced the lag time of the Δ*pagR* Δ*tktCD* strain, matching the growth behavior of the Δ*pagR* / *pagR*$^+$ strain (Fig. 3, magenta diamonds *vs* yellow squares). However, ectopic expression of the *tktCD* genes in the Δ*pagR* Δ*tktCD* / pTktCD strain was deleterious in that it extended the lag phase by ~8 h, and finally, if the *tktCD* genes were ectopically expressed in a Δ*pagR tktCD*$^+$ strain (JE25520), growth was completely arrested (Fig. A.3, black squares).

PagR (STM2345) represses the expression of the *tktCD* genes. Given the close proximity of *pagR* to the *tktCD* operon (Fig. A.1), we investigated the possibility that PagR repressed the expression of the *tktC* and *tktD* genes. We isolated RNA from three strains, namely *pagR*$^+$ / vector, *pagR*::*kan*$^+$ / vector, and *pagR*::*kan*$^+$ / pPagR, and quantified the amount of *tktC* and *tktD* mRNA using qRT-PCR. For this purpose, cells were grown on LB plus ampicillin and 100 μM arabinose and RNA was isolated when cultures reached an OD_{600} of 0.5. As shown in figure A.4A and A.4B, very little expression of *tktC* or *tktD* occured in the *pagR*$^+$ or Δ*pagR* / pPagR complemented strain (blue and black bars, respectively). In contrast, expression of *tktC* and *tktD* expression increased ~3 fold in the Δ*pagR* strain (red

bars). These data suggested that PagR was likely a transcriptional repressor of the *tktCD* genes, a conclusion that was consistent with its annotated homology to LacI family of repressors.

The absence of PagR triggers high levels of expression of TPP biosynthetic genes, but during growth with succinate, exogenous thiamine or thiamine precursors do not restore growth of a Δ*pagR* strain. We posited that the phenotype of the Δ*pagR* strain might be due to the sudden high demand for TPP, a demand the cell could not meet. The growth behavior of the Δ*pagR* strain devoid of TktCD activity (strain JE25523) supported this idea because the Δ*pagR* Δ*tktCD strain* grew as well as the *pagR*$^{+}$ strain in succinate medium supplemented with thiamine (100 nM) (Fig. A.5, compare black circles to blue circles and grey squares).

The fact that the absence of TktCD allowed growth of the Δ*pagR* strain suggested that when express over physiological levels, TktCD could somehow reduce the level of TPP available to other TPP-dependent enzymes. Surprisingly, supplementation of thiamine (Fig. A.5A), TPP precursors like 4-methyl-5-hydroxyethyl)thiazole (THZ) plus 4-amino-5-hydroxymethyl-2-methylpyrimidine (HMP) (Fig. A.5B), or TPP (Fig. A.5C), all failed to restore growth of the Δ*pagR* strain when succinate was provided as the sole source of carbon and energy (grey triangles in all panels)

Addition of intermediates of the purine biosynthetic pathway rescues growth of a Δ*pagR* strain on succinate. In *S. enterica,* the synthesis of thiamine-diphosphate (TPP) shares the first five steps of the purine biosynthetic pathway Fig. A.6A).

As shown in figure 6C, in *S. enterica*, the last shared intermediate between inosine and TPP biosynthesis is aminoimidazole ribotide (AIR) {Bazurto, 2011 #24440}. Given that neither thiamine, its precursors nor TPP restored growth of the Δ*pagR* strain with succinate, we took an alternative strategy. We supplemented the medium with compounds that would increase flux through the purine biosynthetic pathway and presumably through the branch of the TPP pathway that synthesizes the precursor HMP-P (Fig. A.6A). The first compound we tested was L-glutamine, which in combination with phosphoribosyl pyrophosphate is converted into 5-phosphoribosylamine (PRA) by the amidophosphoribosyltransferase (PurF) enzyme {Bazurto, 2011 #24440} (Fig. A.6A). As shown in figure A.6B, the addition of L-glutamine substantially stimulated the growth of a Δ*pagR* strain during growth on succinate (compare grey triangles to red squares). To ascertain whether the observed difference in growth kinetics could be due to a limitation in ribose, we added 500 μM of ribose to the succinate medium. The addition of such a limited amount of ribose had a striking effect on both the lag and the growth rate of the Δ*pagR* strain (compare Fig. A.2 *vs* Fig. A.6C, grey triangles). However, a discernable growth phenotype was also observed between Δ*pagR* and the *pagR*$^{+}$ strain even when both were grown in the presence of ribose (0.5 mM) (Fig. 6C black triangles *vs* grey squares). That

difference in growth kinetics was abrogated by the ectopic expression of *pagR*$^+$ (Fig. A.6C, grey triangles *vs* open circles). Phosphoribosyl pyrophosphate has the possibility of feeding into the pentose phosphate pathway and be turned into D-ribose-5-phosphate, which could be acted upon by TktCD. This may explain the remaining growth defect seen, meanwhile glutamine feeds directly into AIR synthesis.

Overexpression of TPP-dependent enzymes rescues growth of a D*pagR* strain. Our initial hypothesis was that TPP-dependent enzymes were not saturated for TPP due to the higher-than-normal levels of TktCD. However, we never saw growth defects with the Δ*pagR* strain on minimal medium supplemented with pyruvate as the sole source of carbon and energy. This result was unexpected because the pyruvate dehydrogenase complex requires TPP to function, and yet it was not affected by the hypothesized TPP sequestration by TktCD. However, as shown in figure A.7, a growth defect of a Δ*pagR* strain growing with pyruvate could be observed only if the *tktCD* genes were overexpressed ectopically. We observed progressively deleterious effects as the concentration of inducer (L-(+)-arabinose) increased (Fig. 7, green, teal, magenta squares).

These results prompted us to test the idea that succinate catabolism could stall if the a-ketoglutarate dehydrogenase complex, which contains the TPP-dependent SucA enzyme, would be sub-saturated with TPP when TktCD was overproduced. In support of this idea, when *sucA* was overexpressed ectopically in a Δ*pagR* strain, growth with pyruvate was restored to near *pagR*$^+$ levels (Fig. A.8).

The *tktCD* genes are part of a 5-gene operon. To this point, all the data gathered strongly suggest that PagR repressed *tktCD* gene expression. We investigated the transcriptional context of the *tktCD* genes to assess whether PagR would bind to the putative promoter of *tktC* and *tktD*, blocking their transcription. Bioinformatics analysis suggested that genes *stm2340, stm2341*, *stm2342* and *stm2343* comprised a polycistronic operon (Biocyc.org). To test this possibility, we performed Operon PCR with RNA isolated from a Δ*pagR* strain. The RNA was treated with DNase, and cDNA was generated from purified total RNA. Genomic DNA, cDNA, and RNA were PCR amplified with four primer sets to generate products of four potential regions that overlap in neighboring putative genes (Fig A.10A). As shown in figure 10B, the use of such primers generated amplicons for the cDNA and the gDNA, but none amplified for the RNA, which served as our negative control. PCR amplification of all overlapping regions between genes *stm2340* and *stm2344* suggested these genes were a part of a polycistronic unit.

PagR binds directly to the its own promoter and to the promoter for the polycistronic mRNA encoding *tktCD*. To elucidate whether PagR directly bound to the promoter driving the expression of the *tktCD* genes, or was acting indirectly to repress the expression of the operon, PagR protein was purified to homogeneity and tested for its ability to bind to the putative promoter region of the *tktCD* operon (as depicted in figure A.10A). The putative promoter fragment of the polycistronic operon comprised of the *stm2340-stm2344* genes was incubated with increasing

concentrations of purified PagR protein then separated by native polyacrylamide electrophoresis (Fig. A.11).

DISCUSSION

We speculate that the pyruvate dehydrogenase complex may have a lower K_d for TPP than TktCD, but the data show that increasing levels of TktCD increase the lag time before the onset of exponential growth, a result that would be consistent with slower catabolism of pyruvate due to limiting TPP.

Figure A.1. Genomic context of the *pagR* and *tktC, tktD* genes. PagR activates the transcription of the five-gene operon shown in the figure. The genes encoding the TktCD transketolase are the most promoter-distal genes of the operon. All genes are at scale. *tktD* = 953 nt, *tktC* = 830 nt, *stm2342* = 1391 nt, *stm2343* = 272 nt, *stm2344* = 443 nt, *pagR* = 1019 nt. At present, the location of the promoters is hypothetical. The operon spans the region from 2,257,129 to 2,454, 101 nt of the chromosome. PagR spans the region from 2,457,382 to 2,458401 nt.

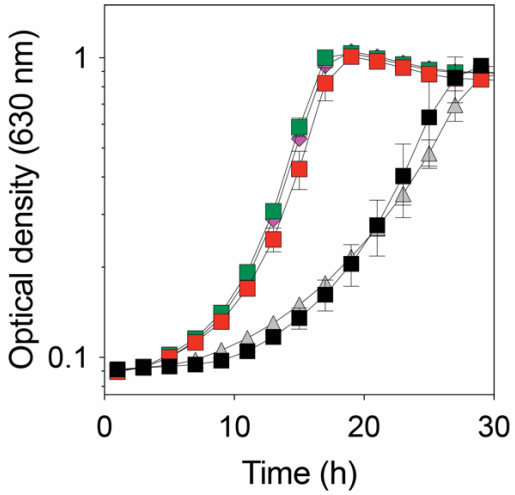

Figure A.2: Deletion of *pagR* causes a growth delay when succinate is the sole carbon and energy source, but a deletion of *tktCD* reverts the phenotype. *S. enterica* was grown on NCE (no-carbon essential) minimal medium {Berkowitz, 1968 #11328} supplemented with sodium succinate (30 mM) as the sole source of carbon and energy. The strains used were *pagR*$^{+}$ / vector (JE22070, red squares), *pagR*::*kan*$^{+}$ / pPagR (JE21577, green squares), *pagR*::*kan*$^{+}$ / vector (JE21566, grey triangles), Δ*pagR* D*tktCD* / vector (JE25523, magenta diamonds), Δ*pagR* / pTktCD (JE25520, black squares); 'vector' stands for the empty cloning vector pCV1 that contains an arabinose-inducible promoter {VanDrisse, 2016 #25211}. Ectopic gene expression was induced with 100 μM of L-(+)-arabinose. This experiment was conducted in technical triplicate of biological triplicates three independent times. Error bars represent one standard deviation from the mean. Error bars that are not visible are smaller than the symbol. Vector = pCV1 {VanDrisse, 2016 #25211}.

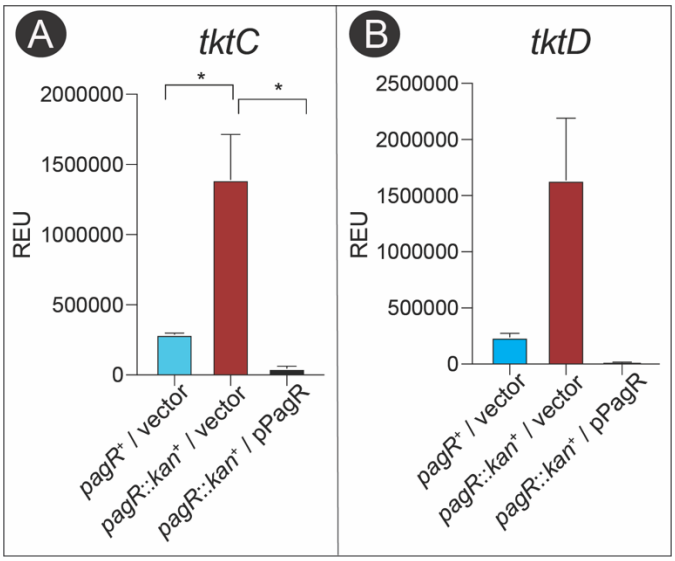

Figure A.4. qRT-PCR analysis of *tktC* and *tktD* expression as a functuion of PagR. **A.** Effect of PagR on the expression of *tktC* as shown by qRT-PCR. **B.** Control of *tktD* expression by PagR. Cultures were grown LB. Details of the qRT-PCR conditions used can be found in the *Experimental procedures* section. Vector = pCV.

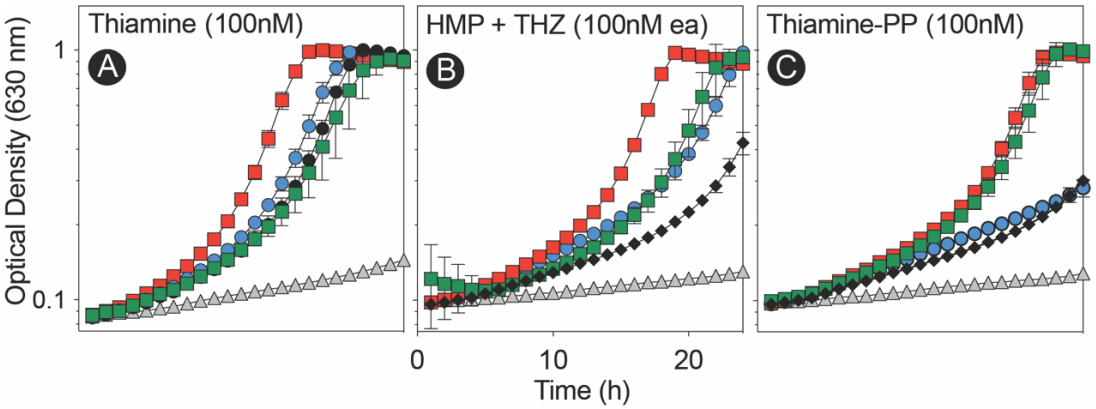

Figure A.5. Neither thiamine, thiamine precursors nor thiamine pyrophosphate restored growth of a D*pagR* strain during growth with succinate. A. succinate + thiamine (100 nM), **B.** succinate + HMP + THZ (100 nM each), **C.** succinate + thiamine-PP (100 nM). In each plot of each panel strains are identified as follows: *pagR*$^{+}$ / vector (red squares); Δ*pagR* / vector (grey triangles); D*pagR* / pPagR (blue circles); Δ*pagR* Δ*tktCD* / vector (black diamonds); Δ*pagR* Δ*tktCD* / pTktCD (green squares). The expression of genes carried by plasmids was induced with arabinose (100 μM). Growth conditions are detailed under *Experimental procedures.*

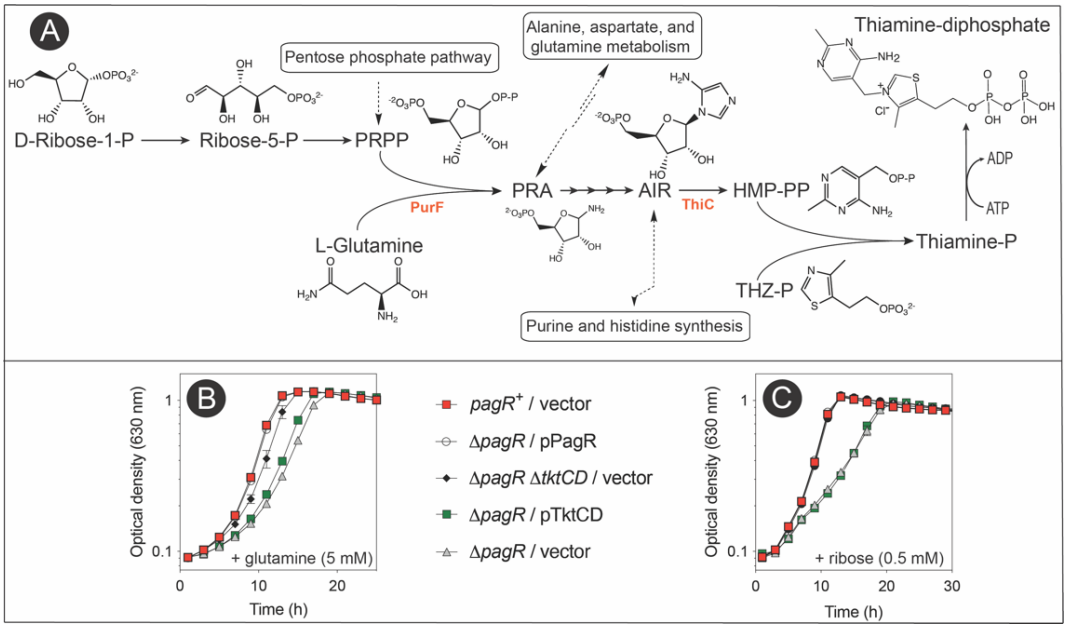

Figure A.6. Metabolites needed for inosine biosynthesis stimulate growth of a Δ*pagR* strain with succinate as the sole source of carbon and energy. A. Metabolic connections relevant to the growth behavior of a Δ*pagR* strain growing with succinate. **B.** Stimulatory effect of glutamine on the growth kinetics of Δ*pagR* strains relative to that of the *pagR*$^{+}$ strain. **C.** Stimulatory effect of a low concentration of ribose (0.5 mM) on the growth kinetics of Δ*pagR* strains relative to that of the *pagR*$^{+}$ strain. Experiments shown in panels B and C are representative of biological triplicates performed in technical triplicates. Error bars represent one standard deviation of the mean. Error bars that are not visible are covered by the size of the symbol. Vector = pCV1 {VanDrisse, 2016 #25211}.

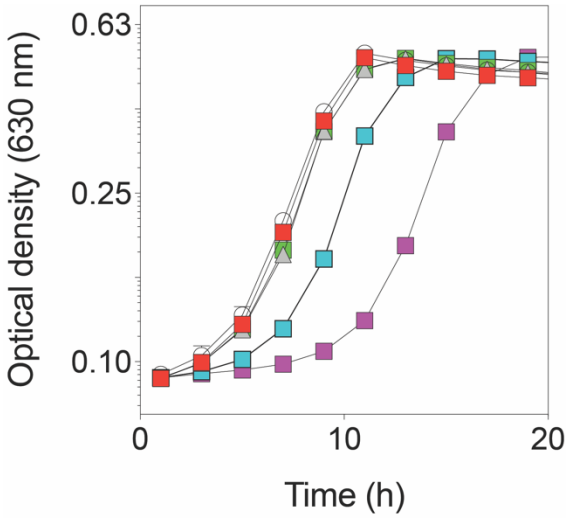

Figure A.7. Overexpression of the *tktCD* genes has a deleterious effect on growth with pyruvate. All cultures were grown in minimal medium supplemented with sodium pyruvate (20 mM). Other components of the medium are found in the *Experimental procedures* section. The strains used in this experiment were: JE22070 (*pagR*$^{+}$ / vector, red squares), JE21566 (Δ*pagR* / vector, grey triangles), (JE21577 (Δ*pagR* / pPagR, open circles), JE25520 (Δ*pagR* / pTktCD + 100 μM arabinose, green squares), JE25520(Δ*pagR* / pTktCD + 500 mM arabinose, teal squares), JE25520 (Δ*pagR* / pTktCD + 1000 μM arabinose, magenta squares). This experiment was performed in biological and technical triplicates. The error bars represent one standard deviation from the mean. Some error bars are not visible because they are smaller than the symbol. Vector = pCV1 {VanDrisse, 2016 #25211}.

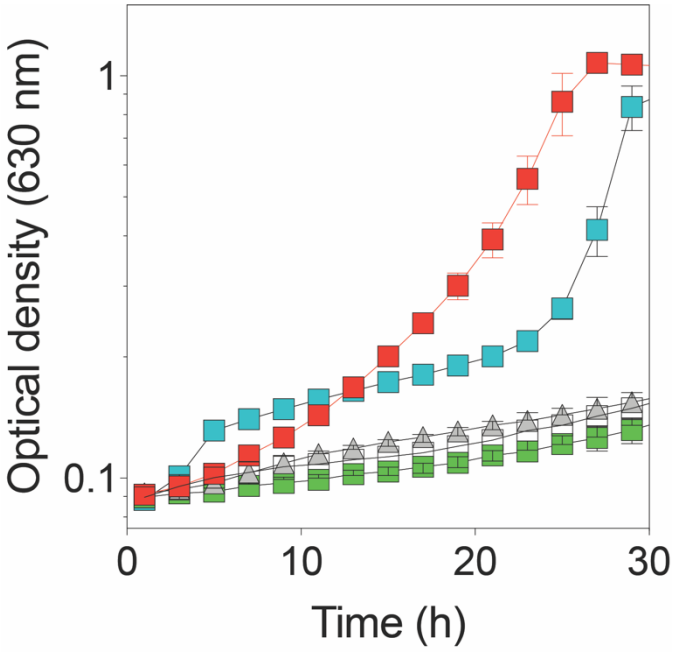

Figure A.8. Positive effect of the overexpression of the SucA component of the a-ketglutarate dehydrogenase complex (a TPP-dependent enzyme) on the growth of a Δ*pagR* strain with succinate. Strains used in this experiment were grown in minimal medium supplemented with succinate (30 mM). The strains used were: JE22070(*pagR*$^{+}$ / vector), red squares), JE21566(D*pagR* / vector), grey triangles), JE25981 (Δ*pagR* / pSucA, no inducer, green squares), JE25981 (Δ*pagR* / pSucA, 100 μM arabinose, open squares), JE25981 (Δ*pagR* / pSucA, 500 μM arabinose, teal squares). This experiment was performed in biological and technical replicates. Error bars represent one standard deviation from the mean. Some error bars are not visible because they are smaller than the symbol. Vector = pCV1{VanDrisse, 2016 #25211}.

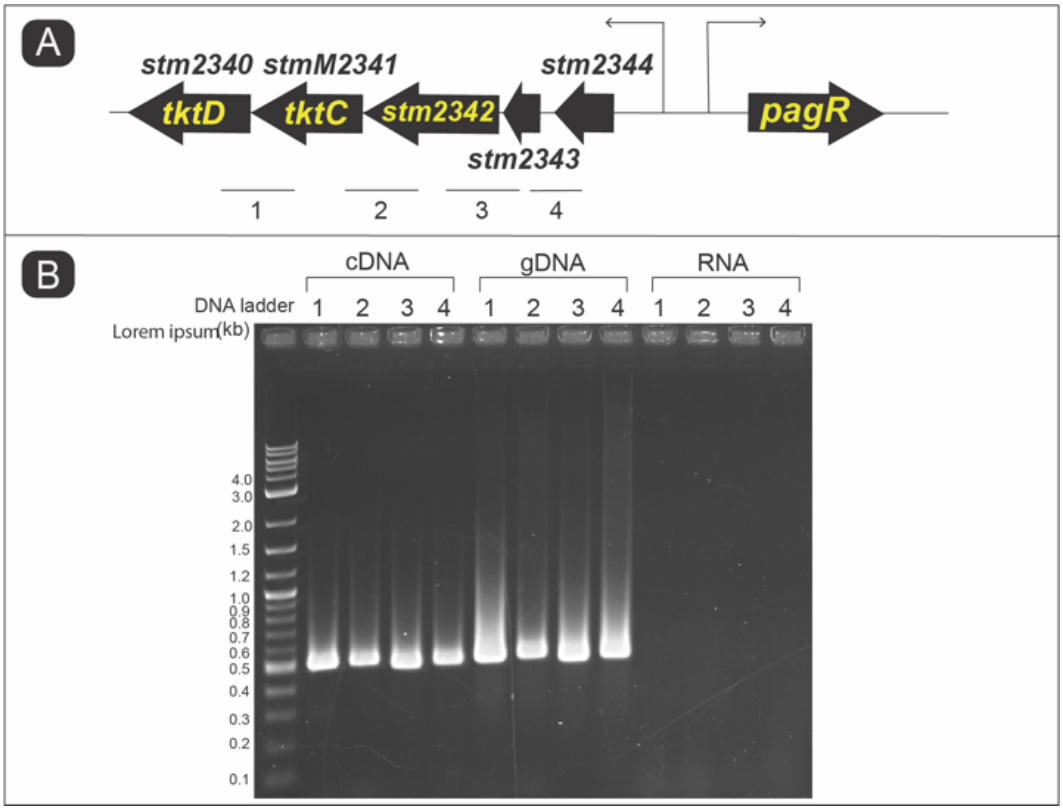

Figure A.10. Genes encoding TktCD are promoter-distal in a 5-gene operon. A. The *pagR* gene is divergently transcribed from the operon containing the genes encoding the two subunits of TktCD. The first three genes of the operon are annotated as putative PTS-type transporters. **B.** Operon PCR of the putative polycistronic operon containg the transketolase genes *tktC* and *tktD*. RNA, cDNA and gDNA were PCR amplified using primers specific to overlapping regions between genes of the putative operon designated 1-4 in panel A. Positive controls are amplification of gDNA; negative controls are no amplification of RNA samples; experimental samples were the cDNA amplifications. Ladder on left shows corresponding amplficiation size (2-Log DNA Ladder, NEB).

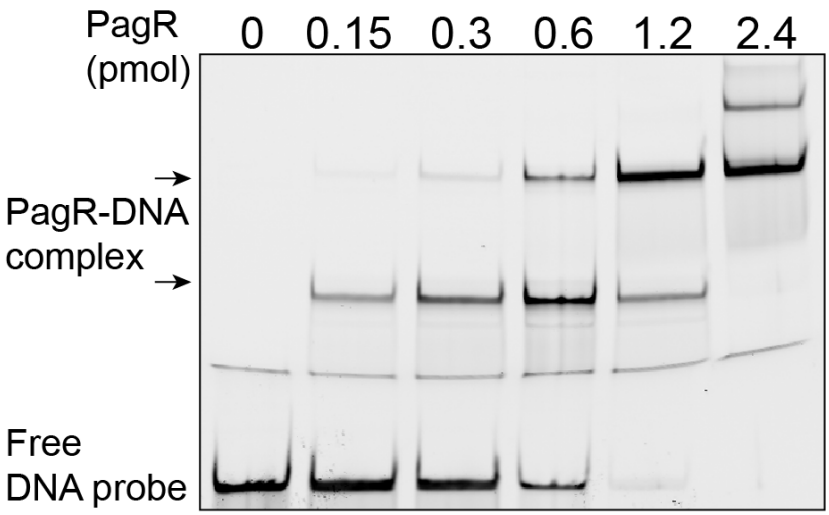

Figure A.11. PagR binds the region between *pagR* and the five-gene operon that includes *tktCD.* The mobility of the DNA fragment [include the exact limits of the probe you used, e.g., between 2,456,686 nt (the start of stm2344) and 2,457,382 nt (the start of stm2345)] that included the putative promoter region decreased as a function of increasing concentrations of PagR protein as compared to the no-PagR-protein-added control. This suggested that the oligomerization state of PagR changed as its concentration increased and/or that multiple PagR-binding sites were present in the DNA fragment. The possibility of multiple PagR-binding sites on the DNA fragment was likely due to the fact that this fragment of DNA also included the putative promoter region for the *pagR* gene. To better understand PagR binding to the DNA and the regulation of the *stm2340-stm2344* genes *versus* the *pagR* gene**.**

APPENDIX B

PAGR ACTS AS MASTER REGULATOR OF IRON ACQUISITION PATHWAY IN SALMONELLA ENTERICA[8]

[8]Parks A.R., Lewis Z. A., and Escalante-Semerena J.C. To be submitted.

INTRODUCTION

These data are a preliminary taste of six years of trying to get the ChIP-seq experiment to work in order to elucidate in vivo targets for the transcriptional regulator, PagR. Dr. Zack Lewis performed all of the data analysis for this manuscript and I would like to thank him for all of his time and efforts.

This ChIP-seq dataset suggests that PagR regulates genes involved in iron acquisition in *S. enterica*. A select few of the promoter targets identified were tested further by electrophoretic mobility shift assays (EMSAs) and β-galactosidase fusions of their promoters. These results are preliminary, but I hope you can get a feeling for the manuscript to come from these data.

METHODS

Bacterial strains and growth conditions. All strains were derivatives of Salmonella enterica subsp. Enterica serovar Typhimurium LT2 (hereafter *S. enterica*) using the Wanner and Datsenko protocol that uses l Red recombinase to engineer in-frame deletion. All bacterial strains and plasmids used in this study are listed in Table SX and SX. Growth of *S. enterica* for experiments using N-medium was conducted as follows: bacterial cultures were grown from a single colony in nutrient broth (NB [Difco]) over 12 hours shaking at 180 rpm at 37C. In the after 12 hours, cells were subcultured1:100 into N-medium containing 0.1% casamino acids, 38 mM glycerol, 10uM $MgCl_2$, 1uM KH_2PO_4, 7.5mM $(NH_4)_2SO_4$, 100mM Tris-HCl pH 7.4, 5mM KCl, 0.5mM K_2SO_4, 15nM cyanocobalamin, and the designated amount of L-(+)-arabinose as cited in figure legends. When used,

ampicillin was added at 100ug/ml, chloramphenicol at 20ug/ml, and kanamycin at 50ug/ml. Cells were grown shaking at 180 rpm at 37C to optical density 600nm of 0.5.

Strain construction. All strains constructed were *metE ara-9* derivatives of *S. enterica*. Primers used in this study were synthesized by Integrated DNA Technologies, Inc. (IDT, Coralville, IA) and are listed in the Supplemental Material. The *pagR*::*kan*$^{+}$ was engineered using the Wanner and Datsenko (2000) protocol as follows: PFU Ultra II Fusion DNA polymerase (Stratagene) was using to amplify flanking regions of plasmid pkD4 (*kan*$^{+}$ marker) using primers containing 36-base pair overhangs corresponding to the beginning or the end of the *pagR* gene. Polymerase chain reaction (PCR) amplifications were visualized by post-staining with ethidium bromide (0.5 μg /ml) for 10 min. Correctly amplified PCR products were cleaned with the Wizard SV gel and PCR clean up kit (Promega). *S. enterica* strain JE6692 (*metE* 205 ara-9 / pKD46 *bla*$^{+}$) was grown to OD600 nm of 0.5 at 30C to maintain the temperature sensitive plasmid pKD46 and was transformed with ~200ng of PCR product using a 0.2-cm electroporation cuvette (MidiSci) and a Bio-Rad Laboratories MicroPulser electroporator on Ec2 setting. Transformed cells were recovered by incubation at 37C with shaking and played on lysogeny broth (LB) agar supplemented with 25ug/ml of kanamycin. Resulting colonies were PCR verified for correct antibiotic insertions then moved into strain JE6583 by P22-mediated transduction as described elsewhere (Roth). Transduction colonies were streaked to isolation and select colonies were tested for phage sensitivity and

Sanger sequenced to verify correct insertion location and sequence using primers flanking gene of interest.

Plasmid construction All plasmids used in this study are listed in Table SX. Primers used in this study were synthesized by Integrated DNA Technologies, Inc. (IDT; Coralville, IA), and are listed in Table SX. The plasmid pPAGR-14 containing PAGR with a N-terminally fused hexahis-tag was designed using the NEBuilder HiFi DNA assembly.

The high-efficiency cloning method of BspQI digestion (REF) was used to clone the assembled product into the arabinose-inducible complementation vector pCV1.

Chromatin immunoprecipitation. Strain JE24646 (*pagR*::*kan*$^{+}$ / pHis$_6$-PagR) was grown in biological triplicate overnight in nutrient broth (2 ml; NB, Difco) with shaking at 37C. After incubation, cells were diluted 1:100 into 50ml of N-medium (as described above) supplemented with 15nM cyanocobalamin and 100ug/ml of ampicillin in a 250ml flask. Cells were grown shaking at 180 rpm at 37C until optical density 600 nm of 0.6 was reached. Plasmids encoding his-tagged PagR were then induced with the addition L-(+)-arabinose (0.1%, [6.6mM]) and cells continued to grow for 30 more minutes shaking at 180 rpm at 37C. After plasmid induction, cells were crosslinked with 1% formaldehyde (from Pierce^{TM} 16% Formaldehyde (w/v), methanol free) for 20 minutes shaking at 100 rpm on a table shaker. Excess formaldehyde was quenched with the addition of 0.5M glycine, crosslinked cells were transferred to 50mL conical tubes, and cells were centrifuged at 4,000 x g in a centrifuge for 10 minutes. The supernatant was discarded and cells were washed

with 1ml of 1X PBS (phosphate buffered saline) pH 7.4. Washed cells were split in half, 500ul was snap frozen at -80C. The remaining cells were centrifuged at 6,000 x g for 5 minutes in 1.7ml ependorf tubes and were resuspended in 1ml of ChIP lysis buffer (20mM HEPES, pH 7.9; 50mM KCl; 0.5mM DTT; 500mM NaCl; 10mM imidazole; 1% BSA; 1/8th of a cOmplete^{TM} mini EDTA-free protease inhibitor cocktail tablet; 400uM phenylmethylsulfonyl fluoride [PMSF]). Cells were lysed with the addition of lysozyme (final concentration of 1mg/ml) and incubated on ice for 30 minutes. Lysed cells were transferred to 15ml conical tubes and DNA was sheared four rounds of sonication of 15 minutes, 30 seconds on 30 seconds off using a water bath Diagenode Bioruptor. While DNA was shearing, 100ul per sample of MagneHisTM Ni-Particle beads were pre-equilibrated with 500ul of ChIP lysis buffer by incubation rotating end over end for 30 minutes at room temperature. After sonication, 40ul of sheared cell lysate was removed and frozen as the “input” samples, while the remaining cell lysate was added to 100ul of pre-equilibrated MagneHis^{TM} Ni-Particle beads and were incubated together with end-over-end mixing for 30 minutes at room temperature.

To isolate PagR-bound DNA fragments, Ni-particles with His-PagR bound were washed five times with 500ul of wash buffer (100mM HEPES pH 7.5; 10mM imidazole; 500mM NaCl; 1% BSA; 1/8th cOmplete^{TM} mini EDTA-free protease inhibitor cocktail tablet), using the MagneRack magnetic bead separator. To elute the His-tagged PagR, 100ul of elution buffer (100mM HEPES pH 7.5; 500mM imidazole) was added to the beads and incubated together stationary for 30 minutes at room temperature with occasional mixing. Eluted DNA-protein

complexes were separated from the beads and added to fresh PCR tubes. The input samples were thawed and brought to the same volume as the ChIP samples through the addition of MQ water. SDS (final concentration of 1.25%) was added and samples were decrosslinked overnight at 65C in a heated-lid PCR machine. After decrosslinking, 8ul of Rnase A (10mg/ml) was added and samples were incubated for two hours at 50C. Proteinase K (4ul at 20mg/ml) was added and samples were incubated for 2 hours at 50C. Samples were cleaned of protein using the Qiagen MinElute Reaction Cleanup kit (cat# 28204). Final input and ChIP DNA samples were quantified using the Qubit dsDNA HS DNA Assay kit on a Qubit 4 fluormeter.

ChIP-seq library construction and sequencing. 10ng of DNA was used to construct each chip sequencing library starting with end repair of DNA fragments using the NEB Ultra II End Repair Module (cat# E7546S) according to manufacturers instructions. Adaptor ligation of the iTruSeq Illumina adaptor (2.5ul of 0.15uM stock) to DNA fragments was performed using the NEBNext Ultra II Ligation Module (cat# E7595) according to manufacturer's instructions. Self-ligated adaptors were removed from ligated libraries through magnetic binding and washing of DNA using Sera-Mag SpeedBeads (cat# 65152105050250) at a ratio of 1:1 DNA/beads volume. Cleaned adaptor-ligated DNA fragments were amplified using NEB Ultra II Q5 hotstart polymerase (cat# M0543S) and a unique dual index combination of i7/i5 index primer mix. The PCR condition was according to manufacturer except with 11 cycles with annealing temperature of 55C and final extension of 3 minutes at 72C. PCR amplified chip libraries were cleaned again

using the Sera-Mag SpeedBeads and final quantification was conducted using the Qubit dsDNA HS DNA Assay kit on a Qubit 4 fluormeter.

ChIP-seq data analysis

Indexed libraries were diluted to 5ng/ul and pooled equally to a volume of at least 20ul and were sequenced were sequenced on an Illumina NextSeq2000 instrument at the Georgia Genomics and Bioinformatics Core. Resulting FastQ files were processed to remove low quality reads and adaptor sequences using TrimGalore! (https://github.com/FelixKrueger/TrimGalore). Trimmed reads were then mapped to the *Salmonella enterica* subsp. *enterica* serovar Typhimurium LT2 genome (Genbank accession#: GCF_000006945.2) using BWA [1]. Output files were sorted and converted to bam format using SAMtools [2]. Bigwig files were generated using the deeptools bamCoverage command (--binSize 25 --smoothLength 50) [3] and visualized using the Integrative Genomics Viewer [4]. Peaks were called for each replicate using MACS2 [5] (-q =0.001). Peak files generated by MAC2 were filtered using awk to remove peaks with enrichment values < 2.0, and a list of consensus peak files from all three replicates was generated using MSPC [6].

β-Galactosidase assays β-galactosidase assays were conducted as described previously [7]. Three independent cultures were grown overnight in LB (2 ml; LB, Difco) containing 100 μg /ml of ampicillin. In the morning, cultures were subcultured 1:100 into fresh LB plus ampicillin (100μg/ml). Complementation

plasmids were induced with 1mM of L-(+)-arabinose unless otherwise stated. Cells were grown shaking at 180 rpm at 37C until an OD600 nm of 0.4-0.6.

Protein purification PagR protein was purified to homogeneity from plasmid pPagR-8 encoding a PagR protein with a maltose binding protein-hexahistidine (MBP-H6) N-terminal tag. The tag was removed after incubation with recombinant tobacco etch virus (rTEV) protease since the plasmid used to produce MBP-H_6-PagR (pTEV19) contained a rTEV protease cleavage site [11]. A sample (10 ml) of an overnight culture of *E. coli* C41 (DE3) / pPagR-8 strain was used to inoculate one liter of lysogeny broth (LB) containing 100 mg/ml of ampicillin. Cells were grown shaking at 125 rpm at 37 °C until the culture reached an optical density (OD) at 600 nm of 0.7. Expression of genes of interest encoded by the plasmids was induced by the addition of IPTG (0.25 mM) followed by ~12 h of overnight incubation at 37 °C. The next morning cells were harvested by centrifugation at 6,000 x g for 15 min using a refrigerated Beckman-Coulter Avanti J-20-XPI centrifuge equipped with a JLA 8.1 rotor. Cell pellets were resuspended in 20 ml of buffer A [(4-(2-hydroxyethyl-1-piperazineethanesulfonic acid (HEPES) buffer (50 mM, pH 7.5 at 4 ºC) containing NaCl (0.5 M), glycerol (20% v/v), and imidazole (20 mM)] and were sonicated thrice for 30-s intervals, and during each interval, sonication was on for 2 s and off for 2 s, at 60% amplitude. The resulting whole-cell lysates were centrifuged for 30 min at 40,000 x g and the supernatants were filtered with a 0.45-m filter to remove large particulates. Each filtered lysate was applied onto a 1-ml nitrilotriacetic acid (NTA) affinity chromatography column pre-

equilibrated with buffer A. Fractions were collected by gravity at 4 °C. The purification was performed as follows: After all the lysate was applied, the column was washed with 10 column volumes (CV, *i.e.,* 10 mL) of buffer A, 7 CV (*i.e.,* 7 mL) of buffer A containing 4% elution buffer B [HEPES buffer (50 mM, pH 7.5 at 4 ºC) containing NaCl (0.5 M), glycerol (20% v/v), and imidazole (0.5M)], and finally, MBP-H_6-PagR was eluted in two fractions, first with 1 CV (*i.e.,* 1mL) of 100% elution buffer B, and the second fraction being 4 CV (*i.e.,* 4mL) of 100% elution buffer B. Both elution fractions were pooled and MBP-H_6-PagR was cleaved with rTEV protease at a concentration of 1 mg rTEV:100 mg of PagR protein while dialyzing at room temperature in HEPES buffer (50 mM, pH 7.5 at 4 ºC) containing NaCl (0.5M), glycerol (10% v/v), and dithiothreitol (DTT, 1 mM). Cleaved protein was dialyzed twice more at 4 °C in buffer A. Cleaved PAGR protein was run back over a 2-ml NTA column to remove MBP-His and rTEV protease. Tag-less PAGR protein did not interact with the NTA resin, and was collected in the flow-through fraction. To further remove MBP from PagR protein, flow-through fractions from the second NTA purification were run over a 1-mL amylose resin to remove contaminating MBP. Pure PagR was dialyzed overnight HEPES buffer (50 mM, pH7.5 at 4ºC) containing NaCl (150 mM) and glycerol (20% v/v) at 4˚C. Dialyzed PagR was stored at -80˚C in 15-ml aliquots that were flash frozen with liquid nitrogen. Protein concentration was determined by Nanodrop using the extinction coefficient and molecular mass of PagR.

Electrophoretic mobility shift assay Electrophoretic mobility shift assays (EMSAs) were performed to quantify PagR binding to DNA. EMSAs were performed as follows: PagR protein purified to >95% homogeneity was incubated at the concentrations in pmol of protein labeled in the figures with a 5'-6-FAM-labeled DNA probe at 50 ng [*i.e.,* 0.3 pmol] was added). EMSA buffer [HEPES buffer (50 mM, pH 7.5 at 4ºC) containing NaCl (150 mM), and glycerol (10% v/v)] was added to the reaction mixture (total volume = 25 μl) and DNA and protein were incubated at room temperature for 40 min. During incubation, a 7.5% Tris-boric acid-EDTA (TBE) polyacrylamide gel was pre-developed at 100 V for 40 min in 0.5X TBE buffer at 4˚C. After incubation, 5 μl of glycerol (50% v/v) was added to the reaction mixtures, and 20 μl of each reaction mixture was resolved by the polyacrylamide gel. A lane of xylene cyanol and bromophenol blue dye was added as a tracking indicator, and the gel was run until bromophenol blue reached the bottom of the gel. The gel was imaged using a Typhoon Trio imager at 525 nm with the 488 (Blue) filter.

REFERENCES

1. Li, H. and R. Durbin, *Fast and accurate short read alignment with Burrows-Wheeler transform.* Bioinformatics, 2009. **25**(14): p. 1754-60.
2. Danecek, P., et al., *Twelve years of SAMtools and BCFtools.* Gigascience, 2021. **10**(2).

3. Ramirez, F., et al., *deepTools2: a next generation web server for deep-sequencing data analysis.* Nucleic Acids Res, 2016. **44**(W1): p. W160-5.

4. Robinson, J.T., et al., *Integrative genomics viewer.* Nat Biotechnol, 2011. **29**(1): p. 24-6.

5. Zhang, Y., et al., *Model-based analysis of ChIP-Seq (MACS).* Genome Biol, 2008. **9**(9): p. R137.

6. Jalili, V., et al., *Using combined evidence from replicates to evaluate ChIP-seq peaks.* Bioinformatics, 2018. **34**(13): p. 2338.

7. Miller, F.D. and C.L. Hershberger, *A quantitative beta-galactosidase alpha-complementation assay for fusion proteins containing human insulin B-chain peptides.* Gene, 1984. **29**(1-2): p. 247-50.

8. Stead, M.B., et al., *RNAsnap: a rapid, quantitative and inexpensive, method for isolating total RNA from bacteria.* Nucleic Acids Res., 2012. **40**: p. e156.

9. Rocha, D.J., C.S. Santos, and L.G. Pacheco, *Bacterial reference genes for gene expression studies by RT-qPCR: survey and analysis.* Antonie Van Leeuwenhoek, 2015. **108**: p. 685-693.

10. Livak, K.J. and T.D. Schmittgen, *Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method.* Methods, 2001. **25**: p. 402-408.

11. VanDrisse, C.M. and J.C. Escalante-Semerena, *New high-cloning-efficiency vectors for complementation studies and recombinant protein overproduction in Escherichia coli and Salmonella enterica.* Plasmid, 2016. **86**: p. 1-6.

Table B.1. N-medium low phosphate low Mg2+ results

start	end	length	abs_summit	fold_enrichment	#NAME?	gene
3379672	3380457	786	3380068	13.6756	3390.51	yqjH-yqjI
4216872	4217753	882	4217233	7.00537	933.89	glnA-typA
655831	656384	554	656119	6.20667	826.413	fepB-entC
3714911	3715796	886	3715517	6.20638	790.848	yhhx-yhhY
643727	644169	443	643937	4.82504	462.181	fepA-fes
2464965	2466799	1835	2465435	4.55189	409.129	hisJ-argT-ubiX
223497	224014	518	223715	4.46459	403.533	fhuA
896312	897663	1352	897153	4.27136	413.802	glnH
2984762	2985293	532	2985002	3.99599	327.027	ygaA-STM2840
653420	653848	429	653615	3.753	270.05	fepD
2298443	2299159	717	2298745	3.53649	242.853	cirA
1673527	1674038	512	1673745	3.52931	234.639	STM1586_yncD
4808777	4809240	464	4808991	3.49347	230.312	fhuF
3597316	3598490	1175	3598172	3.34801	220.644	STM3444
1945538	1946182	645	1945957	3.33048	204.629	STM1849
836899	837357	459	837166	3.27407	197.63	gmpA
2925628	2926095	468	2925881	3.26787	226.219	iroN
1289910	1290548	639	1290136	3.25567	211.697	fhuE
4215083	4215718	636	4215366	3.21331	187.616	glnL and glnA_typA
411242	411773	532	411414	3.12099	174.587	STM0363
539528	540147	620	539910	2.99672	157.567	yabN
1350113	1350623	511	1350348	2.87165	151.085	yeaS
4145836	4146745	910	4146537	2.86535	140.276	hemC-cyaA
3183861	3184451	591	3184180	2.85044	157.552	rcnR-yohM
174666	175500	835	175222	2.79107	144.313	aroP-pdhR
139511	140361	851	140046	2.74514	126.218	yabB
310228	311318	1091	311017	2.72062	156.844	clpV T6SS stm0271-stm0272

2472320	2473203	884	2472789	2.71248	133.17	dokD-STM2361
3120259	3120864	606	3120501	2.69579	143.057	sdaC
1877222	1878008	787	1877768	2.63456	111.719	prsA
2362298	2362639	342	2362481	2.62391	110.458	yojI/STM2263
2521951	2523506	1556	2522247	2.61325	109.204	STM2405-STM2406
2707559	2708231	673	2707886	2.61325	109.204	yfhK
4266282	4266656	375	4266449	2.60514	119.233	yiiZ-sodA
515936	516477	542	516186	2.59905	107.539	cdsH-ybaO
2546055	2547004	950	2546798	2.58841	111.85	ccr
3874705	3875418	714	3875114	2.58029	125.879	yibF-mtlA
4493067	4493853	787	4493445	2.53788	121.082	
253704	254602	899	253942	2.53514	100.169	map-rpsB
37134	39931	2798	39582	2.50002	232.037	
729396	730173	778	729961	2.46531	94.0287	
3684568	3684994	427	3684781	2.46413	92.212	
4703355	4703752	398	4703532	2.45945	97.574	
2515348	2515984	637	2515755	2.44206	91.1271	
2040255	2040597	343	2040393	2.39196	95.0943	
3135292	3136048	757	3135842	2.38163	87.9253	
1756161	1756553	393	1756373	2.37448	88.0061	
3780380	3780910	531	3780620	2.36108	90.414	
2346537	2346844	308	2346674	2.35846	93.9864	
3251341	3251859	519	3251624	2.33138	88.1954	
466194	467121	928	466395	2.31855	76.7099	
1935601	1936867	1267	1936182	2.31242	87.2832	
4096696	4097204	509	4097014	2.31081	77.6903	
1425041	1425486	446	1425259	2.3008	74.8944	
1381688	1382273	586	1382054	2.28593	79.4613	
2731431	2731725	295	2731569	2.27712	99.426	
2447451	2448323	873	2447676	2.27281	78.0316	
889078	889446	369	889235	2.26174	70.9566	

2524834	2526002	1169	2525028	2.25623	81.0206	
2597268	2597534	267	2597400	2.22429	79.9142	
1649024	1649660	637	1649466	2.21203	66.0616	
2951521	2951986	466	2951840	2.21203	66.0616	
782063	782717	655	782554	2.19154	78.4369	
4258655	4259326	672	4258915	2.19073	64.0045	
4642608	4643626	1019	4643196	2.19073	64.0045	
1052815	1054000	1186	1053821	2.14859	72.0514	
2082132	2082667	536	2082515	2.14707	60.5973	
3149807	3150203	397	3150045	2.13747	58.9785	
978680	979095	416	978889	2.13692	70.643	
450434	451172	739	450774	2.11892	61.3322	
3457486	3459082	1597	3458893	2.10511	63.6028	
1205704	1206295	592	1205884	2.10211	61.9652	
695641	696211	571	695926	2.09906	58.6407	
1892989	1893824	836	1893446	2.08651	62.8024	
2021437	2023157	1721	2022681	2.08644	62.3324	
3736338	3737111	774	3736898	2.08421	54.1236	
3527655	3529362	1708	3528652	2.07353	59.5594	
4855872	4857272	1401	4856607	2.07137	61.7774	
4815454	4816166	713	4815691	2.06812	54.2793	
3295668	3296094	427	3295867	2.06665	57.8038	
1285808	1286474	667	1286176	2.05226	51.2915	
4053710	4054194	485	4053996	2.05155	55.6003	
4334948	4335469	522	4335155	2.04074	55.2163	
4204911	4206394	1484	4205954	2.03233	61.0704	
4140851	4141882	1032	4141549	2.0274	49.1301	
4241456	4242169	714	4241752	2.02619	57.6671	
526024	527586	1563	526282	2.0248	58.7716	
354973	355520	548	355154	2.02414	50.904	
3587894	3591969	4076	3588154	2.02385	48.824	

1372031	1372841	811	1372502	2.00965	47.6091	
260336	260666	331	260490	2.00255	47.0066	
3585506	3587436	1931	3586860	2.00057	49.3029	

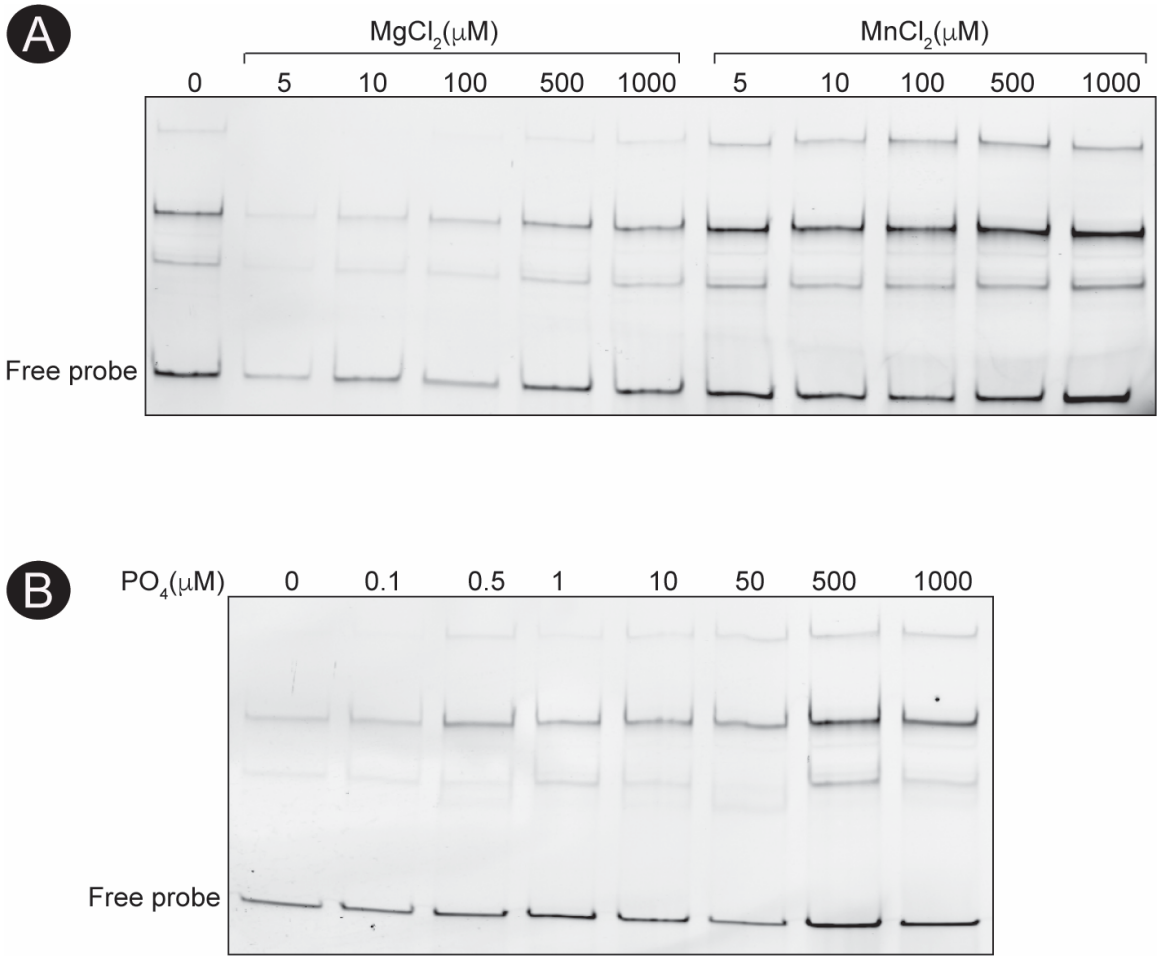

Figure B.1. Metals affect PagR DNA binding to the *pagR* promoter. Each lane contains PagR protein at 4pmol and P*pagR* promoter DNA of 0.3pmol.

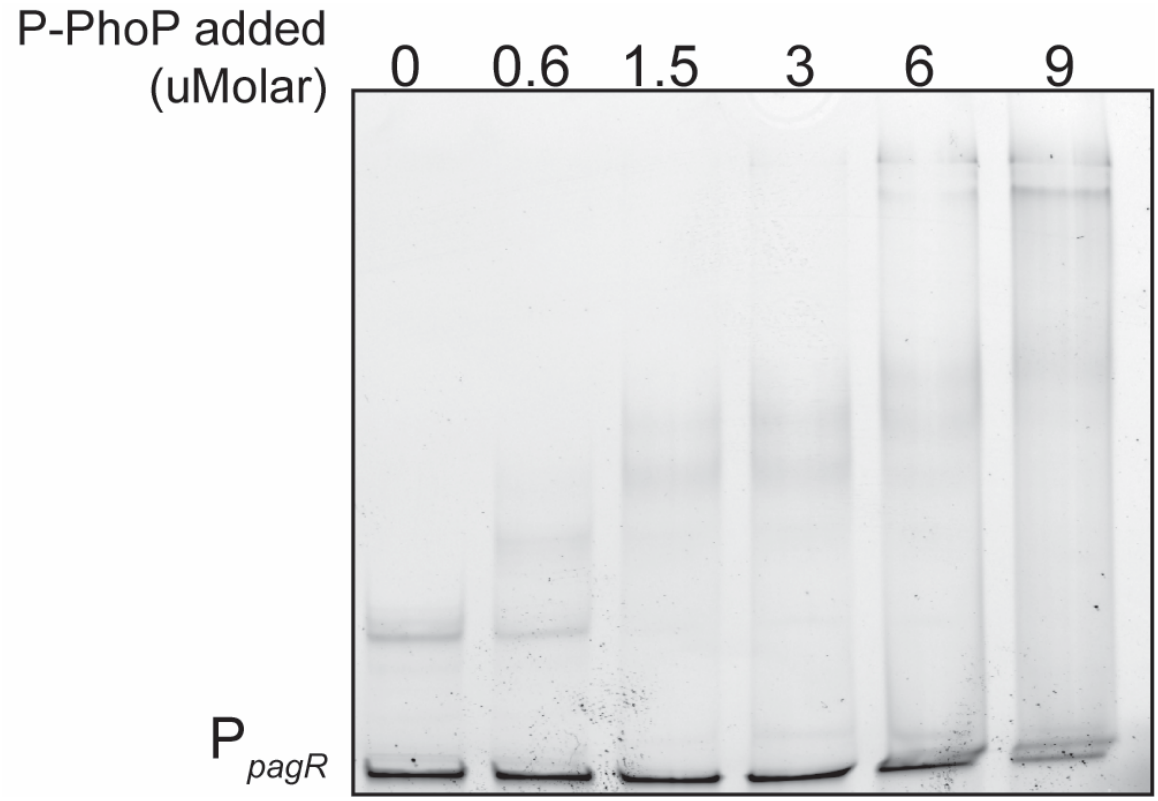

Figure B.2. Phosphorylated PhoP binds to the P_{pagR} promoter. PhoP-Hix6 was purified to homogeneity, phosphorylated with a homogeneously purified PhoQ soluble truncation, and tested for the ability to bind to P_{pagR} by EMSA.

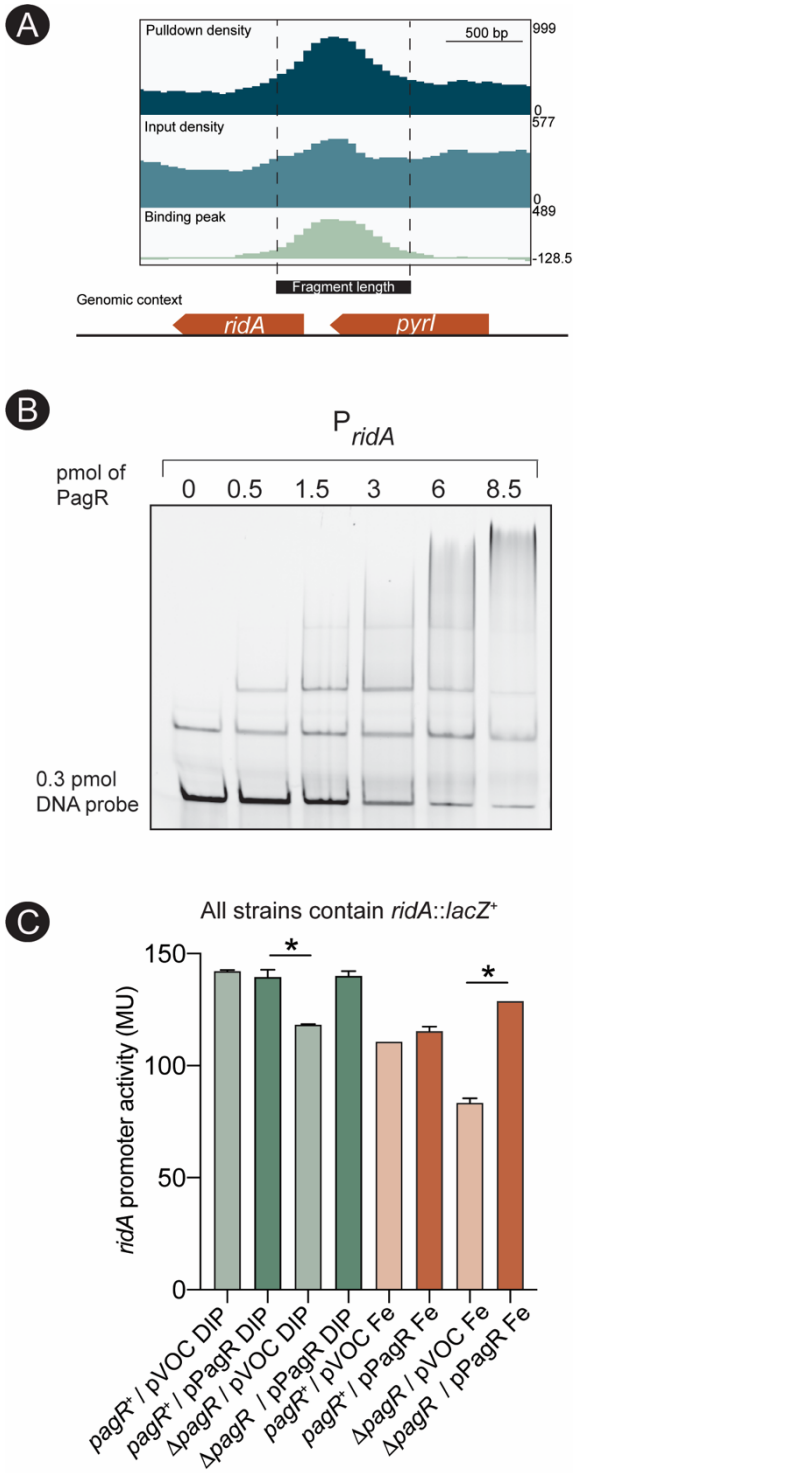

Figure B.3. PagR regulates the transcription of the 2-aminoacrylate deaminase RidA in *S. enterica.* A) Peak density from N-medium ChIP-seq. B) EMSA of PagR binding to P_{ridA}. C) β-galactosidase assay of the effect of PagR on P_{ridA}.

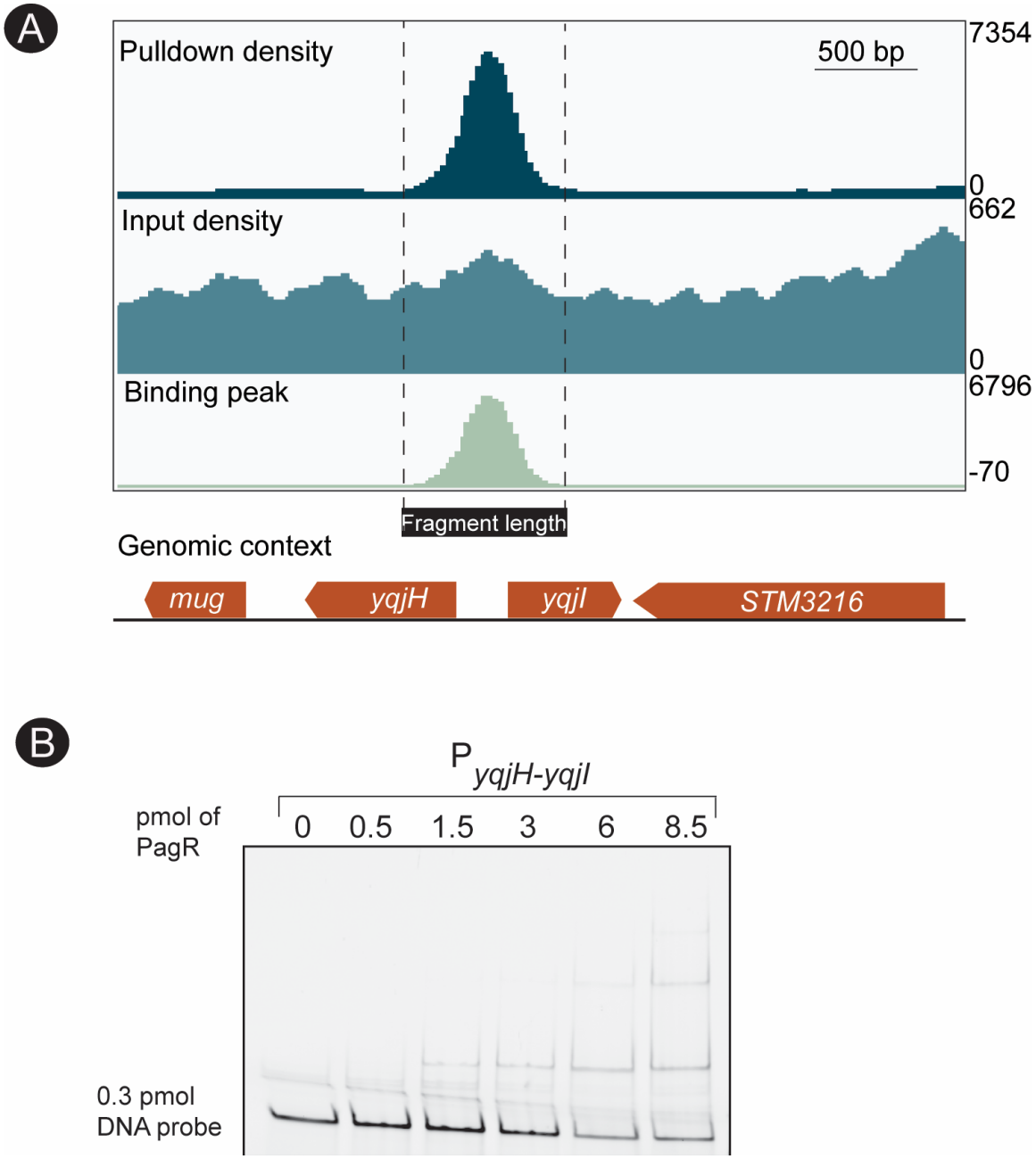

Figure B.4. PagR binds to the *yqjH-yqjI* promoter region, genes that encode for a ferric siderophore reductase (*yqjH*) and its transcriptional regulator (*yqjI*).

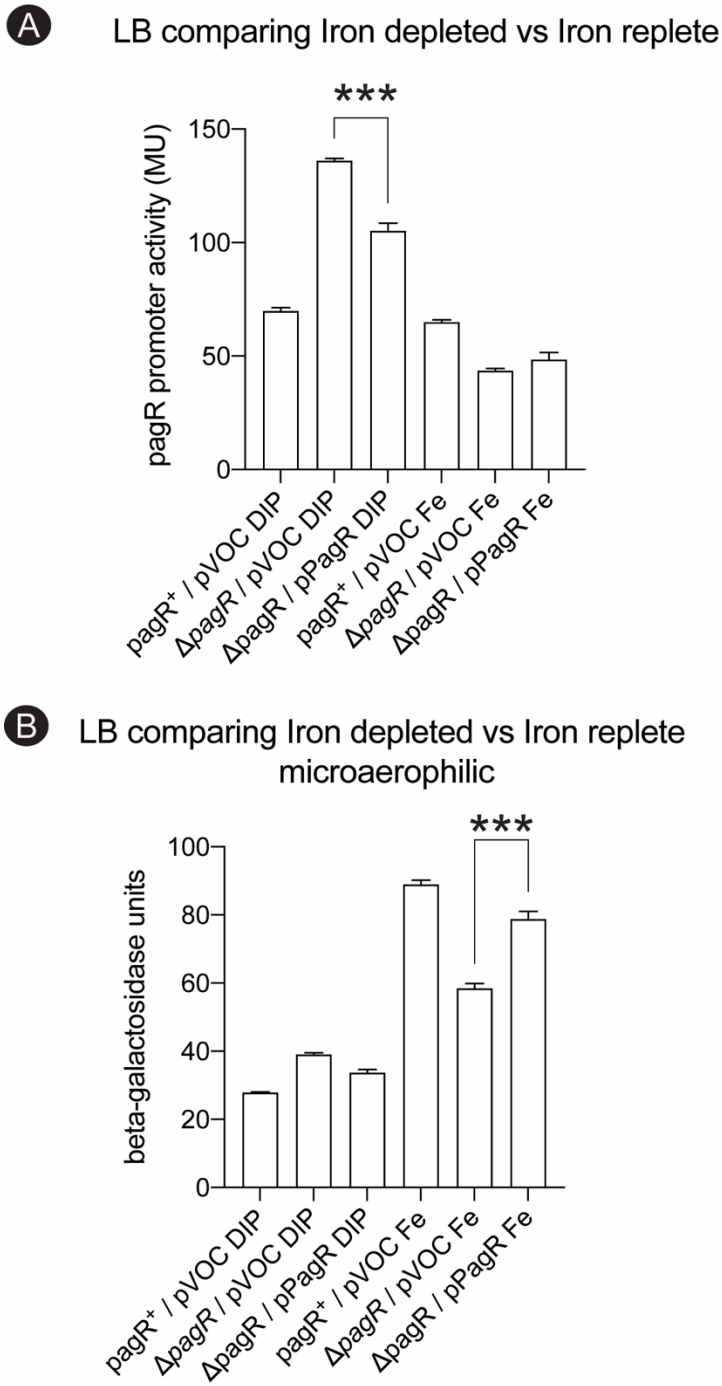

Figure B.5. PagR regulates its own promoter in an Iron and oxygen dependent manner. DIP stands for 2,2-dipyridyl, added at 200μM to be an iron chelator in order to mimic iron deplete conditions. Fe stands for the addition of 100μM $FeCl_3$ in order to mimic an iron replete condition.

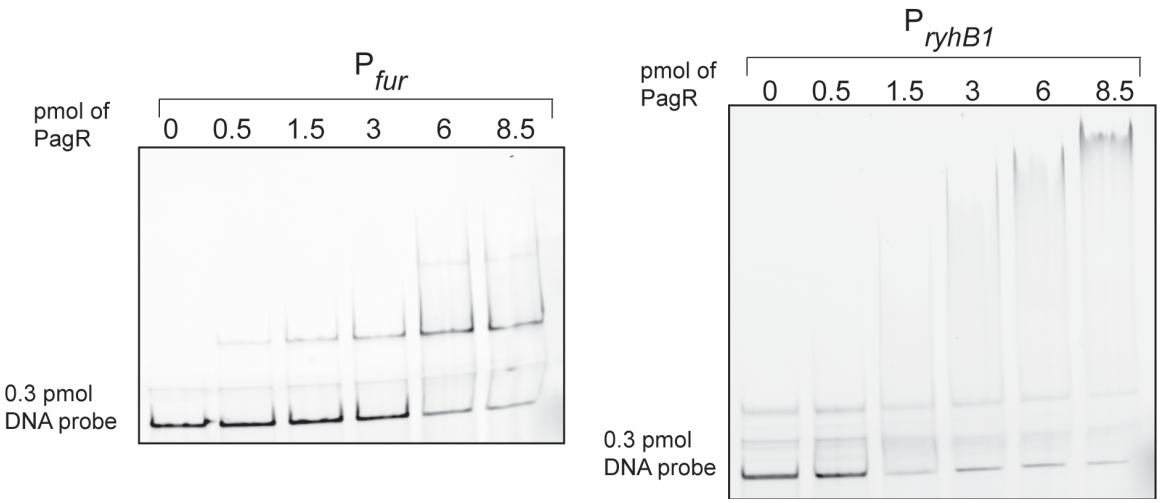

Figure B.6. PagR binds to the promoter of Fur and the small RNA *ryhB1*, targets identified from the N-medium ChIP-seq data set.

APPENDIX C

DYSREGULATION OF CULRI SYNTHESIS VIA DISRUPTION OF MASTER IRON REGULATOR PAGR IN S. ENTERICA[9]

[9]Parks A.R., Lewis Z. A., and Escalante-Semerena J.C. To be submitted.

INTRODUCTION

These data are a preliminary taste of six years of trying to elucidate *in vivo* targets for the transcriptional regulator, PagR. We initially noticed that a deletion of *pagR* causes a loss of congo red dye pigment retention on salt-free agar plates. Congo red dye is typically used to screen for the loss of bacterial curli.

This RNA-seq dataset suggests that the genes involved in the synthesis of curli are down regulated in a *pagR* deletion strain as compared to the *pagR*$^{+}$ strain. We also found that thiamine synthesis and import genes are upregulated in the *pagR* deletion. Finally, we found that deletion of the tktCD genes that are regulated by pagR (refer to appendix A), restores pigment retention in a pagR deletion strain. ChIP-seq was recently performed under the congo red agar condition in order to idenfity in vivo targets. Those data, with the help of Dr. Zack Lewis, will be incorporated into this manuscript soon. These results are preliminary, but I hope you can get a feeling for the manuscript to come from these data.

METHODS

Bacterial strains and growth conditions. All strains were derivatives of Salmonella enterica subsp. Enterica serovar Typhimurium LT2 (hereafter *S. enterica*) using the Wanner and Datsenko protocol that uses l Red recombinase to engineer in-frame deletions (REF Wanner). All bacterial strains and plasmids used in this study are listed in Table SX and SX. Growth of *S. enterica* for experiments using congo red medium was conducted as follows: bacterial cultures were grown from a single colony in salt-free Luria Bertani (LB) broth over 12 hours shaking at

180 rpm at 37C. After 12 hours, 4ul of cells were spotted onto salt-free LB agar plates containing congo red dye (40ug/ml) and brilliant blue dye (20ug/ml) (Hereafter refered to as congo red agar plates) containing the designated amount of L-(+)-arabinose as cited in figure legends. The recipe for congo red plates is as follows: When used, ampicillin was added at 100ug/ml, chloramphenicol at 20ug/ml, and kanamycin at 50ug/ml.

Strain construction. All strains constructed were *metE ara-9* derivatives of *S. enterica*. Primers used in this study were synthesized by Integrated DNA Technologies, Inc. (IDT, Coralville, IA) and are listed in the Supplemental Material. The *pagR*::*kan*$^{+}$ was engineered using the Wanner and Datsenko (2000) protocol as follows: PFU Ultra II Fusion DNA polymerase (Stratagene) was using to amplify flanking regions of plasmid pkD4 (*kan*$^{+}$ marker) using primers containing 36-base pair overhangs corresponding to the beginning or the end of the *pagR* gene. Polymerase chain reaction (PCR) amplifications were visualized by post-staining with ethidium bromide (0.5 μg /ml) for 10 min. Correctly amplified PCR products were cleaned with the Wizard SV gel and PCR clean up kit (Promega). *S. enterica* strain JE6692 (*metE* 205 ara-9 / pKD46 *bla*$^{+}$) was grown to OD600 nm of 0.5 at 30C to maintain the temperature sensitive plasmid pKD46 and was transformed with ~200ng of PCR product using a 0.2-cm electroporation cuvette (MidiSci) and a Bio-Rad Laboratories MicroPulser electroporator on Ec2 setting. Transformed cells were recovered by incubation at 37C with shaking and played on lysogeny broth (LB) agar supplemented with 25ug/ml of kanamycin. Resulting colonies were

PCR verified for correct antibiotic insertions then moved into strain JE6583 by P22-mediated transduction as described elsewhere (Roth). Transduction colonies were streaked to isolation and select colonies were tested for phage sensitivity and Sanger sequenced to verify correct insertion location and sequence using primers flanking gene of interest.

Plasmid construction All plasmids used in this study are listed in Table SX. Primers used in this study were synthesized by Integrated DNA Technologies, Inc. (IDT; Coralville, IA), and are listed in Table SX. The plasmid pPagR-14 containing STM2345 with a N-terminally fused hexahis-tag was designed using the NEBuilder HiFi DNA Assembly.

The high-efficiency cloning method of BspQI digestion (REF) was used to clone the assembled product into the arabinose-inducible complementation vector pCV1.

Chromatin immunoprecipitation. Strain JE24646 (*pagR*::*kan*$^{+}$ / pHis$_6$-PagR) was grown in biological triplicate overnight in salt-free LB (2 ml) with shaking at 37C. After incubation, 4ul of cells were spotted onto congo red agar plates containing 100ug/ml of ampicillin and 100uM of L-(+)-arabinose. Cells were grown at 25C for 72 hours, then each spotted colony was scraped using a sterile loop and deposited into 15 ml falcon tubes. Scraped cells were in 10ml of 1X PBS and were crosslinked with 1% formaldehyde (from Pierce^{TM} 16% Formaldehyde (w/v), methanol free) for 20 minutes shaking at 100 rpm on a table shaker. Excess formaldehyde was quenched with the addition of 0.5M glycine, crosslinked cells were transferred to 50mL conical tubes, and cells were centrifuged at 4,000 x g in

a centrifuge for 10 minutes. The supernatant was discarded and cells were washed with 1ml of 1X PBS (phosphate buffered saline) pH 7.4. Washed cells were split in half, 500ul was snap frozen at -80C. The remaining cells were centrifuged at 6,000 x g for 5 minutes in 1.7ml ependorf tubes and were resuspended in 1ml of ChIP lysis buffer (20mM HEPES, pH 7.9; 50mM KCl; 0.5mM DTT; 500mM NaCl; 10mM imidazole; 1% BSA; 1/8th of a cOmplete^{TM} mini EDTA-free protease inhibitor cocktail tablet; 400uM phenylmethylsulfonyl fluoride [PMSF]). Cells were lysed with the addition of lysozyme (final concentration of 1mg/ml) and incubated on ice for 30 minutes. Lysed cells were transferred to 15ml conical tubes and DNA was sheared four rounds of sonication of 15 minutes, 30 seconds on 30 seconds off using a water bath Diagenode Bioruptor. While DNA was shearing, 100ul per sample of MagneHisTM Ni-Particle beads were pre-equilibrated with 500ul of ChIP lysis buffer by incubation rotating end over end for 30 minutes at room temperature. After sonication, 40ul of sheared cell lysate was removed and frozen as the “input” samples, while the remaining cell lysate was added to 100ul of pre-equilibrated MagneHis^{TM} Ni-Particle beads and were incubated together with end-over-end mixing for 30 minutes at room temperature.

To isolate PagR-bound DNA fragments, Ni-particles with His-PagR bound were washed five times with 500ul of wash buffer (100mM HEPES pH 7.5; 10mM imidazole; 500mM NaCl; 1% BSA; 1/8th cOmplete^{TM} mini EDTA-free protease inhibitor cocktail tablet), using the MagneRack magnetic bead separator. To elute the His-tagged PagR, 100ul of elution buffer (100mM HEPES pH 7.5; 500mM imidazole) was added to the beads and incubated together stationary for 30

minutes at room temperature with occasional mixing. Eluted DNA-protein complexes were separated from the beads and added to fresh PCR tubes. The input samples were thawed and brought to the same volume as the ChIP samples through the addition of MQ water. SDS (final concentration of 1.25%) was added and samples were decrosslinked overnight at 65C in a heated-lid PCR machine. After decrosslinking, 8ul of Rnase A (10mg/ml) was added and samples were incubated for two hours at 50C. Proteinase K (4ul at 20mg/ml) was added and samples were incubated for 2 hours at 50C. Samples were cleaned of protein using the Qiagen MinElute Reaction Cleanup kit (cat# 28204). Final input and ChIP DNA samples were quantified using the Qubit dsDNA HS DNA Assay kit on a Qubit 4 fluormeter.

ChIP-seq library construction and sequencing. 10ng of DNA was used to construct each chip sequencing library starting with end repair of DNA fragments using the NEB Ultra II End Repair Module (cat# E7546S) according to manufacturers instructions. Adaptor ligation of the iTruSeq Illumina adaptor (2.5ul of 0.15uM stock) to DNA fragments was performed using the NEBNext Ultra II Ligation Module (cat# E7595) according to manufacturer's instructions. Self-ligated adaptors were removed from ligated libraries through magnetic binding and washing of DNA using Sera-Mag SpeedBeads (cat# 65152105050250) at a ratio of 1:1 DNA/beads volume. Cleaned adaptor-ligated DNA fragments were amplified using NEB Ultra II Q5 hotstart polymerase (cat# M0543S) and a unique dual index combination of i7/i5 index primer mix. The PCR condition was according to manufacturer except with 11 cycles with annealing temperature of 55C and final

extension of 3 minutes at 72C. PCR amplified chip libraries were cleaned again using the Sera-Mag SpeedBeads and final quantification was conducted using the Qubit dsDNA HS DNA Assay kit on a Qubit 4 fluormeter.

ChIP-seq data analysis

Indexed libraries were diluted to 5ng/ul and pooled equally to a volume of at least 20ul and were sequenced were sequenced on an Illumina NextSeq2000 instrument at the Georgia Genomics and Bioinformatics Core. Resulting FastQ files were processed to remove low quality reads and adaptor sequences using TrimGalore! (https://github.com/FelixKrueger/TrimGalore). Trimmed reads were then mapped to the *Salmonella enterica* subsp. *enterica* serovar Typhimurium LT2 genome (Genbank accession#: GCF_000006945.2) using BWA [1]. Output files were sorted and converted to bam format using SAMtools [2]. Bigwig files were generated using the deeptools bamCoverage command (--binSize 25 --smoothLength 50) [3] and visualized using the Integrative Genomics Viewer [4]. Peaks were called for each replicate using MACS2 [5] (-q =0.001). Peak files generated by MAC2 were filtered using awk to remove peaks with enrichment values < 2.0, and a list of consensus peak files from all three replicates was generated using MSPC [6].

β-Galactosidase assays β-galactosidase assays were conducted as described previously [7]. Three independent cultures were grown overnight in LB (2 ml; LB, Difco) containing 100 μg /ml of ampicillin. In the morning, cultures were

subcultured 1:100 into fresh LB plus ampicillin (100μg/ml). Complementation plasmids were induced with 1mM of L-(+)-arabinose unless otherwise stated. Cells were grown shaking at 180 rpm at 37C until an OD600 nm of 0.4-0.6.

RNA-Seq. Three biological replicates of strains JE6583 and JE21107 were grown shaking overnight at 37C in 2ml of salt-free LB. 4ul of full density cultures were spotted onto congo red agar, and cells were grown at 25C for 48 hours. Colony formations were scraped off the congo red agar plates and transferred to 1.5ml microcentrifuge tubes, then snap frozen in liquid nitrogen and stored at -80C. Cell pellets were sent to GeneWiz for RNA siolation, rRNA depletion using the Ribo-Zero kit by Illumin, and library preparation for Illumina HiSeq 2x150bp configuration, single index strain-specific RNA-sequencing processing. RNA integrit was assessed by Agilent Tapestation and concentration of RNA yield was determined by Qubit assay.

RNA isolation for qRT-PCR. Strains JE22070 (*pagR*$^{+}$ / pVOC), JE21566 (*pagR*::*kan*$^{+}$ / pVOC), JE21577 (*pagR*::*kan*$^{+}$ / pPagR), were grown overnight in triplicate in nutrient broth (2 ml; NB, Difco) with shaking at 37ºC. After incubation, strains were diluted 1:100 into 10 ml N-medium (as described above) supplemented with L(+)-arabinose (100uM), 100ug/ml ampicillin, and 15nM cyanocobalamin. Cultures were grown shaking at 37 ºC to an optical density (600 nm) of 0.6, then 10 ml of each sample were quickly centrifuged in 15 mL Falcon tubes at 4000 x g, supernatant was removed, and pellets were flash-frozen in liquid

nitrogen and kept on dry ice. RNA was isolated following the RNAsnap™ protocol [8]. Pellets were re-suspended in 150 μl of boil solution (ethylenediaminetretraacetic acid (EDTA, 18 mM), SDS (0.025%, w/v) formamide (95%, v/v; RNA grade), 2-mercaptoethanol (1% v/v) in RNase-free water) and were vortexed vigorously to break up the cell pellet. Pellets were incubated at 95 ºC for 7 min and centrifuged at 16,000 x g for 5 min at room temperature; 100 μl of supernatant was transferred to a fresh tube. A sodium acetate/ethanol RNA precipitation was then conducted by the addition of 400 μl of RNase-free water, 50 μl of sodium acetate (3 M, pH 5.2; final concentration of 0.3 M), and finally 1650 μl of ice-cold absolute ethanol (100%), with mixing briefly before the addition of the next reagent. The mixture was incubated at -80C for one hour, centrifuged at 16,000 x g for 30 min at 4ºC, and ethanol was decanted off. Ethanol (300 μl of cold 70% v/v) was added and pellets were centrifuged at 8,000 x g for 5 min at 4ºC in an Eppendorf 5415D centrifuge. Ethanol was removed and pellets were allowed to dry. RNA pellets were re-suspended in RNase-free water. Subsequent RNase-free DNase I treatment was conducted using the Ambion Turbo DNA-free kit according to manufacturer's instructions (ThermoFisher Scientific). After DNA cleavage, a final sodium acetate/ethanol precipitation was performed as described above, except using 360ul of water, 50ul of 3M sodium acetate, and 1500ul of cold 100% ethanol. After one hour incubation at -80C, RNA was centrifuged at 16,000 x g for 30 minutes then washed with 300ul of cold 70% ethanol. RNA decanted of ethanol and pellets were allowed to dry for 20 minutes at room temperature. The final resuspension of RNA was resuspended in 100ul of water. A small aliquot of

each sample was quantified using the RNA Broad Range (BR) Assay kit by Qubit on a Qubit 4 fluometer. A small amount of each sample was also tested for quality and integrity using the Qubit RNA IQ Assay on a Qubit 4 fluometer. Primers for qPCR were designed using Primer 3.

cDNA synthesis and RT-qPCR. Total RNA (972 ng) from each sample was used for the synthesis of cDNA using the iScript^{TM} cDNA synthesis Kit from BioRad Laboratories according to manufacturer's protocol. Each cDNA reaction was then diluted to 5 ng/μl and used as template for PCR. For real-time PCR, 20 μl reactions were prepared with 10 μl of 2X FastSYBR Green master mix (Applied Biosystems), 500 nM of each gene-specific primer (1μl of 10 μM primer stock), and 10 ng of cDNA (2 μl of 5 ng/μl cDNA). The real-time PCR reaction was performed using a 7500 Fast real-time PCR system (Applied Biosystems). The threshold cycle values of *rpoB* and *gyrB* were checked first to ensure that both genes were optimal for use as reference genes for these strains under the conditions chosen for RT-qPCR. Cycle threshold (C_T) data were normalized to the *rpoB* gene [9]. These normalized values (ΔC_T) were transformed using $2(e\text{-}\Delta Ct)/10^{-6}$ [10], and were reported as arbitrary gene expression units (EU), or the gene expression ratio of the mutant strains/the parent strain (JE22070 *pagR*$^{+}$). Mean EU values were used to calculate the standard error of the mean (SEM) using Prism8 from three biological replicates that were each tested in technical triplicate. Differences in EU between mutant strains and JE22070, were compared using Welch's t-test with the Graphpad Prism8 software.

Protein purification PagR protein was purified to homogeneity from plasmid pPagR-8 encoding a PagR protein with a maltose binding protein-hexahistidine (MBP-H6) N-terminal tag. The tag was removed after incubation with recombinant tobacco etch virus (rTEV) protease since the plasmid used to produce MBP-H_6-PagR (pTEV19) contained a rTEV protease cleavage site [11]. A sample (10 ml) of an overnight culture of *E. coli* C41 (DE3) / pPagR-8 strain was used to inoculate one liter of lysogeny broth (LB) containing 100 mg/ml of ampicillin. Cells were grown shaking at 125 rpm at 37 °C until the culture reached an optical density (OD) at 600 nm of 0.7. Expression of genes of interest encoded by the plasmids was induced by the addition of IPTG (0.25 mM) followed by ~12 h of overnight incubation at 37 °C. The next morning cells were harvested by centrifugation at 6,000 x g for 15 min using a refrigerated Beckman-Coulter Avanti J-20-XPI centrifuge equipped with a JLA 8.1 rotor. Cell pellets were resuspended in 20 ml of buffer A [(4-(2-hydroxyethyl-1-piperazineethanesulfonic acid (HEPES) buffer (50 mM, pH 7.5 at 4 ºC) containing NaCl (0.5 M), glycerol (20% v/v), and imidazole (20 mM)] and were sonicated thrice for 30-s intervals, and during each interval, sonication was on for 2 s and off for 2 s, at 60% amplitude. The resulting whole-cell lysates were centrifuged for 30 min at 40,000 x g and the supernatants were filtered with a 0.45-m filter to remove large particulates. Each filtered lysate was applied onto a 1-ml nitrilotriacetic acid (NTA) affinity chromatography column pre-equilibrated with buffer A. Fractions were collected by gravity at 4 °C. The purification was performed as follows: After all the lysate was applied, the column

was washed with 10 column volumes (CV, *i.e.,* 10 mL) of buffer A, 7 CV (*i.e.,* 7 mL) of buffer A containing 4% elution buffer B [HEPES buffer (50 mM, pH 7.5 at 4 ºC) containing NaCl (0.5 M), glycerol (20% v/v), and imidazole (0.5M)], and finally, MBP-H_6-PagR was eluted in two fractions, first with 1 CV (*i.e.,* 1mL) of 100% elution buffer B, and the second fraction being 4 CV (*i.e.,* 4mL) of 100% elution buffer B. Both elution fractions were pooled and MBP-H_6-PagR was cleaved with rTEV protease at a concentration of 1 mg rTEV:100 mg of PagR protein while dialyzing at room temperature in HEPES buffer (50 mM, pH 7.5 at 4 ºC) containing NaCl (0.5M), glycerol (10% v/v), and dithiothreitol (DTT, 1 mM). Cleaved protein was dialyzed twice more at 4 °C in buffer A. Cleaved STM2345 protein was run back over a 2-ml NTA column to remove MBP-His and rTEV protease. Tag-less STM2345 protein did not interact with the NTA resin, and was collected in the flow-through fraction. To further remove MBP from PagR protein, flow-through fractions from the second NTA purification were run over a 1-mL amylose resin to remove contaminating MBP. Pure PagR was dialyzed overnight HEPES buffer (50 mM, pH7.5 at 4ºC) containing NaCl (150 mM) and glycerol (20% v/v) at 4˚C. Dialyzed PagR was stored at -80˚C in 15-ml aliquots that were flash frozen with liquid nitrogen. Protein concentration was determined by Nanodrop using the extinction coefficient and molecular mass of PagR (37 kDa).

Electrophoretic mobility shift assay

To determine PagR binding to promoter sites, EMSAs were conducted as follows: PagR protein purified to homogeneity was incubated at increasing concentrations

of protein with 5'-6-FAM labeled DNA probe (50ng added). EMSA buffer containing 1X of 50mM HEPES pH 7.5, 150mM NaCl, 10% glycerol) was added for a total volume of 25ul. PagR and DNA were incubated at room temperature for 40 minutes. During incubation, a 7.5% Tris-boric acid-EDTA (TBE) polyacrylamide gel (BioRad, Criterion) was pre-electrophoresed at 100 volts for 40 minutes in 0.5X TBE buffer at 4C. After incubation 5ul of 50% glycerol was added to reactions and 20ul of each reaction was loaded onto the gel. Xylene cyanol and bromophenol blue was added to an empty lane to use as a tracking indicator and the gel was electrophoresed until the bromophenol blue reached the bottom of the gel. The EMSA was imaged using a Typhoon Trio imager at 525 nm with the 488 (Blue) filter.

REFERENCES

1. Li, H. and R. Durbin, *Fast and accurate short read alignment with Burrows-Wheeler transform.* Bioinformatics, 2009. **25**(14): p. 1754-60.
2. Danecek, P., et al., *Twelve years of SAMtools and BCFtools.* Gigascience, 2021. **10**(2).
3. Ramirez, F., et al., *deepTools2: a next generation web server for deep-sequencing data analysis.* Nucleic Acids Res, 2016. **44**(W1): p. W160-5.
4. Robinson, J.T., et al., *Integrative genomics viewer.* Nat Biotechnol, 2011. **29**(1): p. 24-6.

5. Zhang, Y., et al., *Model-based analysis of ChIP-Seq (MACS).* Genome Biol, 2008. **9**(9): p. R137.

6. Jalili, V., et al., *Using combined evidence from replicates to evaluate ChIP-seq peaks.* Bioinformatics, 2018. **34**(13): p. 2338.

7. Miller, F.D. and C.L. Hershberger, *A quantitative beta-galactosidase alpha-complementation assay for fusion proteins containing human insulin B-chain peptides.* Gene, 1984. **29**(1-2): p. 247-50.

8. Stead, M.B., et al., *RNAsnap: a rapid, quantitative and inexpensive, method for isolating total RNA from bacteria.* Nucleic Acids Res., 2012. **40**: p. e156.

9. Rocha, D.J., C.S. Santos, and L.G. Pacheco, *Bacterial reference genes for gene expression studies by RT-qPCR: survey and analysis.* Antonie Van Leeuwenhoek, 2015. **108**: p. 685-693.

10. Livak, K.J. and T.D. Schmittgen, *Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method.* Methods, 2001. **25**: p. 402-408.

11. VanDrisse, C.M. and J.C. Escalante-Semerena, *New high-cloning-efficiency vectors for complementation studies and recombinant protein overproduction in Escherichia coli and Salmonella enterica.* Plasmid, 2016. **86**: p. 1-6.

Table C.1. Strains and plasmids used in this work		
***S. enterica* strain**	**Relevant genotype**	**Description; Source, reference**
JE6583	*metE205 ara-9*	Laboratory collection
Derivatives of JE6583		
JE22070	*metE205 ara-9* / pCV1	
JE21107	*pagR::kan*+	
JE21566	*pagR::kan*+ / pCV1	
JE21577	*pagR::kan*+ / pPagR-7	
JE25373	*stm2340-stm2341::cat*+	
JE25443	*pagR::kan*+ *stm2340-stm2341::cat*+	
JE6692	*metE205 ara-9* / pKLD46	Laboratory collection
Plasmid	**Genotype**	**Description; Source, Reference**
pCV1	*araC*+ *bla*+	pBAD24 with BspQI MCS
pSTM2345-7	*pagR*+ *bla*+	*stm2345*+ cloned into pCV1
pPagRX	*pagR-6xHis bla*+	*Stm2345*-6xHis fusion cloned into pCV1

STM2345^{+} *tktC*$^{+}$ | Δ*STM2345* | Δ*STM2345*/ *p*STM2345

Figure C.1. A *pagR* deletion strain does not retain congo red dye on salt free LB supplemented with congo red.

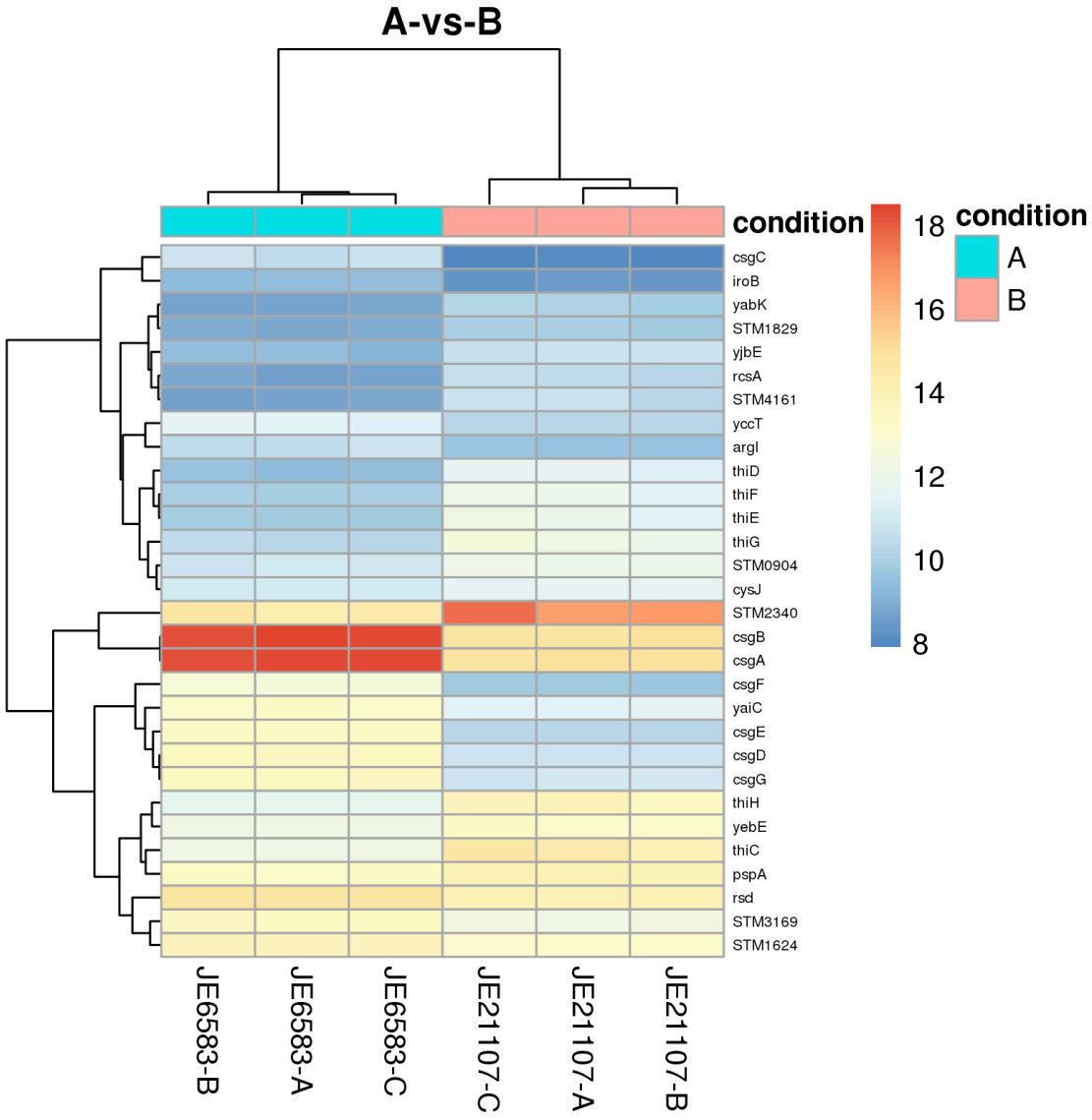

Figure C.2. Curli biosynthesis genes are down-regulated and thiamine-related genes are upregulated in a p*agR* deletion grown on salt-free LB. RNA-seq box plot analysis of genes with significant mRNA transcript changes in a pagR deletion (JE21107, A-C) as compared to the parental strain control (JE6583, A-C). The color intensity bar to the right indicates scale of down-regulated to up-regulated genes, where more red signifies upregulation and blue signifies downregulation.

STM2345^{+} *tktC*$^{+}$ | Δ*STM2345* | Δ*STM2345*/ *p*STM2345 | Δ*STM2345* Δ*tktC*

Figure C.3. Reversion of *pagR* pigment phenotype occurs when *tktCD* is deleted in a pagR deletion strain.

APPENDIX D

PAGR REGULATES BOTH THE HIGH AND LOW AFFINITY ACETATE METABOLIC PATHWAYS IN S. ENTERICA[10]

[10]Parks A.R., and Escalante-Semerena J.C. To be submitted.

INTRODUCTION

These data are a continuation of the work conducted by Dr. Kristy Hentchel in the Escalante lab. She found that the transcriptional repressor for myo-inositol degradation activates the transcription of the protein lysine acetyltransferase gene, pat. Pat is required for modulation of *S. enterica* growth using low concentrations of acetate (10mM), known as the high-affinity acetate pathway. The low-affinity acetate pathway is regulated by the enzymes Pta and AckA. Here we have found that PagR binds to the promoter of *ackA* and alters its transcription in an iron-dependent manner. PackA was also identified as an a potential in vivo target in the ChIP-seq data set of Appendix B. We also found that PagR binds to the promoter of iolR, and indirectly measured the affect of PagR on P_{iolR} by analyzing its known target, P_{pat}. These data suggest that PagR is able to directly regulate the high-affinity acetate pathway and indirectly regulate the low affinity acetate pathway in *S. enterica*. These data are preliminary but I hope they can give you a glimpse of the manuscript to come.

METHODS

Bacterial strains and growth conditions. All strains were derivatives of Salmonella enterica subsp. Enterica serovar Typhimurium LT2 (hereafter *S. enterica*) using the Wanner and Datsenko protocol that uses l Red recombinase to engineer in-frame deletions (REF Wanner). Growth of *S. enterica* for experiments using N-medium was conducted as follows: bacterial cultures were grown from a single colony in nutrient broth (NB [Difco]) over 12 hours shaking at 180 rpm at

37C. In the after 12 hours, cells were subcultured1:100 into N-medium containing 0.1% casamino acids, 38 mM glycerol, 10uM $MgCl_2$, 1uM KH_2PO_4, 7.5mM $(NH_4)_2SO_4$, 100mM Tris-HCl pH 7.4, 5mM KCl, 0.5mM K_2SO_4, 15nM cyanocobalamin, and the designated amount of L-(+)-arabinose as cited in figure legends. When used, ampicillin was added at 100ug/ml, chloramphenicol at 20ug/ml, and kanamycin at 50ug/ml. Cells were grown shaking at 180 rpm at 37C to optical density 600nm of 0.5.

Strain construction. All strains constructed were *metE ara-9* derivatives of *S. enterica*. Primers used in this study were synthesized by Integrated DNA Technologies, Inc. (IDT, Coralville, IA) and are listed in the Supplemental Material. The *pagR*::*kan*$^+$ was engineered using the Wanner and Datsenko (2000) protocol as follows: PFU Ultra II Fusion DNA polymerase (Stratagene) was using to amplify flanking regions of plasmid pkD4 (*kan*$^+$ marker) using primers containing 36-base pair overhangs corresponding to the beginning or the end of the *pagR* gene. Polymerase chain reaction (PCR) amplifications were visualized by post-staining with ethidium bromide (0.5 □g /ml) for 10 min. Correctly amplified PCR products were cleaned with the Wizard SV gel and PCR clean up kit (Promega). *S. enterica* strain JE6692 (*metE* 205 ara-9 / pKD46 *bla*$^+$) was grown to OD600 nm of 0.5 at 30C to maintain the temperature sensitive plasmid pKD46 and was transformed with ~200ng of PCR product using a 0.2-cm electroporation cuvette (MidiSci) and a Bio-Rad Laboratories MicroPulser electroporator on Ec2 setting. Transformed cells were recovered by incubation at 37C with shaking and played on lysogeny broth (LB) agar supplemented with 25ug/ml of kanamycin. Resulting colonies were

PCR verified for correct antibiotic insertions then moved into strain JE6583 by P22-mediated transduction as described elsewhere (Roth). Transduction colonies were streaked to isolation and select colonies were tested for phage sensitivity and Sanger sequenced to verify correct insertion location and sequence using primers flanking gene of interest.

Plasmid construction All plasmids used in this study are listed in Table SX. Primers used in this study were synthesized by Integrated DNA Technologies, Inc. (IDT; Coralville, IA), and are listed in Table SX. The plasmid pSTM2345-14 containing STM2345 with a N-terminally fused hexahis-tag was designed using the NEBuilder HiFi DNA Assembly as follows: GET FROM NOTEBOOK.

The high-efficiency cloning method of BspQI digestion (REF) was used to clone the assembled product into the arabinose-inducible complementation vector pCV1.

β-Galactosidase assays β-galactosidase assays were conducted as described previously [1]. Three independent cultures were grown overnight in LB (2 ml; LB, Difco) containing 100 μg /ml of ampicillin. In the morning, cultures were subcultured 1:100 into fresh LB plus ampicillin (100μg/ml). Complementation plasmids were induced with 1mM of L-(+)-arabinose unless otherwise stated. Cells were grown shaking at 180 rpm at 37C until an OD600 nm of 0.4-0.6.

RNA isolation Strains JE22070 (*pagR*$^{+}$ / pVOC), JE21566 (*pagR*::*kan*$^{+}$ / pVOC), JE21577 (*pagR*::*kan*$^{+}$ / pPagR), were grown overnight in triplicate in nutrient broth

(2 ml; NB, Difco) with shaking at 37ºC. After incubation, strains were diluted 1:100 into 10 ml N-medium (as described above) supplemented with L(+)-arabinose (100uM), 100ug/ml ampicillin, and 15nM cyanocobalamin. Cultures were grown shaking at 37 ºC to an optical density (600 nm) of 0.6, then 10 ml of each sample were quickly centrifuged in 15 mL Falcon tubes at 4000 x g, supernatant was removed, and pellets were flash-frozen in liquid nitrogen and kept on dry ice. RNA was isolated following the RNAsnap™ protocol [2]. Pellets were re-suspended in 150 μl of boil solution (ethylenediaminetretraacetic acid (EDTA, 18 mM), SDS (0.025%, w/v) formamide (95%, v/v; RNA grade), 2-mercaptoethanol (1% v/v) in RNase-free water) and were vortexed vigorously to break up the cell pellet. Pellets were incubated at 95 ºC for 7 min and centrifuged at 16,000 x g for 5 min at room temperature; 100 μl of supernatant was transferred to a fresh tube. A sodium acetate/ethanol RNA precipitation was then conducted by the addition of 400 μl of RNase-free water, 50 μl of sodium acetate (3 M, pH 5.2; final concentration of 0.3 M), and finally 1650 μl of ice-cold absolute ethanol (100%), with mixing briefly before the addition of the next reagent. The mixture was incubated at -80C for one hour, centrifuged at 16,000 x g for 30 min at 4ºC, and ethanol was decanted off. Ethanol (300 μl of cold 70% v/v) was added and pellets were centrifuged at 8,000 x g for 5 min at 4ºC in an Eppendorf 5415D centrifuge. Ethanol was removed and pellets were allowed to dry. RNA pellets were re-suspended in RNase-free water. Subsequent RNase-free DNase I treatment was conducted using the Ambion Turbo DNA-free kit according to manufacturer's instructions (ThermoFisher Scientific). After DNA cleavage, a final sodium acetate/ethanol precipitation was

performed as described above, except using 360ul of water, 50ul of 3M sodium acetate, and 1500ul of cold 100% ethanol. After one hour incubation at -80C, RNA was centrifuged at 16,000 x g for 30 minutes then washed with 300ul of cold 70% ethanol. RNA decanted of ethanol and pellets were allowed to dry for 20 minutes at room temperature. The final resuspension of RNA was resuspended in 100ul of water. A small aliquot of each sample was quantified using the RNA Broad Range (BR) Assay kit by Qubit on a Qubit 4 fluometer. A small amount of each sample was also tested for quality and integrity using the Qubit RNA IQ Assay on a Qubit 4 fluometer. Primers for qPCR were designed using Primer 3.

cDNA synthesis and RT-qPCR. Total RNA (972 ng) from each sample was used for the synthesis of cDNA using the iScript^{TM} cDNA synthesis Kit from BioRad Laboratories according to manufacturer's protocol. Each cDNA reaction was then diluted to 5 ng/μl and used as template for PCR. For real-time PCR, 20 μl reactions were prepared with 10 μl of 2X FastSYBR Green master mix (Applied Biosystems), 500 nM of each gene-specific primer (1μl of 10 μM primer stock), and 10 ng of cDNA (2 μl of 5 ng/μl cDNA). The real-time PCR reaction was performed using a 7500 Fast real-time PCR system (Applied Biosystems). The threshold cycle values of *rpoB* and *gyrB* were checked first to ensure that both genes were optimal for use as reference genes for these strains under the conditions chosen for RT-qPCR. Cycle threshold (C_T) data were normalized to the *rpoB* gene [3]. These normalized values (ΔC_T) were transformed using $2(e\text{-}\Delta Ct)/10^{-6}$ [4], and were reported as arbitrary gene expression units (EU), or the gene expression ratio of

the mutant strains/the parent strain (JE22070 *pagR*$^{+}$). Mean EU values were used to calculate the standard error of the mean (SEM) using Prism8 from three biological replicates that were each tested in technical triplicate. Differences in EU between mutant strains and JE22070, were compared using Welch's t-test with the Graphpad Prism8 software.

Protein purification PagR protein was purified to homogeneity from plasmid pPagR-8 encoding a PagR protein with a maltose binding protein-hexahistidine (MBP-H6) N-terminal tag. The tag was removed after incubation with recombinant tobacco etch virus (rTEV) protease since the plasmid used to produce MBP-H_6-PagR (pTEV19) contained a rTEV protease cleavage site [5]. A sample (10 ml) of an overnight culture of *E. coli* C41 (DE3) / pPagR-8 strain was used to inoculate one liter of lysogeny broth (LB) containing 100 mg/ml of ampicillin. Cells were grown shaking at 125 rpm at 37 °C until the culture reached an optical density (OD) at 600 nm of 0.7. Expression of genes of interest encoded by the plasmids was induced by the addition of IPTG (0.25 mM) followed by ~12 h of overnight incubation at 37 °C. The next morning cells were harvested by centrifugation at 6,000 x g for 15 min using a refrigerated Beckman-Coulter Avanti J-20-XPI centrifuge equipped with a JLA 8.1 rotor. Cell pellets were resuspended in 20 ml of buffer A [(4-(2-hydroxyethyl-1-piperazineethanesulfonic acid (HEPES) buffer (50 mM, pH 7.5 at 4 ºC) containing NaCl (0.5 M), glycerol (20% v/v), and imidazole (20 mM)] and were sonicated thrice for 30-s intervals, and during each interval, sonication was on for 2 s and off for 2 s, at 60% amplitude. The resulting whole-

cell lysates were centrifuged for 30 min at 40,000 x g and the supernatants were filtered with a 0.45-m filter to remove large particulates. Each filtered lysate was applied onto a 1-ml nitrilotriacetic acid (NTA) affinity chromatography column pre-equilibrated with buffer A. Fractions were collected by gravity at 4 °C. The purification was performed as follows: After all the lysate was applied, the column was washed with 10 column volumes (CV, *i.e.,* 10 mL) of buffer A, 7 CV (*i.e.,* 7 mL) of buffer A containing 4% elution buffer B [HEPES buffer (50 mM, pH 7.5 at 4 ºC) containing NaCl (0.5 M), glycerol (20% v/v), and imidazole (0.5M)], and finally, MBP-H_6-PagR was eluted in two fractions, first with 1 CV (*i.e.,* 1mL) of 100% elution buffer B, and the second fraction being 4 CV (*i.e.,* 4mL) of 100% elution buffer B. Both elution fractions were pooled and MBP-H_6-PagR was cleaved with rTEV protease at a concentration of 1 mg rTEV:100 mg of PagR protein while dialyzing at room temperature in HEPES buffer (50 mM, pH 7.5 at 4 ºC) containing NaCl (0.5M), glycerol (10% v/v), and dithiothreitol (DTT, 1 mM). Cleaved protein was dialyzed twice more at 4 °C in buffer A. Cleaved STM2345 protein was run back over a 2-ml NTA column to remove MBP-His and rTEV protease. Tag-less STM2345 protein did not interact with the NTA resin, and was collected in the flow-through fraction. To further remove MBP from PagR protein, flow-through fractions from the second NTA purification were run over a 1-mL amylose resin to remove contaminating MBP. Pure PagR was dialyzed overnight HEPES buffer (50 mM, pH7.5 at 4ºC) containing NaCl (150 mM) and glycerol (20% v/v) at 4˚C. Dialyzed PagR was stored at -80˚C in 15-ml aliquots that were flash frozen with liquid

nitrogen. Protein concentration was determined by Nanodrop using the extinction coefficient and molecular mass of PagR (37 kDa).

Electrophoretic mobility shift assay Electrophoretic mobility shift assays (EMSAs) were performed to quantify PagR binding to DNA. EMSAs were performed as follows: PagR protein purified to >95% homogeneity was incubated at the concentrations in pmol of protein labeled in the figures with a 5'-6-FAM-labeled DNA probe at 50 ng [*i.e.,* 0.3 pmol] was added). EMSA buffer [HEPES buffer (50 mM, pH 7.5 at 4ºC) containing NaCl (150 mM), and glycerol (10% v/v)] was added to the reaction mixture (total volume = 25 μl) and DNA and protein were incubated at room temperature for 40 min. During incubation, a 7.5% Tris-boric acid-EDTA (TBE) polyacrylamide gel was pre-developed at 100 V for 40 min in 0.5X TBE buffer at 4˚C. After incubation, 5 μl of glycerol (50% v/v) was added to the reaction mixtures, and 20 μl of each reaction mixture was resolved by the polyacrylamide gel. A lane of xylene cyanol and bromophenol blue dye was added as a tracking indicator, and the gel was run until bromophenol blue reached the bottom of the gel. The gel was imaged using a Typhoon Trio imager at 525 nm with the 488 (Blue) filter.

REFERENCES

1. Miller, F.D. and C.L. Hershberger, *A quantitative beta-galactosidase alpha-complementation assay for fusion proteins containing human insulin B-chain peptides.* Gene, 1984. **29**(1-2): p. 247-50.

2. Stead, M.B., et al., *RNAsnap: a rapid, quantitative and inexpensive, method for isolating total RNA from bacteria.* Nucleic Acids Res., 2012. **40**: p. e156.

3. Rocha, D.J., C.S. Santos, and L.G. Pacheco, *Bacterial reference genes for gene expression studies by RT-qPCR: survey and analysis.* Antonie Van Leeuwenhoek, 2015. **108**: p. 685-693.

4. Livak, K.J. and T.D. Schmittgen, *Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method.* Methods, 2001. **25**: p. 402-408.

5. VanDrisse, C.M. and J.C. Escalante-Semerena, *New high-cloning-efficiency vectors for complementation studies and recombinant protein overproduction in Escherichia coli and Salmonella enterica.* Plasmid, 2016. **86**: p. 1-6.

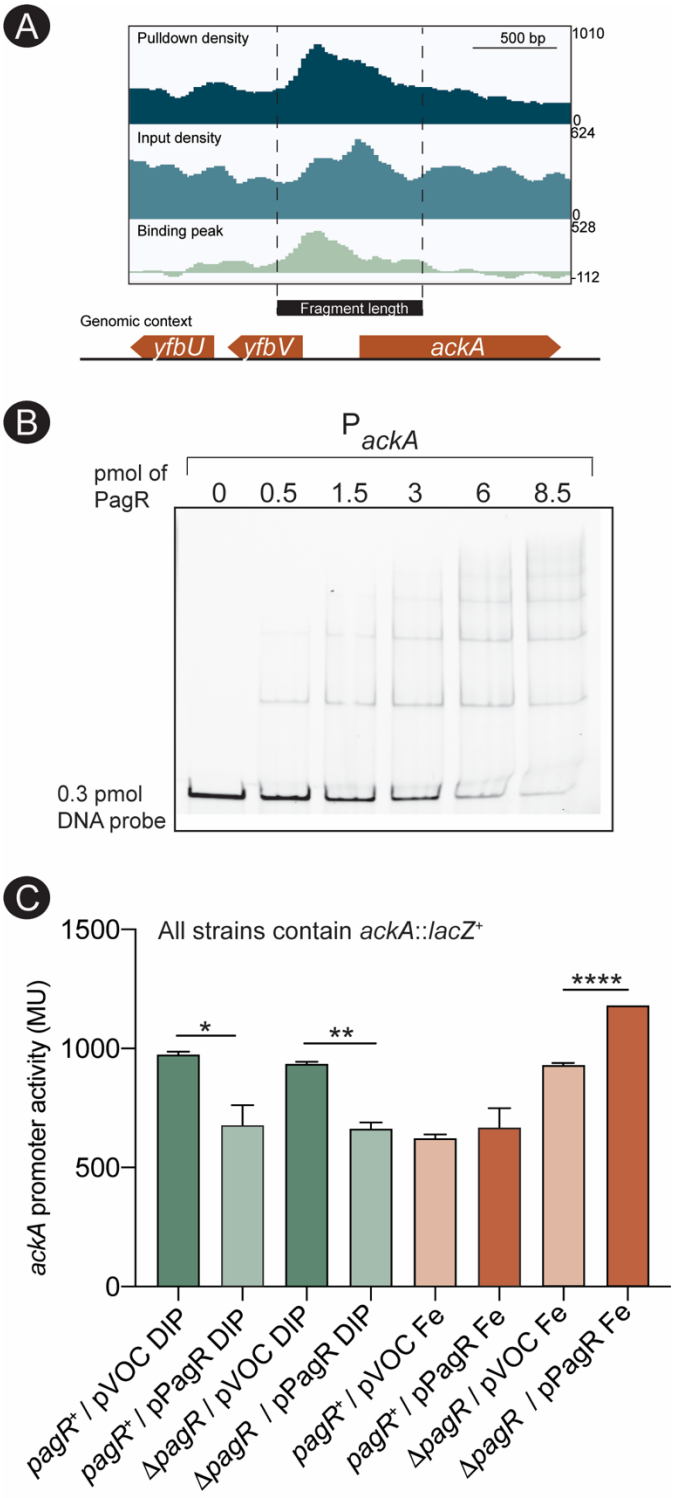

Figure D.1. PagR binds to the *ackA* promoter and regulates P_{ackA} in an iron-dependent manner.

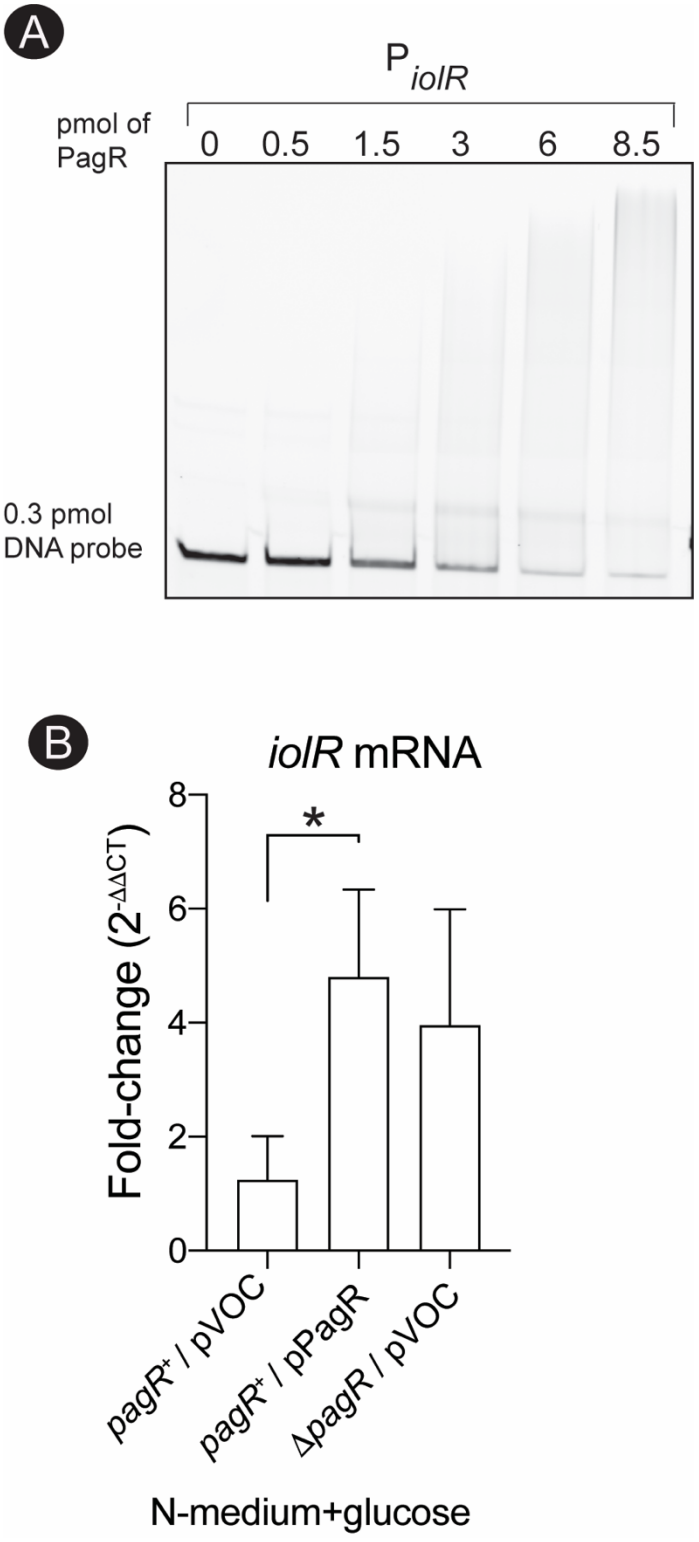

Figure D.2. PagR binds to the P_{iolR} promoter and represses P_{iolR} transcription on low phosphate, low magnesium N-medium.

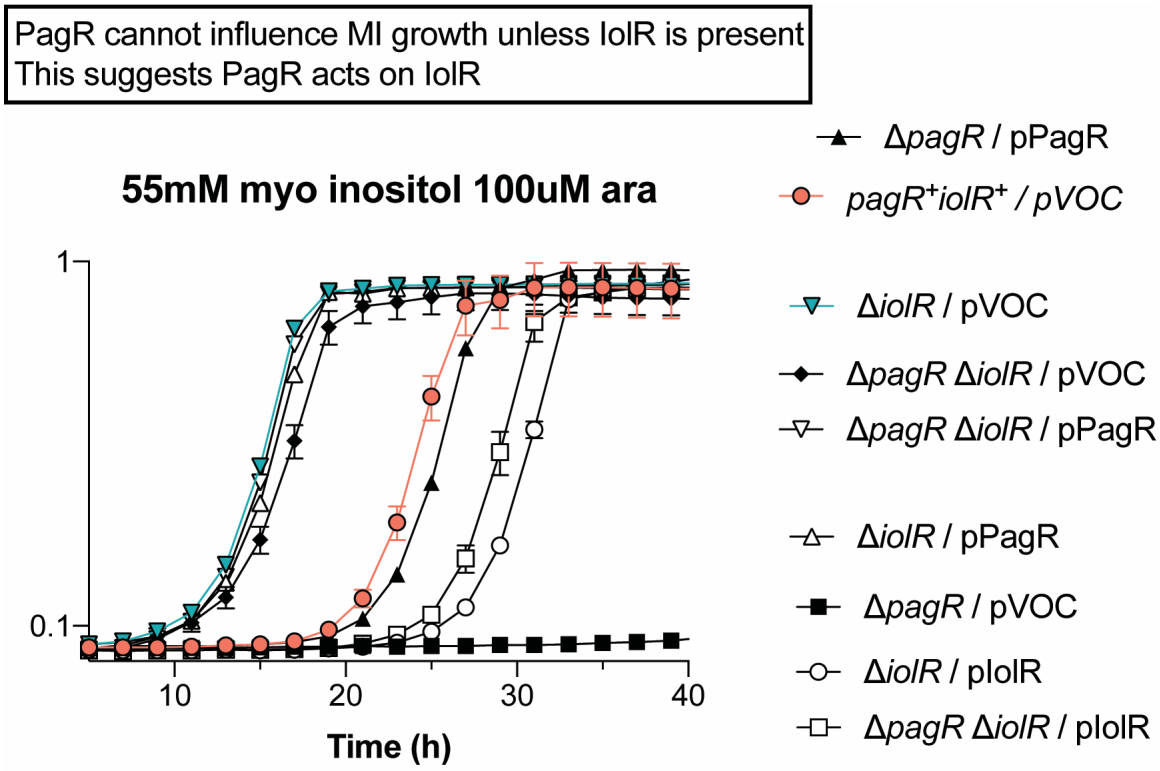

Figure. D.3. *S. enterica* cannot use myo-inositol as a carbon and energy source of *pagR* is deleted. IolR is the master repressor for myo-inositol degradation.

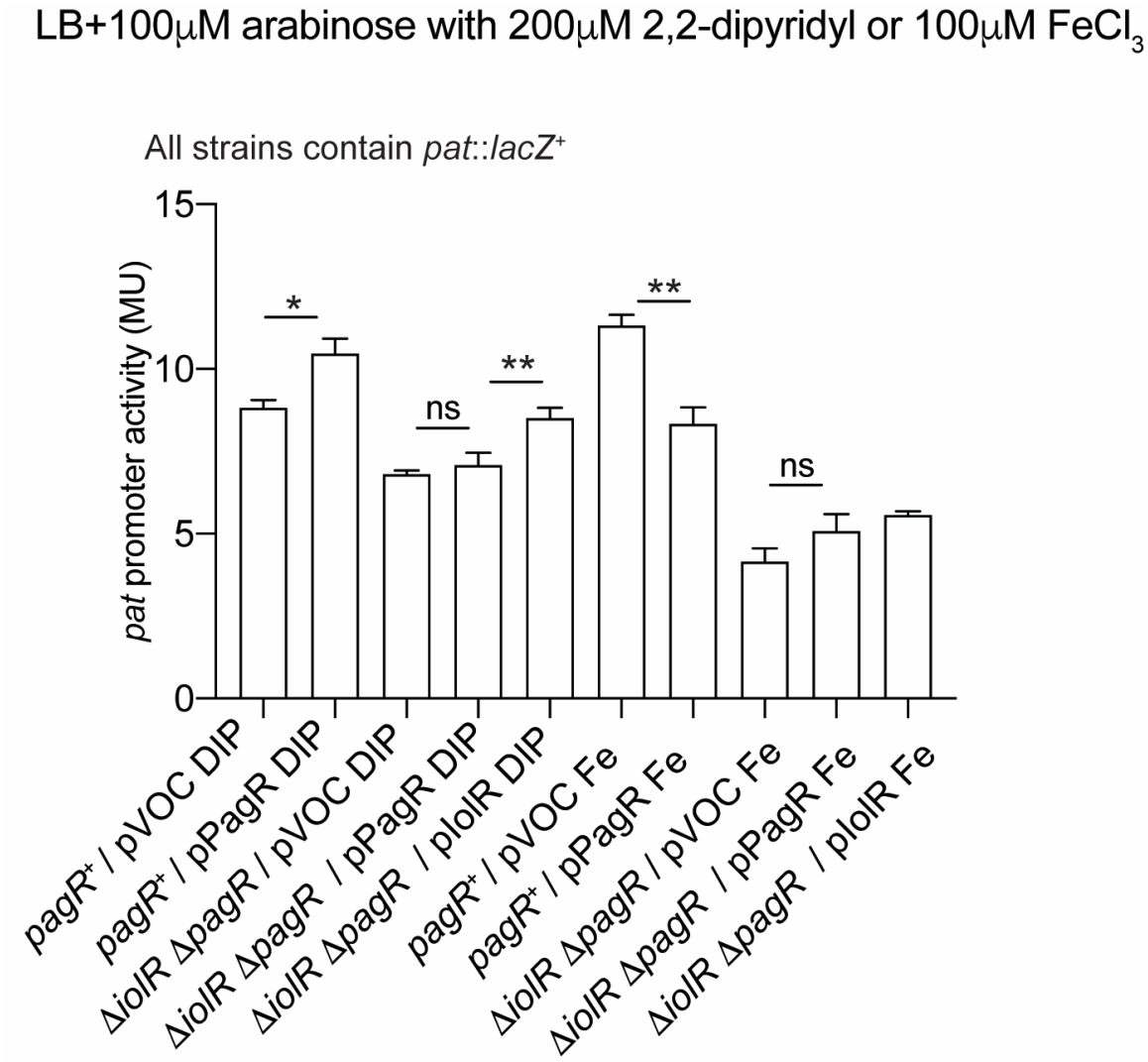

Figure D.4. The expression of *pat* is directly regulated by IolR, and indirectly regulated by PagR via P_{iolR}-mediated regulation.