2018

Control of Salmonella Virulence by Microbiota-Derived and Dietary Fatty Acids

Zhenrun J. Zhang

CONTROL OF *SALMONELLA* VIRULENCE BY MICROBIOTA-DERIVED AND DIETARY FATTY ACIDS

A Thesis Presented to the Faculty of

The Rockefeller University

in Partial Fulfillment of the Requirements for

the degree of Doctor of Philosophy

by

Zhenrun J. Zhang

June 2018

CONTROL OF *SALMONELLA* VIRULENCE BY MICROBIOTA-DERIVED AND DIETARY FATTY ACIDS

Zhenrun J. Zhang, Ph.D.

The Rockefeller University 2018

The intestinal microbiota plays critical roles in human physiology and diseases. While recent research has revealed many mechanisms by which gut microbiota influences host immunity to defend against invading pathogens, how microbiota directly antagonizes pathogen virulence is less studied. In particular, gut microbiota produces large amounts and varieties of small molecules that may impact both host immunity and pathogen virulence. In this thesis, I describe how fatty acids, derived from both gut microbiota and diet, contribute to attenuation of virulence of enteric pathogen *Salmonella*.

In Chapter 1, I review how dietary and microbiota metabolites affect different aspects of host-microbe interactions. These metabolites are categorized into microbial-associated molecular patterns and microbiota-derived secondary metabolites. Small molecules reviewed in this chapter not only enhances host innate and adaptive immunity, but also directly inhibit virulence of invading pathogens, providing colonization resistance to the host. In some cases, pathogens could exploit these metabolites as environmental signals to enhance its survival and expansion. These findings highlight the importance of understanding the intricate interactions between host and microbiota, and should provide insights in developing microbiota-targeting therapeutics for host physiology, immunity, and pathogen resistance.

In Chapter 2, I describe a mechanism by which microbiota-derived short-chain fatty acids inhibit virulence of *Salmonella* Typhimurium. Short-chain fatty acids can inhibit *Salmonella* virulence, but the molecular mechanism(s) remain poorly characterized. We use a chemical reporter strategy to identify molecular targets of short-chain fatty acids in *Salmonella.* I demonstrate that alkynyl-functionalized short-chain fatty acids can be metabolized and covalently attached to proteins in *Salmonella*. Proteomic analysis reveal that HilA, a key virulence transcription regulator, is short-chain fatty acylated. I employ Amber Suppression Technology and CRISPR-Cas9 genome editing to faithfully mimic butyrylation on endogenous HilA. Biochemical and functional characterization show that acylation of HilA has site-specific effect, and K90 butyrylation affect HilA DNA-binding activity and *Salmonella* invasion in mice. Overall, our results discover a mechanism by which gut microbiota provides resistance against *Salmonella* through short-chain fatty acids.

In Chapter 3, I describe long-chain fatty acylation of HilA and biochemical characterization of HilA. I find that dietary long-chain fatty acids potently inhibit *Salmonella* virulence. Chemical proteomics with alkynyl-functionalized long-chain fatty acids reveal proteins that are long-chain fatty acylated in *Salmonella*, including HilA. Modification by long-chain fatty acids on HilA is post-translationally *N*-linked. Moreover, with photo-crosslinking unnatural amino acid, we discover that HilA forms homo-oligomers in *Salmonella*. Our data suggest that dietary long-chain fatty acids may interfere pathogenesis of *Salmonella* through post-translational modification, and further structural characterization of HilA may reveal novel target for treatment of *Salmonella* infection.

The projects described in this thesis underscore the important roles microbiota and dietary metabolites have played in host immunity and enteric pathogen restriction.

DEDICATION

I dedicate this thesis to Theobald Smith,

who first discovered *Salmonella enterica* and after whom Smith Hall is named,

and to Simon Flexner,

after whom *Shigella flexneri* and Flexner Hall are named,

both of whom worked at Rockefeller Institute for Medical Research.

ACKNOWLEDGMENTS

It has been a great journey and a pleasure to do my Ph.D. studies in the lab of Prof. Howard C. Hang. Howard has supported my exploration in the research and provided needed guidance and critical comments to my projects. Your breadth of knowledge inspire me to expand my interest in science, and the high standard you set always encourages me to push my limit and achieve better accomplishment. I will continue to see you as a role model for the rest of my academic career.

I am also grateful for Prof. Luciano A. Marraffini for serving as my thesis committee chair, Prof. C. David Allis for serving as my thesis committee member, and Prof. Jorge E. Galan for agreeing to be the external member of my thesis committee. Your critical comments and suggestions have helped my project greatly.

I want to thank all the Hang Lab members as well. Dr. Tao Peng and Dr. Virginia Pedicord are instrumental in helping my projects become what they are now. My research would not be possible without your scientific guidance and technical support. Dr. Nathan Westcott, Dr. Emmanuelle Thinon, Dr. Yen-Chih Wang, Dr. Byungchul Kim, Dr. Qiang Li, Dr. Charles Hespen, Dr. Xinglin Yang, Dr. Rayshonda Hardy, Dr. Ruina He, Dr. Xiaonan (Elsa) Zhao, Dr. Jordan RoseFigura, Dr. Srinivasan Ramakrishnan, and Dr. Lisa Ambrosini Vadola, you have made Hang Lab a great place and thank you for your support in the lab. Dr. Kavita Rangan, Dr. Xiaoqiu Yuan, Dr. Avital Percher, Emma Garst, Taku Tsukidate, Tandrilla Das, Juliel Espinosa, it has been a pleasure to have you as fellow students in Hang Lab.

I thank Dr. Wenyan Jiang, Zhen Chen, Dr. Qiao Wang, Cheng Lyu, Zhe Yang, Zikun Wang, Putianqi Wang, and Dr. Xin Jin for your companion at the Rockefeller University. I also thank Dr. Lei Wei, Dr. Kan Lin, Dr. Yuchen Qi, Zeda Zhang, Shijie Zhao, Dr. Liuchuan Tong, Dr. Fanghao Hu, Dr. Lu Wei, Lixue Shi, and Zhilun Zhao, you have enriched my life in New York and in the U.S.

The Rockefeller University has been a heaven for graduate students. I thank all the support from Prof. Sidney Strickland, Dr. Emily Harms, Marta Delgado, Kristen Cullen, Cristian Rosario, and Stephanie Fernandez of Deans Office. I also thank John L. Gerlach for organizing Tri-I Noon Recitals, as well as all the other staffs that has helped make the university a great scientific community.

My academic journey starts at Peking University, and I am very thankful for the mentorship and support from Prof. Peng R. Chen, Dr. Shixian Lin, and Dr. Jie Li. I would not be able to pursue my academic career without your first push when I was at my starting line.

I owe my life to my wonderful parents, who have supported me throughout my journey in the academia. Lastly, I want to thank my beloved wife, Dr. Yihui Shen, for being such a wonderful academic and life companion of mine.

Zhenrun J. Zhang

April 2018

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1.

Chapter 2.

Chapter 3.

LIST OF TABLES

Chapter 1

Mechanisms of dietary and microbiota metabolites on host-microbe interactions

Introduction

Recent advances on microbiota profiling and mechanisms have revealed its critical roles in human health and disease. More than 90% of human microbiota resides in the gastrointestinal tract, making the intestinal microbiota the major target of microbiome research. Intestinal microbiome as a whole broadly affects host physiology and responses to intestinal and systemic diseases. The composition of the intestinal microbiota is dynamic and is influenced by environmental factors including host diet, exposure to drugs, infection as well as genetic factors.

The intestine, placed in the latter part of the digestive tract, is a melting pot of complex constituents. Food in the intestine has to be broken down into smaller pieces, in terms of both physical form and chemical composition. Smaller dietary molecules, derived from plants and animals, could be absorbed by the intestine, or be converted by intestinal microbiota into secondary metabolites. Moreover, indigestible macromolecules could also be utilized by intestinal microflora and be fermented into smaller and accessible nutrients. In fact, intestinal microbiota produces significant amounts of metabolites that can function as signaling molecules to modulate host physiology and disease (**Fig. 1.1**).

This chapter summarizes recent studies of specific metabolites derived from both diet and microbiota, and their potential biochemical effects on both host and microbes. While biopolymers common in all kingdoms of life, including double-stranded DNA and RNA, lipopolysaccharide (LPS), and lipopeptides, are long recognized as ligands for pattern recognition receptors (PRRs), especially Toll-like receptors (TLRs) on host cells, this chapter will focus on small, soluble metabolites specifically from microbiota and diet. Some of these metabolites have been recently reviewed (*1*, *2*), and this chapter aims to expand the scope and include recent advances. A molecular understanding of intestinal microbiota–host interactions is pivotal to medicine and

human health. Insight into the function of dietary and microbiota metabolites could help design targeted therapeutics against a variety of diseases and advance personalized medicine.

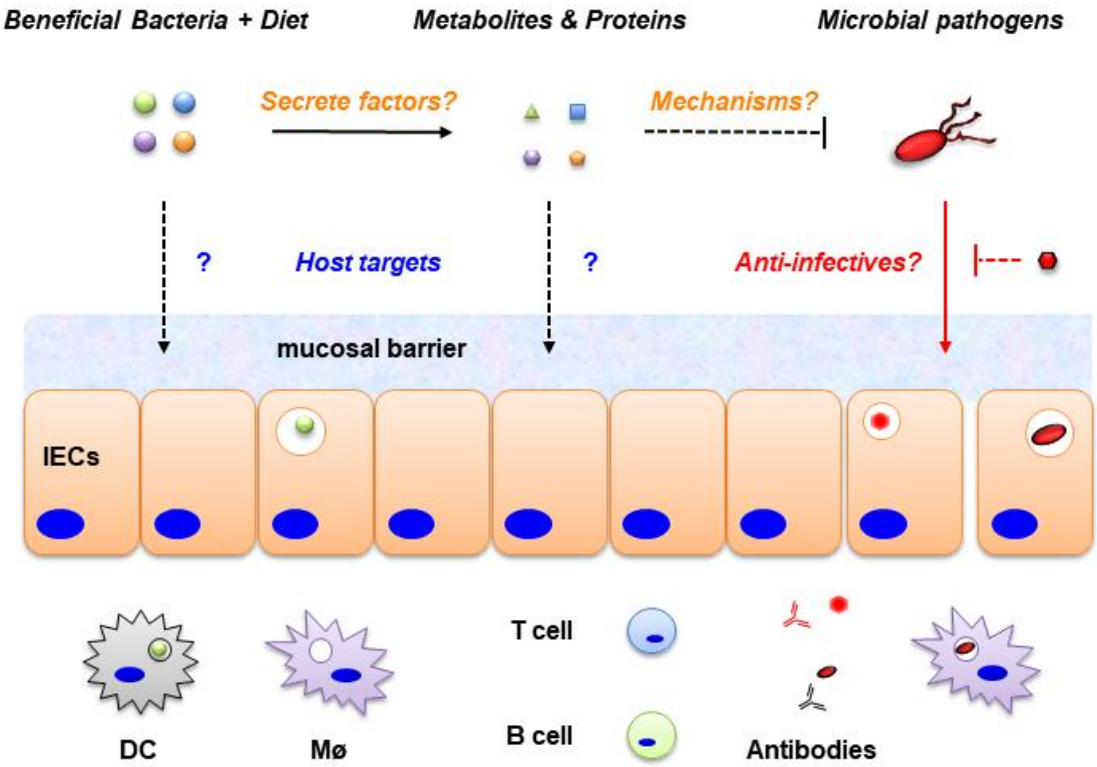


Figure 1.1. Dietary and microbiota-derived metabolites modulate host physiology and immunity and pathogen virulence.

Microbe-Associated Molecular Patterns

Microbe-associated molecular patterns (MAMPs) are molecules derived from microbiota that are specifically recognized by PRRs of the host and could trigger downstream signaling events, leading to host resistance or tolerance towards microbes. While these molecules were first characterized through studies on host-pathogen interactions and initially termed “pathogen-associated molecular patterns (PAMPs)”, they are now accepted as critical molecules from commensal microbes as well. Depending on the context, these molecules could stimulate either

pro- or anti-inflammatory responses. Being the first family of innate PRRs identified, Toll-like receptors (TLRs) and their cognate ligands have been described in detail elsewhere (*3*), so this chapter will focus on other more recently identified MAMPs as well as their receptors in the host.

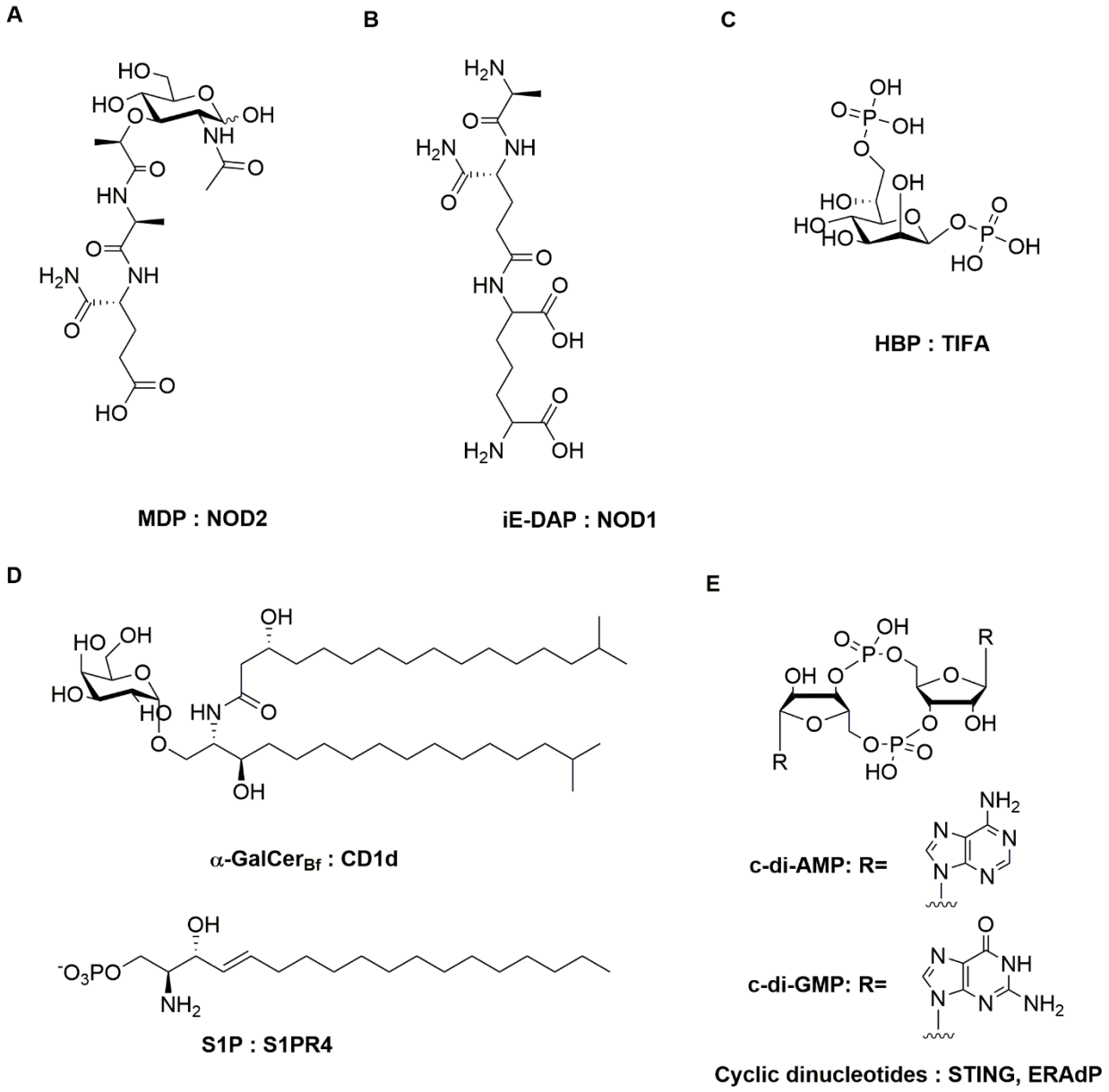


Figure 1.2. Microbe-associated molecular pattern metabolites. Their corresponding host receptors are listed after the colon. (**A**) MDP : NOD2. (**B**) iE-DAP : NOD1. (**C**) HBP : TIFA. (**D**) α-galactosylceramide : CD1d ; sphingosine-1-phosphate : S1PR4. (**E**) c-di-AMP, c-di-GMP: STING, ERAdP.

Muramyl dipeptide (MDP) and γ-D-glutamyl-*meso*-DAP (iE-DAP)

Peptidoglycan (PG) is one of the most abundant macromolecules in a bacterial cell, typically forming a mesh-like structure that encloses the cytoplasmic membrane (*4*–*6*). PG is structurally distinct from cell wall components in archaea and single-celled eukaryotes, making it ideal as a PAMP for mammalian cells. Peptidoglycan is composed of polysaccharide chains with alternating ß-1,4-linked *N*-acetylglucosamine (GlcNAc) and *N*-acetylmuramic acid (MurNAc) residues, which are crosslinked via short peptides (*4*, *5*). These peptides contain D-amino acids such as D-alanine or D-glutamate, as well as unusual non-proteinogenic amino acids, such as *meso*-diaminopimelic acid (*meso*-DAP). The amino acid sequence of the peptides and the structure of crosslinks are variable between bacterial species (*7*–*9*). Although PG is hidden from the innate immune system in Gram-negative bacteria by an outer membrane, soluble PG turnover products and its biosynthesis intermediates could be released from both intact and lysed cells into the surrounding milieu. Mammalian cells utilize multiple PG recognition receptors to detect these MAMPs and initiate an inflammatory response.

Nucleotide-binding oligomerization domain (NOD)-like receptors (NLRs) are cytosolic proteins that play a pivotal role in the regulation of the host innate immune system (*10*). They act as scaffolding proteins that assemble signaling complexes that trigger NF-κB and MAPK signaling pathways. In particular, NOD1 and NOD2 sense PG fragments from bacteria. NOD1 senses peptide fragments of PG, with dipeptide γ-D-glutamyl-*meso*-DAP (iE-DAP) (**Fig. 1.2B**) as the minimal unit (*11*, *12*), which is conserved among most Gram-negative bacteria and some Gram-positive bacteria (*13*). NOD2 senses muramyl dipeptide (MDP) (**Fig. 1.2A**) (*14*), which is found in nearly all Gram-positive and Gram-negative bacteria. Following sensing these microbial PG fragments, NOD1 and NOD2 directly recruit receptor-interacting protein 2 (RIP2) through caspase recruiting domain (CARD)-CARD interactions (*15*, *16*). This leads to activation of IKK complex

and phosphorylation as well as degradation of inhibitor IκB, and further causes NF-κB to translocate to the nucleus and activate the innate immune response (*17*). NOD1 and NOD2 directly senses invading bacterial pathogens *in vitro* and *in vivo*, such as Gram-negative *Shigella flexneri* (*18*), *Salmonella* Typhimurium (*19*), Gram-positive *Mycobacterium tuberculosis* (*20*), and *Listeria monocytogenes* (*21*). Interestly, our laboratory recently showed that a secreted peptidase SagA from *Enterococcus faecium* could protect both worms and mice from *Salmonella* Typhimurium infection, and the protection is mediated through PG fragments generated by SagA enzymatic activity (*22*, *23*). Further experiments showed that SagA generates GlcNAc-β(1–4)-MurNAc dipeptide (GlcNAc-MDP), which could stimulate NOD2 signaling in mammalian cells (Kim B *et al*., unpublished). Indeed, the protective effect by SagA and SagA-expressing bacteria depends on the expression of NOD2 *in vivo* (*23*). These results suggest that commensal bacteria in the gut could harness stimulatory effects of PG fragments to prime the host innate immune system against enteric pathogens.

Heptose-1,7-bisphosphate (HBP)

Lipopolysaccharide (LPS) is a vital cell membrane component in Gram-negative bacteria (*24*). LPS has been well characterized as a ligand for TLR4 and stimulate MyD88-dependent TRIF-dependent innate immune response (*3*). LPS are large molecules consisting of lipid A as the membrane anchor, and a polysaccharide part composed of O-antigen (or O polysaccharide) and core oligosaccharide (*25*). The core oligosaccharide directly attaches to lipid A, and commonly contains sugars such as heptose. Recently, Gaudet *et al.* reported that heptose-1,7-bisphosphate (HBP) (**Fig. 1.2C**), a key biosynthesis intermediate of LPS, could trigger inflammatory response in the mammalian host (*26*). They further identified TRAF-interacting protein with forkhead-associated domain (TIFA) as the critical mediator of signaling axis, although direct sensor of HBP in host cells remains unknown. HBP induces TIFA phosphorylation and oligomerization at

lysosomal compartments, which would then trigger oligomerization of ubiquitinylated TRAF6, and ultimately NF-κB activation pathway as well as innate immune gene expression. Interestingly, while HBP could be released from lysed Gram-negative bacteria, intact *Neisseria* species actively excrete HBP into its surrounding environment. *Neisseria gonnorrhoeae* may use HBP as a stimulator of NF-κB pathway, which could drive HIV gene expression as well as viral shedding and transmission in *Neisseria* and HIV coinfection setting (*27*, *28*). Elucidation of the full signaling pathway of HBP will not only provide essential insights for controlling infection but will also serve as a model for identifying how MAMPs signal to the host and influence host immunity.

Sphingolipids

Sphingolipids are a class of lipids characterized by a long-chain amino alcohol sphingoid backbone with an amide-bound fatty acyl chain. While sphingolipid production is ubiquitous in eukaryotes, only a small subset of bacteria could produce sphingolipids. To date, known sphingolipid-producing bacteria include the majority of the Bacteroidetes phylum together with a few members of the Chlorobi phylum (*29*), as well as a subset of Alphaproteobacteria and Deltaproteobacteria (*30*, *31*). The initial step of sphingolipid synthesis involves the condensation of an amino acid and a fatty acid via the serine palmitoyltransferase (SPT) enzyme, which is highly conserved in both eukaryotes and bacteria (*32*). Within the human gut, members of the Bacteroidetes are known to produce sphigophospholipids that resemble sphingomyelin, an abundant sphingolipid in mammalian membranes (*33*, *34*). This raises the question of whether bacteria in the gut has evolved to exploit sphingolipid signaling pathways in their hosts.

In fact, *Bacteroides fragilis*, a common gut commensal, synthesizes α-galactosylceramide (α-$GalCer_{Bf}$) (**Fig. 1.2D**), which is structurally similar to the synthetic potent CD1d activator KRN7000 (*34*). However, the subtle structural difference renders α-$GalCer_{Bf}$ as CD1d antagonist, which

leads to reduced colonic invariant natural killer T (iNKT) cells and protection against induced colitis (*35*).

Certain sphingolipids from gut microbiota that resembles sphingosine-1-phosphate (S1P) (**Fig. 1.2D**) may also signal through S1P receptor S1PR4, a GPCR that specifically expressed in lymphoid tissue (*36*). S1PR4 signaling induces the chemotaxis of natural killer cells and dendritic cells, as well as modulates Th2 immune responses (*37*). The recent discovery of *N*-acyl amides produced by gut microbiota could also act as specific ligands for S1PR4 (*38*), which suggests that gut microbiota may have a variety of molecules that could modulate host immunity through sphingolipid signaling pathways.

Cyclic dinucleotides

Besides bacterial cell wall components that are essential for bacterial survival, bacteria as a community also use a variety of signaling molecules to communicate with each other for collective behavior. One particular class of signaling molecules are cyclic nucleotides that include cyclic adenosine 3',5'-monophosphate (cGMP), cyclic guanosine 3',5'-monophosphate (cAMP), 3',3'-cyclic dimeric GMP (c-di-GMP) (**Fig. 1.2E**), 3',3'-cyclic GMP-AMP (3',3'-cGAMP), and 3',3'-cyclic dimeric AMP (c-di-AMP) (**Fig. 1.2E**). They mediate numerous critical pathways in the bacterial community, such as biofilm formation, chemotaxis, motility, and virulence (*39*, *40*).

Mammalian hosts have also evolved specific sensors for cyclic nucleotides as means of immune recognition. Indeed, c-di-GMP could be specifically sensed by STING (stimulator of IFN genes; also known as MITA, ERIS, MPYS, and TMEM173), an ER-localized transmembrane protein (*41*, *42*). Its C-terminal cytosolic domain (CTD) could bind to c-di-GMP (*43*), and the binding event triggers recruitment and activation of IRF3. After a series of phosphorylation events mediated by TBK1, IRF3 would dimerize and enter cell nucleus to activate transcription of relevant genes,

resulting in type I interferon production. Interestingly, mammalian cells also sense non-physiological double-stranded DNA (dsDNA) through STING (*44*). The cyclic GMP-AMP synthase (cGAS) in the cytosol could bind to tiny amounts of dsDNA and adopt a conformational change, which in turn catalyze cytosolic GTP and ATP to synthesize 2',3'-cyclic GMP-AMP (2',3'-cGAMP) (*45*), which is structurally similar to c-di-GMP. The endogenous 2',3'-cGAMP could bind even more potently to STING and activate type I interferon response. These findings suggest mammals have used STING as a sensor for bacteria and dsDNA virus infection, all mediated through cyclic dinucleotides.

Since c-di-AMP is also very similar to c-di-GMP and 2',3'-cGAMP, it is thought to bind to STING and trigger innate immune response as well. However, STING affinities towards c-di-AMP and c-di-GMP (K_d = 1 to 5 uM) are much weaker than that towards 2',3'-cGAMP (K_d = 51 nM), making its identity as a bona fide physiological sensor of c-di-AMP in question (*46*). Recently, Xia *et al.* showed that ERAdP, an ER adaptor protein, directly senses c-di-AMP (*46*). C-di-AMP binds to the CTD of ERAdP at high affinity (K_d = 76 nM), which leads to recruitment of TAK1. This would initiate activation of transcription factor NF-κB and induce production of pro-inflammatory cytokines in immune cells. This ERAdP-TAK1 signaling axis is required for the eradication of *Listeria monocytogenes* infection.

Microbiota-derived secondary metabolites

Gut microbiota metabolizes nutrients from host diet, uses them for energy source as well as bacterial cell building blocks. At the same time, microbes also generates a broad range of secondary metabolites. Some may be regarded as side products of bacterial metabolism pathways, and some appear to have no apparent functions on bacterial community at first sight.

As shown in the following, many secondary metabolites may not directly impact microbiota itself, but play pivotal roles in regulating host immunity, and even directly antagonize invading pathogens, thus maintaining the homeostasis of microbiota as well as the host.

Short-Chain Fatty Acids (SCFA)

Undigested complex dietary fibers are abundant substrates for bacterial fermentation in the colon; therefore they are also termed microbiota-accessible carbohydrates (MACs) (*47*). Their main metabolic products are short-chain fatty acids (SCFAs), including acetic acid, propionic acid, and butyric acid (**Fig. 1.3A**). SCFA concentrations in the gut range from 5 mM to 140 mM (*48*), depending on the location of the intestine, microbiota composition, and the MAC content of the host diet. SCFAs are critical energy sources not only for the gut microbiota itself but also for intestinal epithelial cells (IECs), especially colonocytes. In addition to acting as local substrates for energy production, SCFAs have diverse regulatory functions on host physiology and immunity, which recent exciting discoveries continue to reveal.

SCFAs are ligands for G protein-coupled receptors (GPCRs), and thereby act as signaling molecules that influence the expansion and function of many cell lineages. GPCRs responding to SCFAs include GPR43 (also known as FFAR2), GPR41 (also known as FFAR3) and GPR109A (also known as HCAR2), which are expressed by numerous cell types, including immune cells and IECs. GPR43 expression is necessary for SCFA-induced neutrophil chemotaxis (*49*) and the expansion and suppressive function of forkhead box P3 (FOXP3)+ regulatory T (Treg) cells (*50–52*). SCFA-mediated activation of GPR109A, a receptor that responds to both niacin and butyric acid, prevented colitis and colon carcinogenesis through increased expression of anti-inflammatory effector molecules by monocytes and induced differentiation of Treg cells (*53*). Binding of SCFAs to GPR43 and GPR109A on IECs also activated inflammasome assembly and

increased production of the downstream inflammatory cytokine IL-18 (*54*), which prevented a colitogenic phenotype, enhanced IEC integrity, and fortified IEC barrier function (*55*). Outside the gut, SCFA–GPR43 interactions decrease chemotaxis and inflammatory gene expression in neutrophils (*56*), and downregulates gout-associated inflammation by mediating inflammasome assembly and immune cell clearance of monosodium urate crystals (*57*). SCFAs blocked DC maturation through GPR41 signaling and ameliorated allergic airway inflammation (*58*).

The GPR43-dependent effects of SCFAs on host physiology also extend to the central nervous system (CNS). The maturation and function of microglia, which are the resident macrophages of the CNS, were dependent on the gut microbiota, and the maintenance of microglia homeostasis required SCFAs and GPR43 (*59*). However, SCFAs can also exacerbate disease. In an α-synuclein (αSyn)-dependent Parkinson's disease mice model, SCFAs from the gut accelerate αSyn aggregation and microglia activation in mice brain, thus exacerbate motor dysfunction (*60*). Therefore, the immunomodulatory effects of SCFAs depend on the context and cell type under investigation.

SCFA are also inhibitors of histone deacetylases (HDACs). Several studies demonstrate that SCFA-induced HDAC inhibition downregulates of NF-κB activity and pro-inflammatory innate immune responses in neutrophils (*49*, *61*), macrophages (*62*, *63*) and dendritic cells (DCs) (*58*, *64*). Moreover, SCFAs also influence peripheral T cells, particularly regulatory T (Treg) cells, through HDAC inhibition (*50–52*). HDAC inhibition by SCFAs increased FOXP3 expression, leading to amplified Treg cell numbers, increased Treg cell frequency, and enhanced Treg cell suppressive function *in vivo* (*65*). Only high millimolar SCFA concentrations are sufficient to perturb HDAC function (*66*), and their effects may require specific transporters (*64*). SCFA-driven inhibition of HDACs tends to promote a tolerogenic, anti-inflammatory cell phenotype that is crucial for maintaining immune homeostasis, and this activity supports the concept that the microbiota can function as an epigenetic regulator of host physiology.

SCFAs not only modulate host immunity, but also metabolism of host intestinal epithelial cells. Enterocytes use butyrate for β-oxidation, which depletes oxygen diffused from blood vessel and lamina propria (*67*). Butyrate also signals through PPAR-γ to limit the availability of oxidative species in the gut lumen (*68*). This helps maintenance of anaerobic environment in the gut. Depletion of butyrate by antibiotics or gut inflammation switches metabolic program in enterocytes from β-oxidation to glycolysis, leading to accumulation of lactate and oxygen in the gut lumen, both of which drives expansion of facultative anaerobes, including *Salmonella* (*67*).

SCFAs also directly modulate virulence of various pathogens. In particular, SCFAs could differentially regulate expression of *Salmonella* virulence genes. While acetate enhances the virulence gene expression through BarA/SirA two-component system signaling in *Salmonella* (*69*), propionate and butyrate could downregulate the expression (*69*, *70*). Specifically, propionate may post-translationally modify HilD, a key transcription regulator of *Salmonella* virulence, and affect its stability (*71*). Loss of butyrate utilization pathway in *Salmonella* attenuates *Salmonella* virulence (*72*), but the mechanism through which butyrate represses *Salmonella* virulence remains unknown. We are proposing that butyrate dampens *Salmonella* virulence through post-translational modification on HilA, another key transcription regulator of *Salmonella* virulence (Chapter 2). More studies are needed to investigate the immunomodulatory functions and therapeutic potential of SCFAs in health and disease.

Retinoic acid

Vitamin A (VA) is a lipophilic micronutrient obtained by dietary ingestion of pro-vitamin A carotenoids (such as β-carotene) and retinyl esters (RE). Humans are not able to synthesize VA, thus depend on dietary supply for maintenance of multiple physiological processes throughout the human body (reviewed in (*73*)). After intake, carotenes are enzymatically converted in the intestine into retinol, which is transported from the gut lumen into the cell cytoplasm and is rapidly converted to RE or to retinoic acid (RA) (*74*), with *all-trans*-RA (*at*RA) (**Fig. 1.3B**) being the physiologically most abundant and well-studied. RA could be sensed by receptors that are broadly classified into two subgroups, retinoic acid receptor (RAR) and retinoid X receptor (RXR). In the absence of ligands, RAR/RXR heterodimers constitutively bind to retinoic acid response elements (RARE) and suppress the transcription of target genes (*75*). Both *at*RA and 9-*cis*-RA can bind to RAR/RXR, displacing the corepressors and activating target gene expression (*76*–*78*). Most of the RA immune-related functions signal through the RAR/RXR pathway, primarily driven by *at*RA acting through RARα (*79*).

VA is crucial to the establishment of oral immunological tolerance against food antigens (*80*, *81*), and its deficiency might contribute towards food allergies and inflammatory bowel diseases (IBD). Notably, *at*RA participates in multiples steps in the establishment of oral immunological tolerance (*80*). *At*RA could dramatically expand pre-mucosal DCs (pre-μDCs) population that expresses α4β7, a gut homing receptor on lymphocytes (*82*). Furthermore, *at*RA in the lamina propria of intestine enhances differentiation of pre-μDCs into CD103+ DCs, a critical DC population for sensing antigens from the gut lumen (*83*). In combination with CD103+ DCs and TGF-β, *at*TA increases the generation of FOXP3+ Treg cells in mesenteric lymph nodes (MLNs) (*84*), restricting Th17 differentiation (*85*–*87*), as well as induces in T cells expression of α4β7 and CCR9, another gut homing receptor. Treg cells induced in this manner can produce anti-

inflammatory IL-10 and migrate to the intestine to promote tolerance (*88*, *89*). Innate lymphoid cells (ILCs) in the MLN could also be induced to express CCR9 and α4β7 by *at*RA, which leads to ILC migration to the intestine (*90*). The concentrations of *at*RA in the small intestine follows a proximal to distal (i.e., duodenum to colon) decreasing gradient (*91*, *92*), which correlates with the differential distribution of different ILCs population along the gastrointestinal tract (*93*). Therefore, it may be reasonable to speculate that *at*RA might be involved in the regionalization of ILCs in the intestine, likely by controlling their migration, differentiation, and function.

Aromatic acids

Gut lumen is filled with digested nutrients from the diet, including amino acids. In particular, aromatic amino acids (tryptophan, phenylalanine, tyrosine, histidine) could be metabolized by gut microbiota and converted into aromatic acids through a series of oxidative or reductive pathways. Of note, several of them, including indoleacetic acid, indolepropionic acid (IPA) (**Fig. 1.3C**), and indolelactic acid (ILA) (**Fig. 1.3C**), could be absorbed by the intestine and circulate in the mice serum (*94*). Aromatic acids are well-known agonists of aryl-hydrocarbon receptor (AhR), a ligand-activated transcription factor involved in lymphoid system development (*95*, *96*), immune response (*97*), and toxic response (*98*). Aromatic acids also activate pregnane X receptor (PXR), another crucial nuclear receptor involved in intestinal permeability and toxic response (*99*). Therefore, elevated aromatic acids level in systemic sites induced by microbiota have a potentially profound impact on host physiology.

Indeed, *Lactobacillus reuteri* catabolizes L-tryptophan and produces ILA, which activates AhR in CD4+ T cells, allowing downregulation of the transcription factor ThPOK and differentiation into CD4+CD8αα+ double-positive intraepithelial T lymphocytes (DP IELs) (*100*). DP IELs have a regulatory function complementary to that of Tregs and promote tolerance to dietary antigens (*101*). *Lactobacillus* species also produce indole-3-aldehyde, another derivative of tryptophan,

which is also an AhR ligand and contributes to AhR-dependent IL-22 production. This leads to balanced mucosal response, provides colonization resistance to the fungus *Candida albicans*, and mucosal protection from inflammation (*102*). In another study, Wlodarska *et al.* found that several *Peptostreptococcus* species could produce indoleacrylic acid (IAA) (**Fig. 1.3C**), which promotes intestinal epithelial barrier function and mitigates inflammatory responses through AhR (*103*). On the other hand, Venkatesh *et al.* showed that IPA produced by *Clostridium sporogenes* could downregulate enterocyte TNF-α while it promotes tight junction between IECs through activation of PXR *in vivo*, which also depend on TLR4 (*104*). Recently, Dodd *et al.* engineered *C. sporogenes* to make it defective in producing IPA, and comparing to mice colonized with wild-type *C. sporogenes*, mice with *C. sporogenes* mutant have increased intestinal permeability and adaptive immune response to *C. sporogenes* (*94*). These exciting studies highlight how microbial-derived metabolites could significantly change host physiology and immunity.

Riboflavin precursors

Mucosal-associated invariant T (MAIT) cells are innate-like T cells that comprise up to 10% of the peripheral blood T-cell population of humans (*105*), and they are involved in microbial response (*106*, *107*) as well as autoimmune diseases (*108*). MAIT cells can be activated by most bacteria and yeast through antigen presentation by MR1, an MHC-I like molecule (*109*, *110*). However, the specific antigens presented by MR1 had been long elusive. During the structural determination of MR1, Kjer-Nielsen *et al.* found that MR1 could bind to 6-formyl pterin (6-FP), a photodegradation product of folic acid (*111*). Even though 6-FP could not activate MAIT cells, it provided the first hint on what ligands MR1 could potentially bind. Further functional assays revealed that several riboflavin (Vitamin B2) biosynthesis precursors, namely reduced 6-hydroxymethyl-8-*D*-ribityllumazine (rRL-6-CH_2OH) (**Fig. 1.3D**), 7-hydroxy-6-methyl-8-*D*-ribityllumazine (RL-6-Me-7-OH) (**Fig. 1.3D**), and 6,7-dimethyl-8-*D*-ribityllumazine (RL-6,7-diMe)

(**Fig. 1.3D**), could activate MAIT cell TCR in an MR1-dependent manner (*111*). Interestingly, riboflavin itself could not activate MAIT cells, and microbes that do not have riboflavin biosynthesis pathway could not activate MAIT cells (*111*).

In a follow-up study, Corbett *et al.* found that MAIT-cell activation requires microbial genes encoding enzymes that form 5-amino-6-*D*-ribitylaminouracil (5-A-RU), an early intermediate in bacterial riboflavin synthesis (*112*). Although 5-A-RU does not bind MR1 or activate MAIT cells directly, it does undergo non-enzymatic reactions with other small molecules from microbes, such as glyoxal and methylglyoxal. The resulting adducts, 5-(2-oxoethylideneamino)-6-*D*-ribitylaminouracil (5-OE-RU) (**Fig. 1.3D**) and 5-(2-oxopropylideneamino)-6-*D*-ribitylaminouracil (5-OP-RU) (**Fig. 1.3D**), even though unstable, could bind to MR1 as reversible covalent Schiff base complexes (*112*). Thus, MR1 can capture, stabilize and present chemically unstable pyrimidine intermediates as potent antigens to MAIT cells. Since riboflavin is synthesized by plants and most bacteria and yeasts but not by animal cells, while MAIT invariant TCR is evolutionarily conserved among mammals and other vertebrates (*113*), these two findings exemplify co-evolution of invariant TCRs, MHC-I like molecules, and their cognate microbial-specific small molecule antigens.

Secondary bile acids

Bile acids are amphipathic molecules synthesized from cholesterol in the liver. They are physiological detergents that help secrete metabolites into the gastrointestinal tract. In the intestines, bile acids help intestinal absorption of dietary fats, fat-soluble vitamins, and other nutritions (*114*). Cholic acid (CA) and chenodeoxycholic acid (CDCA) are the dominant primary bile acids in humans. Of note, deoxycholic acid (DCA) (**Fig. 1.3E**) is a secondary bile acid formed by bacterial dehydroxylation of cholic acid; lithocholic acid (LCA) (**Fig. 1.3E**) is formed by the analogous dehydroxylation of chenodeoxycholic acid. DCA and LCA could be further conjugated

with taurine or glycine by gut microbiota, generating conjugated bile acids, including taurodeoxycholic acid (TDCA).

In recent decades, bile acids are recognized not only as digestive surfactants, but also as important signaling molecules in a broad range of biological functions, including glucose and lipid metabolism, energy homeostasis, and the modulation of immune response (*114–116*). The regulatory functions of bile acids are mainly the result of activation of a nuclear receptor, the farnesoid X receptor (FXR, NR1H4) (*117–119*), and a cell surface GPCR, TGR5 (GPBAR-1) (*120*, *121*). FXR activation inhibits pro-inflammatory cytokine production *in vivo*, and limits bacterial overgrowth as well as intestinal permeability (*122*). TGR5 activation leads to inhibition of pro-inflammatory NF-κB pathway (*123*, *124*), as well as activation of endothelial iNOS and NO release (*125*, *126*). Collectively, FXR and TGR5 signaling promotes intestinal barrier function and exerts immune modulation in the gut.

Interestingly, while CDCA is the most potent ligand for FXR (*127–129*), LCA and taurine-conjugated LCA are most potent endogenous ligands for TGR5 (*130–132*). This indicates that microbiota-dependent bile acid metabolism plays critical roles in regulating host signaling pathways and physiology outcomes. Indeed, abnormal bile acid metabolism has been associated with liver injury, metabolic disorders, cardiovascular diseases, as well as digestive system diseases, such as inflammatory bowel disease (IBD) and colorectal cancer (*133–135*).

Secondary bile acids also direct influence virulence of enteric pathogens (*136*, *137*). Bile acids could activate Type III Secretion System 2 (T3SS2) in *Vibrio parahaemolyticus* (*138*). Specifically, TDCA binds to VtrA/VtrC complex, activating VtrB and *V. parahaemolyticus* virulence (*139*). Besides, secondary bile acids inhibit spore germination, growth, and toxin production of *Clostridium difficile* (*140*). *Clostridium scindens*, a bile acid 7α-dehydroxylating intestinal bacterium, enhances resistance to *C. difficile* infection in a secondary bile acid dependent manner (*141*). Moreover, bile acids repress invasion gene expression in *Salmonella* by post-translational

destabilization of HilD (*142*). These findings have implications for the rational design of microbiome-based diagnostics and therapeutics for individuals with enteric infections.

Taurine, histamine, and spermine

Besides acids, gut microbiota could also produce amines, including taurine, histamine, and spermine (**Fig. 1.3F**). Taurine is scavenged by gut microbes from taurine-conjugated bile acids produced by the host, while some bacteria could also perform taurine conjugation on bile acids. Histamine and spermine could be synthesized from histidine and ornithine. In inflammasome-deficient mice model, Levy *et al.* discovered that compared to wild-type mice, *Asc*$^{-/-}$ inflammasome-defective mice have decreased the abundance of taurine and increased the abundance of histamine and spermine, which correlate with suppression of inflammasome activity (*143*). Colonic explants demonstrated that taurine activates, while histamine and spermine suppress NLRP6 inflammasome activity. NLRP6 inflammasome could induce intestinal IL-18 production, which orchestrates colonic anti-microbial peptide (AMP) expression. The inflammasome-AMP axis, in turn, could regulate intestinal microbial composition, establishing a stable mutualism between host innate immune system and gut microbiota (*143*).

Long-chain *N*-acyl amides

The human microbiome is believed to encode functions that are important to human health; however, little is known about the specific effector molecules that commensals use to interact with the human host. Functional metagenomics provides a systematic way of surveying commensal DNA for genes that encode effector functions. Cohen *et al.* examined 3,000 Mb of metagenomics DNA cloned from fecal samples of three IBD patients and screened for effector gene clusters that activate NF-κB. One of the effector gene clusters come out of screening was recovered from all

three patient libraries, and it encodes for the production of *N*-acyl-3-hydroxypalmitoyl-glycine (commendamide), which resembles long-chain *N*-acyl amides that function as mammalian signaling molecules (*144*). They further showed that commendamide activates G2A/GPR132, a GPCR that is implicated in the modulation of immune cell function, autoimmunity, and atherosclerosis. In a follow-up study, Cohen *et al.* performed a bioinformatics analysis of human microbiome data to find bacterial enzymes that produce *N*-acyl amides. They identified 143 unique N-acyl synthase genes, which are enriched in gastrointestinal bacteria (*38*). Exogenous expression of these genes in *Escherichia coli* indicated that they produce *N*-acyl amides of six major families that are structurally similar to human GPCR ligands, including *N*-acyl serinol (**Fig. 1.3G**), which resembles the GPR119 ligand oleoylethanolamide. Profiling of some *N*-acyl amides against 240 human GPCRs revealed specific interactions, especially among receptors expressed in the gastrointestinal tract (*38*). Interestingly, mice colonized with bacteria expressing the synthase of *N*-acyl serinols showed reduced blood sugar levels in an oral glucose-tolerance test, consistent with the action on host GPR119 (*38*). These findings represent one of the first examples of microbe-derived small molecules affecting host physiology and highlight the use of functional metagenomics to identify critical microbial effectors.

Dipeptide Aldehydes

Similar to the discovery of microbiota-produced long-chain *N*-acyl amides, Guo *et al.* have used metagenomic bioinformatics to find a new family of nonribosomal peptide synthetase (NRPS) gene clusters from human microbiome (*145*). By expressing some of the most prevalent NRPS gene clusters (>90% samples), they found new molecules were produced, belonging to a family of pyrazinones and dihydropyrazinones, some of which were found with similar approaches from *Staphylococcus aureus* (*146*). Even though pyrazinones have been implicated as inhibitors of host proteases, they have poor *in vitro* affinity and activity. The authors found that these

pyrazinones were initially liberated from NRPSs as dipeptide aldehydes (**Fig. 1.3H**), which would undergo spontaneous cyclization and oxidation in the presence of oxygen (*145*). Indeed, under physiological pH and anaerobic condition, dipeptide aldehydes were stable enough to remain in active form, being highly potent, cell-permeable protease inhibitors. *N*-octanoyl-Met-Phe-H, an *N*-acylated dipeptide aldehyde that could not cyclize and be oxidized, is also a major product. Quantitative activity-based protein profiling (ABPP) showed that Phe-Phe-H, one dipeptide aldehyde representative, specifically inactivates cathepsin L, with minimal cross-reactivity to other cathepsins as well as other host proteins (*145*). These findings again demonstrate interesting interactions between host and microbiota and provide new lines of evidence on mutualism between commensal gut microbiota and their host.

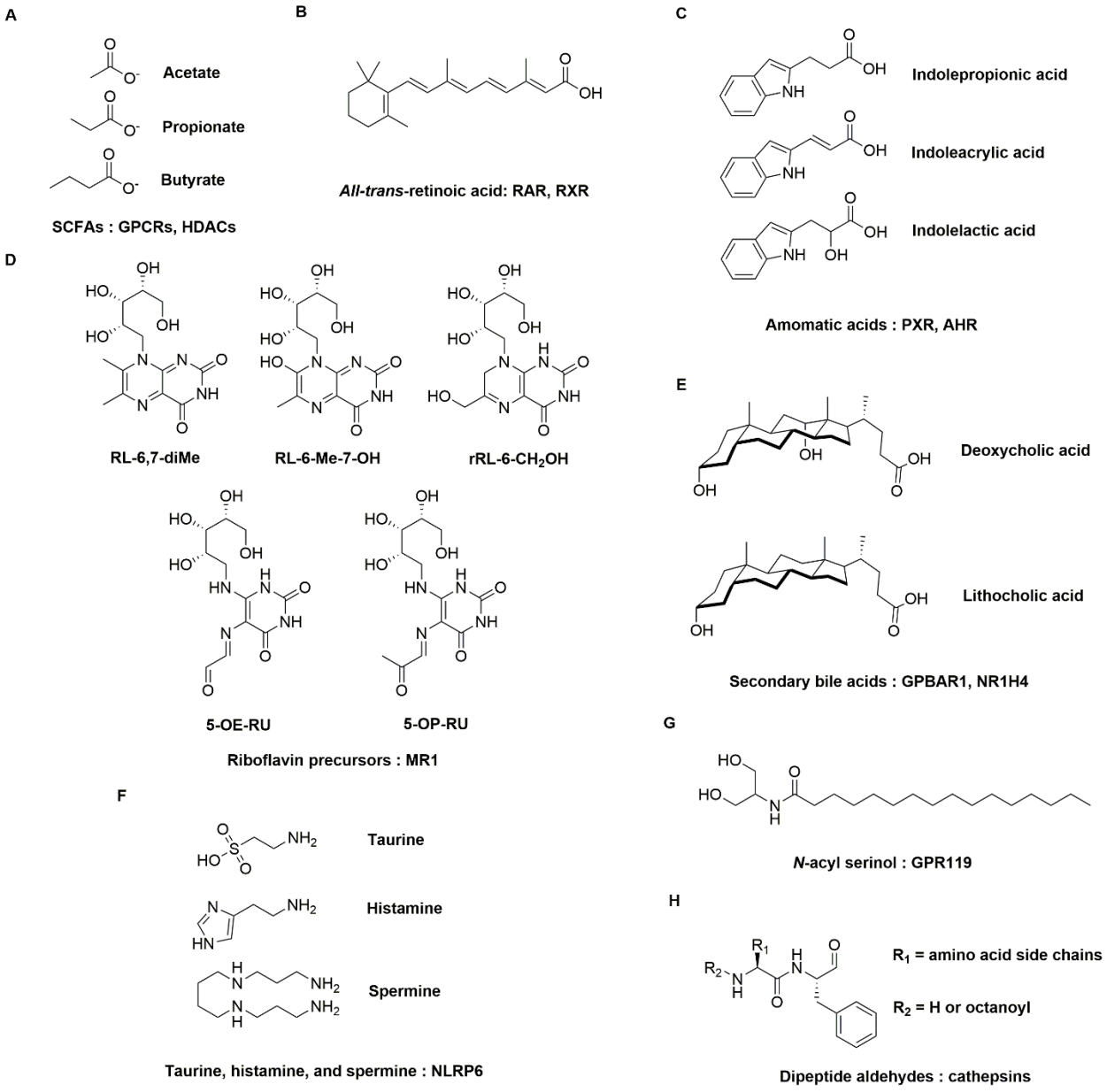


Figure 1.3. Microbiota-derived secondary metabolites. Their corresponding host receptors are listed after the colon. (**A**) Short-chain fatty acids : GPCRs, HDACs. (**B**) *All-trans*-retinoic acid : RAR, RXR. (**C**) Aromatic acids : PXR, AHR. (**D**) Riboflavin precursors : MR1. (**E**) Secondary bile acids : GPBAR1, NR1H4. (**F**) Taurine, histamine, and spermine: NLRP6. (**G**) *N*-acyl serinol : GPR119. (**H**) Dipeptide aldehydes : cathepsins.

Concluding Remarks

All the examples discussed above highlight the intricate interactions between host and its resident gut microbiota, resonating with a comment made more than two millennia ago by Hippocrates, "all disease begins in the gut". Bacterial density reaches a staggering 10^{11} organisms per gram in the colon, making this bacterial community the principal source of microbial metabolites in the human body. O'Hara and Shanahan proposed in 2005 that the gut microbiota represents a ‘microbial organ' (*147*), because it resides as a structural unit and produces metabolites at concentrations that promote health, analogous to an endocrine system. While the ‘microbial organ' concept is generally accepted (*148*), it is problematic to define what a ‘healthy microbial organ' should look like, because the comparison of the gut microbiota composition between different healthy individuals reveals minimal overlap on the species level (*149*). Recently, Byndloss and Baumler proposed that anaerobiosis is the hallmark of mammalian gut (*150*), thereby driving the composition of the microbial community towards a dominance of obligate anaerobes, which is critical for maintaining gut homeostasis (*68*). In contrast, dysbiosis, a state of microbial organ dysfunction, is characterized by failure to limit oxygen level in the gut and expansion of facultative anaerobes, including most enteric pathogens (*150*). This germ-organ theory instructs a shift from microbial community profiling towards understanding host-mediated control of microbial organ ecology. It calls for mechanistic follow-up studies aiming at understanding trophic networks, the influence of host physiology on the microbial ecosystem, and the role that microbiota-derived metabolites have in health and disease. This might be a move in the right direction as we stand at the threshold of a ‘second golden age of microbiology'.

Chapter 2

Microbiota-derived short-chain fatty acids inhibit *Salmonella* virulence through acylation on virulence regulator HilA

Abstract

The intestinal microbiota is important for host metabolism as well as immune development and is associated with human diseases. While the gut microflora is known to protect the host from invading pathogens, the underlying mechanism(s) have been elusive. Of note, short-chain fatty acids (SCFA) produced by commensal bacteria have been shown to inhibit key bacterial virulence pathways, such as type III secretion system (T3SS) in Gram-negative enteric bacterial pathogens, but the molecular mechanism(s) are still under investigation. The major limitation in understanding the functions of SCFA has been identifying the direct biochemical targets of these microbial metabolites. We applied bio-orthogonal alkyne-fatty acid reporters to directly identify the biochemical targets of SCFA in *Salmonella* Typhimurium, a Gram-negative enteric pathogen responsible gastroenteritis in humans. With in-gel fluorescence profiling, click chemistry-mediated enrichment and mass spectrometry-based proteomics, I found that exogenous short-chain fatty acids can inhibit T3SS of *Salmonella* and covalently modify key virulence transcriptional regulator HilA. The modification was not susceptible to enzymatic acylation or deacylation mediated by Pat and CobB. Via amber suppression technology and CRISPR-Cas9 genome editing technique, bio-orthogonal stable lysine acylation mimic was incorporated site-specifically to endogenously expressed HilA in *Salmonella*, and it revealed that fatty-acylation on K90, K324, and K456 of HilA impaired its function to activate virulence gene expression, and decreased corresponding *Salmonella* mutants' infectivity to HeLa cells. In particular, I showed that fatty-acylation on K90 of HilA impaired its DNA-binding activity, decreasing *Salmonella* invasion in mice. These studies are crucial for elucidating fundamental mechanisms of microbiota-mediated resistance on bacterial virulence and should facilitate the development of new anti-infectives to prevent or treat bacterial infections.

Introduction

The intestinal microbiota of mammals is composed of 10^{13}-10^{14} cells, which outnumbers total cell numbers in the host (*151*). It consists of about 1000 different bacterial species that regulate host metabolism (*152*), resistance to gut pathogens (*153*), as well as immune system development and homeostasis (*154*). Dysregulation of the microbiota is associated with a variety of host immune disorders, including inflammatory bowel disease (IBD) (*155*), allergy (*156*), and diabetes (*157*). Fecal transplant from healthy donors to patients has achieved satisfactory results in treating recurrent *C. difficile* infection and other intestinal diseases (*158*). The mechanisms by which commensal bacteria modulate host immune system and protect against enteric pathogens have been difficult to elucidate due to the complex interactions between the host and its microbiota and enteric pathogens.

S. enterica is a Gram-negative intracellular pathogen that causes gastroenteritis and typhoid fever worldwide (*159*). Once ingested, *Salmonella* traverses the gut to the small intestine, where a set of virulence genes are activated to promote gut inflammation as well as invasion of the intestinal epithelia, allowing *Salmonella* to replicate and disseminate throughout the host (*160*). Systemic infection associated with typhoid fever is mediated by two *Salmonella* pathogenicity islands (SPI) that encode Type 3 Secretion Systems (T3SS) for bacterial invasion, dissemination, and replication inside host cells (*159*, *160*). Specifically, SPI-1 is important for *Salmonella* invasion, while SPI-2 is crucial for *Salmonella* replication in host cells (*161*). Genetic and biochemical studies have demonstrated that T3SS form multi-protein complexes to inject a variety of bacterial protein effectors into host cells for *Salmonella* pathogenesis (*159*, *160*). These *Salmonella* virulence programs are regulated by a variety of environmental factors, including the host microbiota (*69*, *162*).

Of note, short-chain fatty acids (SCFA) produced by commensal bacteria, including propionate and butyrate, have been shown to inhibit T3SS in *Salmonella enterica* serovar Typhimurium (Stm) (*69–71*). SCFA have been implicated in inhibiting *Salmonella* virulence through transcriptional (*69*) and post-translational regulation (*71*), but the molecular mechanism(s) are not well studied. Understanding the effects of short-chain fatty acids on Stm will help elucidate the interactions between host, commensal bacteria, and enteric pathogen, and guide new treatment and prevention to enteric bacterial pathogens.

Over the last decade, bio-orthogonal chemistry has emerged as a powerful tool to rapidly investigate biological activities with minimal perturbation to the biological system. Copper-catalyzed azide–alkyne cycloaddition (CuAAC), or more commonly “click chemistry”, is the most widely used bio-orthogonal reaction, in which Cu(I) cation catalyzes cycloaddition between azide and alkyne, forming stable covalent bonds between two molecules attached with azido or alkynyl group. We have developed bio-orthogonal acylation reporters over the past few years that have enabled rapid and robust detection of fatty-acylation of proteins in both bacteria and mammalian cells. LCFA reporters have been applied to detect reversible *S*-palmitoylation in mammalian cells (*163*) and to detect lipoproteins in bacteria (*164*). SCFA reporters are applied to probe acetylation in mammalian cells (*165*). Compared to traditional fatty-acylation detection methods, these fatty-acylation reporters facilitate rapid detection compared to radioactive isotope fatty acids labeling, which may take up to weeks for detection, and robust reproducibility as compared to detection by fatty-acylation-specific antibodies. In this work, I set out to apply SCFA chemical reporter to study how SCFA inhibit the virulence of Stm, specifically the post-translational modification (PTM) of virulence-related factors, and directly address the functional consequence of short-chain fatty-acylation on one of these proteins, HilA.

During the last 15 years, Amber Suppression Technology has been developed to enable incorporation of Unnatural Amino Acids (UAA) site-specifically in proteins of interest *in vivo*. It harnesses an orthogonal pair of $tRNA^{Pyl}$ and aminoacyl-tRNA synthetase derived from *Methanosarcina* species, which can recognize unnatural amino acid (UAA) and incorporate it onto protein-encoding mRNA bearing an amber codon (TAG), a stop codon if naturally read by the translation system. *In vitro* directed evolution has created a variety of synthetase mutants that can incorporate a broad range of functionalized UAAs. Amber Suppression Technology, coupled with CRISPR-Cas9 genome editing technique, allows us to install functional and stable lysine-acylation mimic site-specifically in endogenously expressed HilA in Stm.

Genetic tools in *Salmonella* have been developed over the years and greatly facilitate the genetic research in *Salmonella*. In particular, genome editing techniques based on Lambda Red Recombination System (*166*) have been widely used for genetic knock-out and knock-in in *Salmonella* (*167*). Nevertheless, these editing approaches rely on insertion of an antibiotic resistance cassette in the genome for positive selection. Even if the cassette can be eliminated with FLP recombinase, it inevitably leaves a “scar” sequence on the genome, potentially causing unintended side effects on the cell (*166*). Recently, CRISPR-Cas9 genome editing technique has been widely adopted in scientific community for precise and scarless genome editing in various organisms, from bacteria (*168*) to mammals (*169–171*). Previously it was reported that CRISPR-Cas9 genome editing system could be applied in *Escherichia coli* with high editing efficiency (*168*). Cas9 protein and guide-RNA (sgRNA) were introduced to *E. coli* in two separate plasmids, pCas9 and pCRISPR, along with transformation of editing template and expression of Lambda Red recombinase. Given the similarity between *E. coli* and Stm, we tried to use the same protocol in Stm for genome editing.

In this chapter, I demonstrate that HilA, a key virulence regulator of SPI-1 T3SS in Stm, is short-chain fatty-acylated at five lysine residues *in vivo*. I showed that Lysine Butyrylation Mimic (KBM) UAA is a faithful and stable mimic of lysine butyrylation, and incorporating KBM at different sites of HilA results in varied functional consequence. In particular, I provided evidence that HilA K90 acylation affects its DNA-binding ability, decreases SPI-1 gene expression, and impairs *Salmonella* infectivity in HeLa cells and in mice.

Results

Short-chain fatty acids inhibit *Salmonella* virulence

In mammalian gut, concentrations of SCFA varies, of which median is about 10 mM. I observed that butyrate (**Fig. 2.1A**), a major SCFA present in the gut, did not inhibit Stm growth *in vitro* at concentration of 10 mM (**Fig. 2.1B**). However, 10 mM butyrate decreased Stm invasion ability to HeLa cells (**Fig. 2.1C**), and inhibited expression level of Stm SPI-1 effector genes such as *sipA,* as shown by quantitative reverse–transcription PCR (qRT-PCR) (**Fig. 2.1D**). This suggest that butyrate inhibits Stm virulence through antagonizing transcription and expression of Stm SPI-1 effector proteins. SCFA are metabolized in Stm and may result in the production of a panel of highly reactive acyl intermediates, including acyl-CoA and acyl-phosphate (*172*), which may lead to an extensive increase of post-translational modifications on a variety of proteins, leading to altered regulatory functions (*173*) and enzymatic activities (*174*). Therefore, we hypothesized that SCFA may inhibit Stm virulence through acylation of key virulence regulators, which might affect their normal functions. Similar regulatory mechanisms have been reported in other pathogens (*175*).

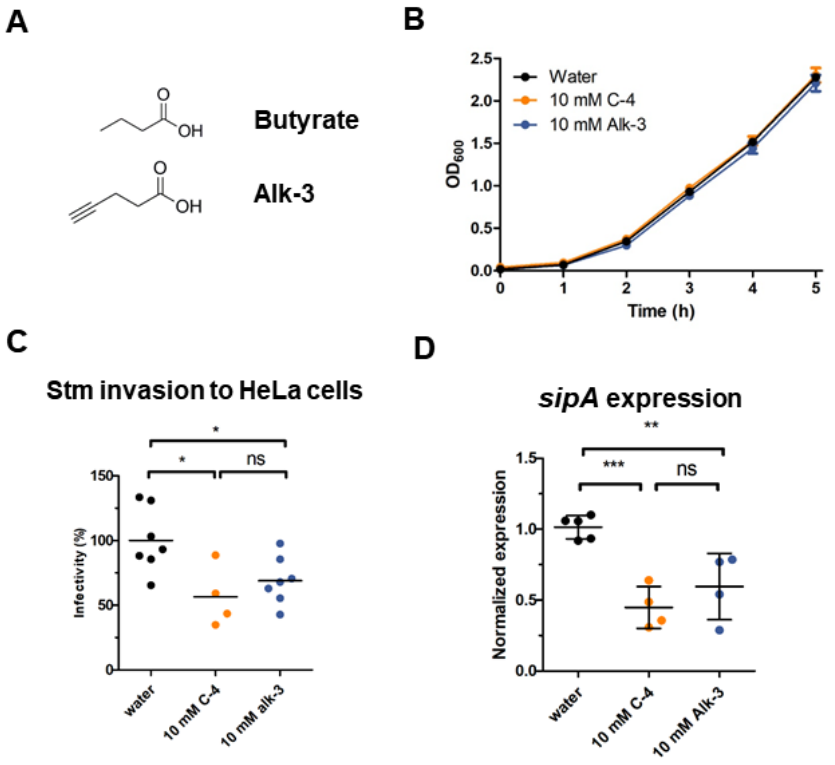


Figure 2.1. Short-chain fatty acids inhibit *Salmonella* virulence. (**A**) Structure of butyrate and Alk-3. (**B**) Growth curve of Stm 14028 WT in SPI-1 inducing LB in the presence of 10 mM C-4, 10 mM mM Alk-3, or same volume of water. (**C**) Gentamicin protection assay of Stm incubated with or without 10 mM Alk-3 infecting HeLa cells at MOI=10. (**D**) Expression of SPI-1 gene *sipA* was measured by qRT-PCR from Stm with or without 10 mM alk-3 incubation.

Proteomic analysis of acylated proteins in *Salmonella*

To identify acylated proteins in *Salmonella*, I employed SCFA chemical reporter alk-3 (pentynoate) (**Fig. 2.1A**). Alk-3 behaved similarly to its natural counterpart butyrate, as it retained the ability to inhibit Stm invasion to HeLa cells (**Fig. 2.1C**) and inhibited expression level of *sipA* (**Fig. 2.1D**) at 10 mM, but did not inhibit Stm growth (**Fig. 2.1B**). This suggests it may have similar mode of action as butyrate to inhibit Stm virulence. To visualize acylated proteins with alk-3, I incubated Stm culture with 10 mM alk-3, and harvested total cell lysates for CuAAC reaction with azide-functionalized Rhodamine (az-Rho, **Fig. 2.2A**). SDS-PAGE followed by in-gel fluorescence

scanning demonstrated that alk-3 metabolically labeled a diverse repertoire of proteins in Stm (**Fig. 2.2C**). To identify these acylated proteins, I performed Label-Free Quantitative (LFQ) Proteomics analysis on Stm proteome with or without alk-3 labeling. Stm cell lysates were reacted with an azido-biotin affinity tag (az-biotin, **Fig. 2.2B**). Alk-3 labeled proteins were enriched by streptavidin beads, and digested by Trypsin/LysC mix on-bead. Digested peptides were processed and identified by Liquid Chromatography–tandem Mass Spectrometry (LC-MS/MS). The resulting spectrum were searched with MaxQuant (*176*) and quantified with Perseus (*177*). I identified 'hits' as proteins enriched for more than 2-fold in alk-3 samples and *P*-value less than 0.05. With this approach, I selectively identified 56 proteins labeled by alk-3 compared to control samples (**Fig. 2.2D**, **Appendix 2.1**). Of these proteins, 34 of the hits (61%) were categorized as metabolism-related proteins. However, only 6 proteins in these hits (11%) were directly related to Stm virulence. Notably, HilA, a master transcriptional activator of Stm SPI-1 virulence (*178*, *179*), was one of the most prominent hits in the data set.

HilA is acylated in *Salmonella*

HilA is a master transcription regulator of SPI-1, belonging to OmpR/PhoB family of regulatory proteins. HilA activates the *inv*/*spa* and *prg* operons, encoding components of the T3SS apparatus (*180*, *181*), and the *sic*/*sip* operon, encoding a chaperone and secreted proteins (*182*). HilA is essential for *Salmonella* virulence, as *hilA* gene deletion in *Salmonella* abolishes its secretion of SPI-1 effectors (*183*) (**Fig. 2.3A**) and its infectivity to cells (*184*) (**Fig. 2.3B**). HilA protein is predicted by NCBI Conserved Domains Search to contain an N-terminus DNA-binding domain (DBD) and a Tetratricopeptide Repeat (TPR) domain near its C-terminus, but no atomic structure of HilA is available to date (**Fig. 2.3C**). It is reported that HilA is acetylated at 5 different lysine residues (*185*), namely K90, K231, K324, K456, and K533. Robetta Full-chain Protein Structure Prediction Server (*186*) predicted that all of these 5 lysine residues located at the

surface of the protein, suggesting that they may be accessible for acylation (**Fig. 2.3D**). Moreover, K57 was predicted to be critical for protein-DNA interaction between HilA and its DNA partner. Therefore, I used HilA K57 mutant as a loss-of-function control. I also chose K527 as a neutral control, as it was not implicated to have any functional role.

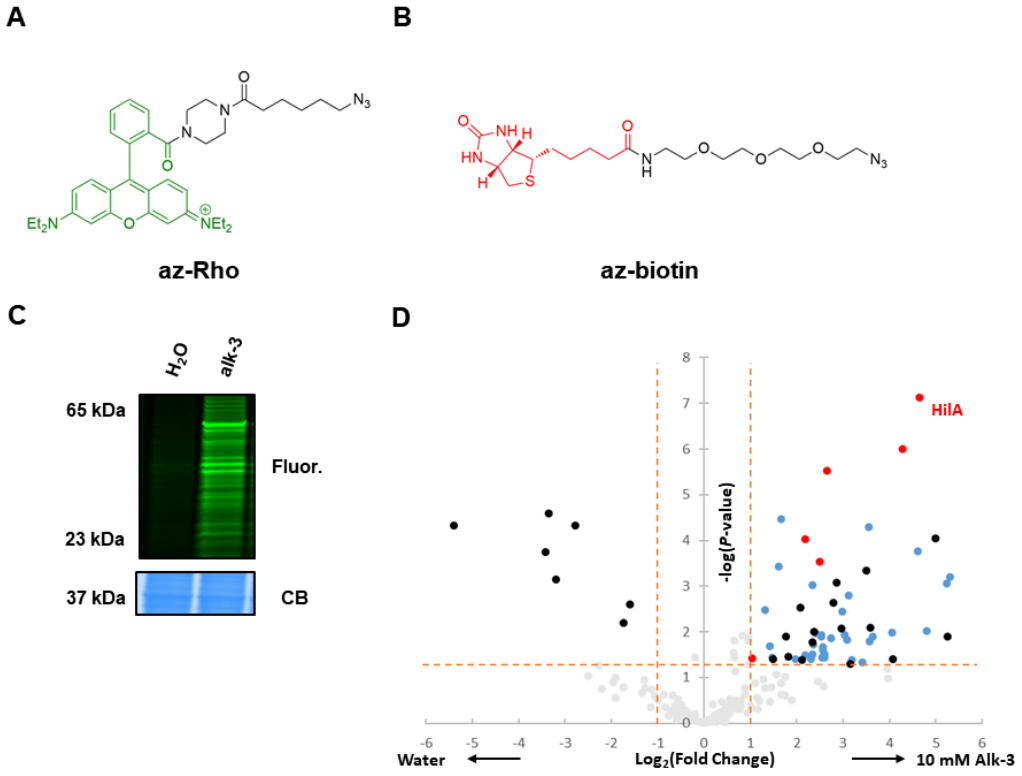


Figure 2.2. Proteomic analysis of acylated proteins in *Salmonella*. (**A**) Molecular structure of azido-Rhodamine (az-Rho). (**B**) Molecular structure of azido-PEG$_3$-biotin (az-biotin). (**C**) *Salmonella* cell lysates were reacted with az-Rho by CuACC, and proteins were separated by SDS-PAGE for visualization by fluorescence gel scanning (top). Coomassie blue staining demonstrates comparable loading (bottom). (**D**) *Salmonella* cell lysates were reacted with az-biotin by CuAAC for the enrichment of alk-3–labeled proteins with streptavidin beads and identification by mass spectrometry. LFQ proteomic analysis Identified proteins that were enriched by alk-3 (top right corner), which were colored according to their annotated biological function. Blue, metabolic enzymes. Red, SPI-1 proteins. Black, other proteins.

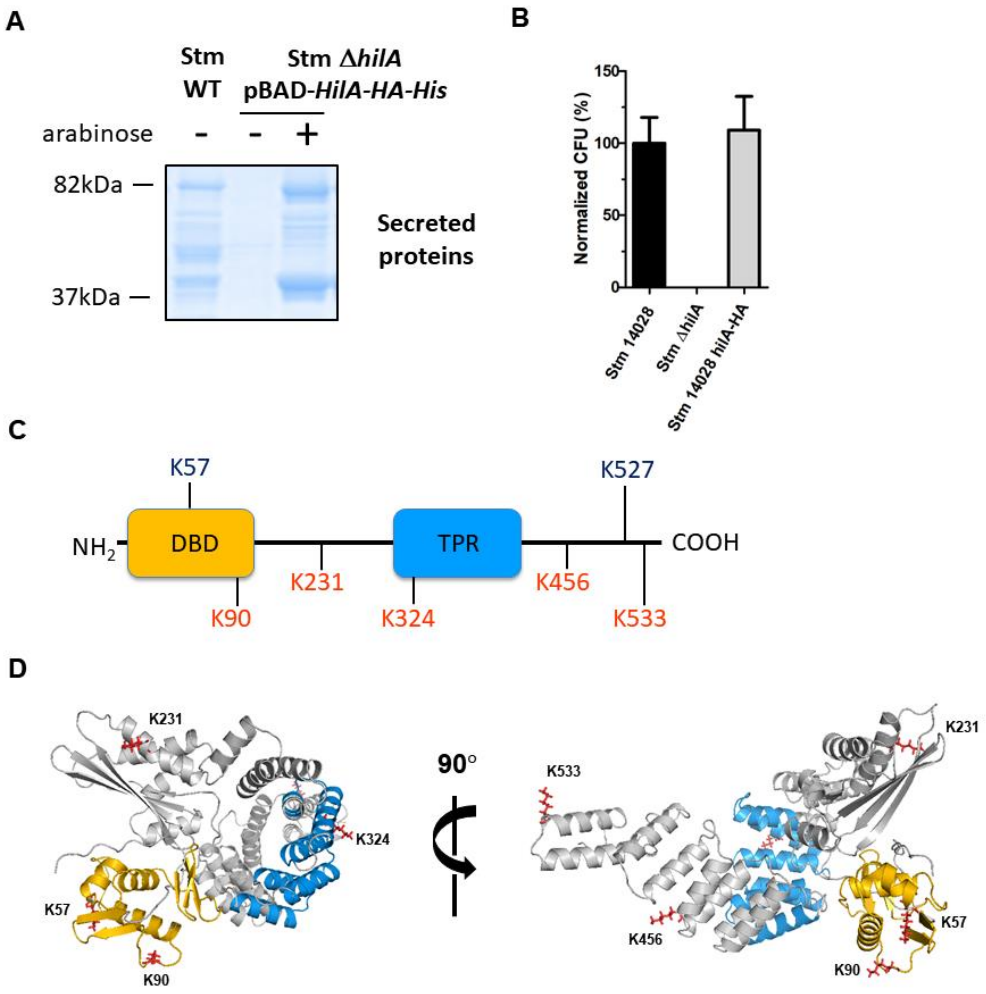
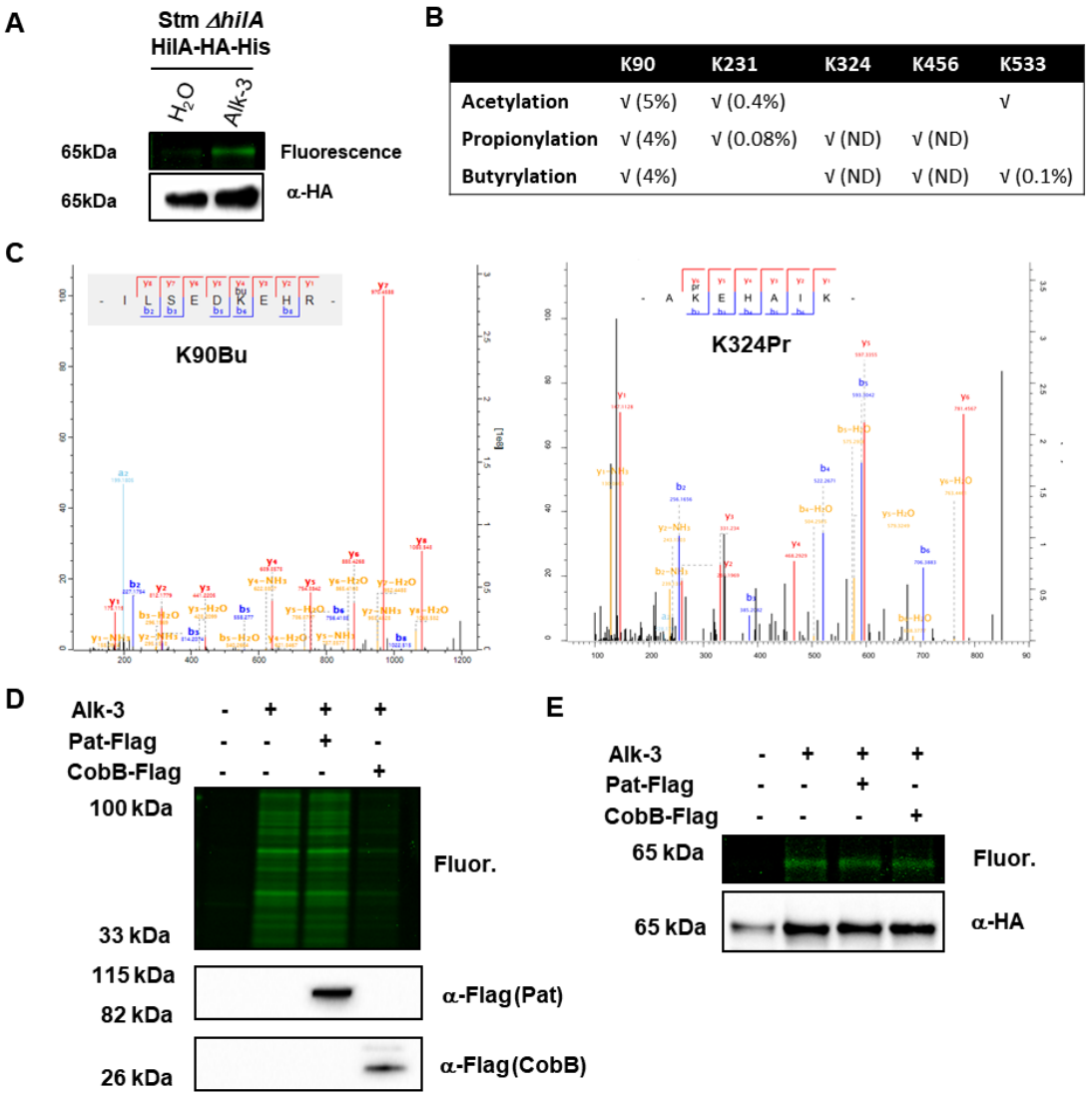


Figure 2.3. HilA is essential for *Salmonella* virulence. (**A**) Secretion assay of Stm WT, Stm *hilA*, and Stm *hilA* overexpressing HilA-HA-His. The secreted SPI-1 effector proteins were precipitated from supernatant with trichloroacetic acid and run on SDS-PAGE, followed by Coomassie Blue staining. (**B**) Gentamicin protection assay of Stm, Stm *hilA*, and Stm HilA-HA infecting HeLa cells at MOI=10. (**C**) Predicted domains of HilA and lysine residues that are reported to be acetylated (orange), as well as K57 and K527 chosen to serve as controls (navy). DBD, DNA-binding domain. TPR, tetratricopeptide repeat domain. (**D**) Predicted structure of HilA by Robetta Full-chain Protein Structure Prediction Server. DBD is labeled in yellow, TPR domain is labeled in blue. Five acylated lysine residues and K57 are labeled in red.

To study HilA in more detail, I subcloned HilA into pBAD plasmid and appended HA tag and His6 tag to the C-terminus of HilA. HilA-HA-His6 was well expressed upon arabinose induction in Stm, and it functionally rescued secretion deficiency in Stm *hilA* mutant (**Fig. 2.3A**). To confirm that HilA was indeed acylated, Stm overexpressing HilA-HA-His6 was grown in medium with alk-3, and CuAAC in-gel fluorescence scanning of cell lysates demonstrated that HilA was labeled by alk-3 (**Fig. 2.4A**). Moreover, I purified HilA-HA-His6 protein from Stm incubated with either propionate or butyrate, and enriched acylated fraction with anti-propionyllysine or anti-butyryllysine antibody. LC-MS/MS analysis of digested peptides from these fractions identified 5 lysine residues out of 34 in HilA were acetylated, propionylated, or butyrylated (**Fig. 2.4B, 2.4C, Appendix 2.2**), consistent with previous report (*185*). Modification occupancies were estimated based on area of modified and unmodified peptides in LC-MS/MS (**Fig. 2.4B**). K324 and K456 unmodified peptides, when fully digested, are too short to be detected, therefore the modification occupancy were not determined.

Acylation is a dynamic process, which may involve both acyltransferases and deacylases 'writing' and 'erasing' acylation on proteins. To investigate whether acylation on HilA is regulated by acyltransferase and deacylase in *Salmonella*, HilA-HA-His6 was co-expressed with Gcn5-like Protein Acyltransferase (Pat), or the only known protein deacylase in *Salmonella*, CobB. CuAAC in-gel fluorescence scanning showed that while CobB could decrease alk-3 labeling level on *Salmonella* proteome (**Fig. 2.4D**), alk-3 labeling level on HilA was not affected by either Pat or CobB (**Fig. 2.4E**).



	K90	K231	K324	K456	K533
Acetylation	√ (5%)	√ (0.4%)			√
Propionylation	√ (4%)	√ (0.08%)	√ (ND)	√ (ND)	
Butyrylation	√ (4%)		√ (ND)	√ (ND)	√ (0.1%)

Figure 2.4. HilA is acylated in *Salmonella*. (**A**) Stm overexpressing HilA-HA-His were incubated with water or 10 mM Alk-3 before CuAAC with az-Rho, SDS-PAGE in-gel fluorescence scanning (top), and immunoblotting (bottom). (**B**) Table of lysine residues (column) and short-chain acylation types (row) identified by LC-MS/MS (tick), and their estimated modification occupancy in parentheses. ND, not determined. (**C**) MS/MS spectrum of HilA K90 butyrylated peptide and K324 propionylated peptide as representative MS/MS spectrums of acylated HilA peptide. (**D**)(**E**) *Salmonella* overexpressing HilA-HA-His, as well as Pat-Flag or CobB-Flag, were incubated with or without alk-3 during overexpression. Total cell lysates (**D**) and anti-HA immunoprecipitated samples (**E**) were analyzed with SDS-PAGE in-gel fluorescence scanning (top), and anti-Flag (**D**) or anti-HA (**E**) immunoblotting (bottom).

Installing stable acylation mimic on HilA in *Salmonella*

Next I set out to pinpoint the effects of lysine fatty-acylation on the function of HilA. Conventional approaches include mutating lysine residues to glutamines as mimicry to lysine acylation. However, glutamine bears a polar terminal acylamine and is relatively small, which is significantly different from the relatively nonpolar internal secondary amide and bulky steric hindrance in *N*-acyllysine. To achieve better mimicry of lysine acylation on HilA protein, we decided to apply amber suppression technology to site-specifically engineer HilA.

First I tested incorporation of a panel of acyllysines, namely acetyllysine (AcK), propionyllysine (PrK), and butyryllysine (BuK) (**Fig. 2.5A**), which have been reported to be incorporated into proteins in bacteria (*187*, *188*). I found that in *Salmonella*, AcK and PrK were not detected to be incorporated at K90 in HilA (**Fig. 2.5B**). BuK was poorly incorporated, and could be barely detected by anti-HA and pan anti-BuK immunoblotting (**Fig. 2.5B**). I also raised a custom site-specific polyclonal antibody against HilA butyrylated K90 (anti-HilAK90Bu), and HilA-K90BuK was poorly detected as well (**Fig. 2.5B**).

To incorporate an unnatural amino acid that can be efficiently incorporated into proteins in *Salmonella*, and faithfully mimics lysine acylation at the same time, we decided to use a lysine analog UAA, Lysine Butyrylation Mimic (KBM, *Nε*-pent-4-ynyloxy-carbonyl-*L*-Lysine) (**Fig. 2.5A**) (*189*). It bears carbamate at -*N* and an alkynyl group at the terminus. This unique structure serves as lysine acylation mimic at -*N* on lysine residues, and protects itself from being deacylated by endogenous deacylases, while at the same time enables robust detection via CuAAC. I demonstrated that KBM could be readily incorporated via *Methanosarcina barkeri* pyrrolysine-tRNA synthetase ASF mutant and $tRNA^{CUA}$, and into over-expressed HilA at K90 at similar levels compared to wild type, which could be detected by immunoblotting (**Fig. 2.5B**) and

in-gel fluorescence scanning (**Fig. 2.5C**). Importantly, HilAK90KBM protein could be immunoblotted by both pan anti-BuK antibody and anti-HilAK90Bu antibody (**Fig. 2.5B**), but not by anti-PrK or anti-AcK antibody, suggesting that KBM faithfully mimics lysine butyrylation on proteins.

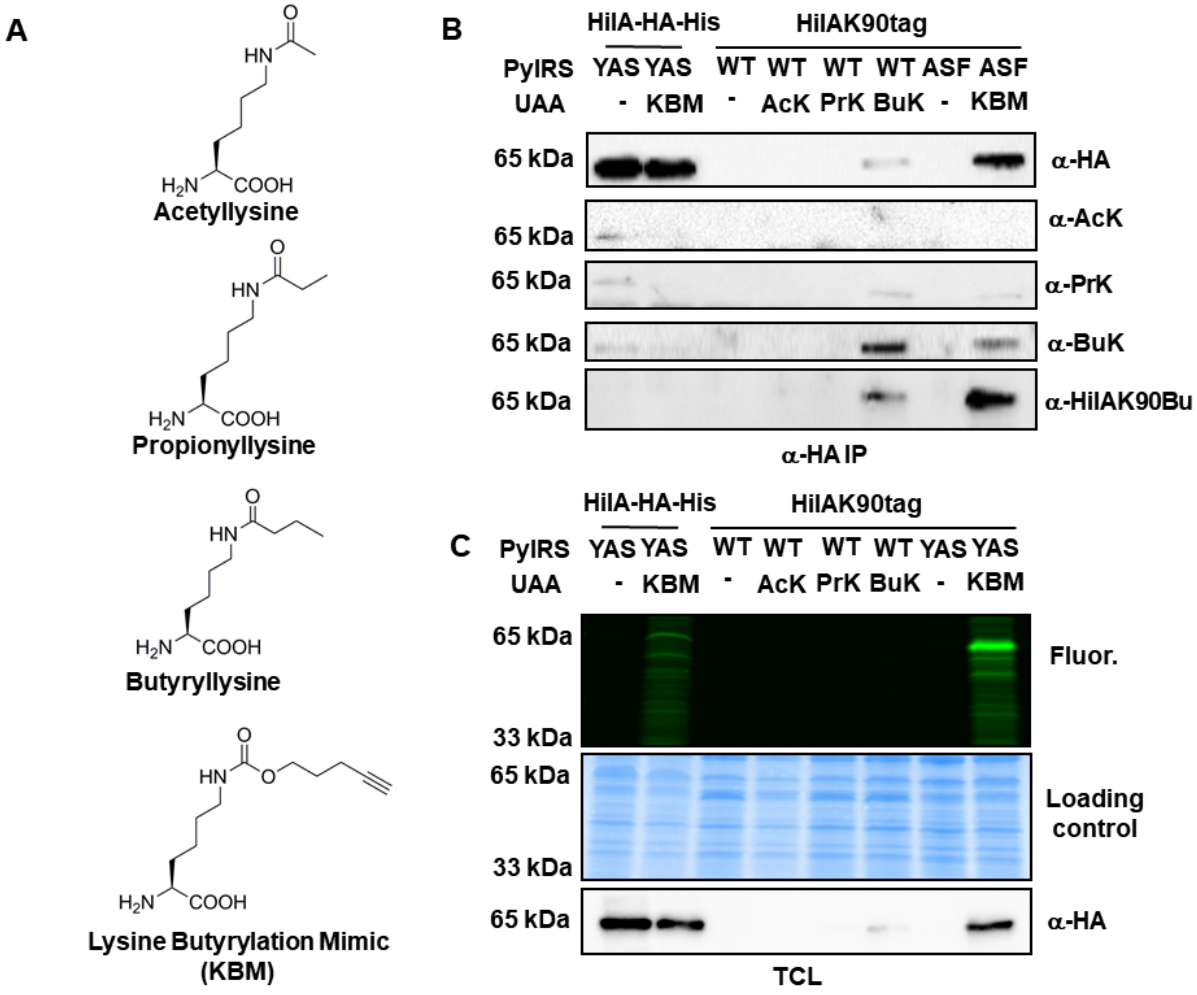


Figure 2.5. Installing Stable acylation mimic on HilA in *Salmonella*. (**A**) Molecular structure of acetyllysine (AcK), propionyllysine (PrK), butyryllysine (BuK), and lysine butyrylation mimic (KBM). (**B**) (**C**) *Salmonella* with HilA-HA-His or HilAK90tag-HA-His construct, as well as PylRS-WT or –ASF mutant plasmid, were incubated with or without AcK, PrK, BuK, or KBM during overexpression. Total cell lysates (**B**) were analyzed with SDS-PAGE in-gel fluorescence scanning (top), Coomassie Blue staining (middle), and anti-HA immunoblotting (bottom). Anti-HA immunoprecipitated samples (**C**) were immunoblotted for HA, AcK, PrK, BuK, or HilAK90Bu.

Optimizing CRISPR-Cas9 Genome Editing in *Salmonella*

To assay effects of acylation on HilA in a more physiologically relevant setting, we sought to incorporate KBM into endogenously expressed HilA protein. Therefore, we would need site-specific amber codon mutations in the *hilA* gene in *Salmonella* genome, with no additional mutations to complicate the phenotype outcome. We decided to use CRISPR-Cas9 genome editing in *Salmonella*, a technique that could generate scarless and precise mutations in genome.

I tried to directly employ CRISPR-Cas9 genome editing protocol in *E. coli* (*168*) to Stm. Briefly, sgRNA sequence is subcloned into pCRISPR, and Stm expressing Lambda Red Recombination System (pKD46 plasmid) is transformed with pCas9, pCRISPR-sgRNA, and a double-stranded DNA (dsDNA) editing template. Bacteria is selected on agar plates for colonies containing both pCas9 and pCRISPR-sgRNA. However, I found that this protocol resulted in very high false-positive rate and editing efficiency was extremely low, with estimation that less than 5% of colonies on the plate were successfully edited.

To optimize the CRISPR-Cas9 genome editing system in Stm, we first deleted a guide RNA sequence with unknown function in pCas9 to minimize off-target effect of Cas9. This new plasmid, pWJ297, contains coding sequence of *Streptococcus pyogenes* Cas9 (SpCas9), Chloramphenicol resistance gene, and direct repeats flanking two BsaI restriction sites for subcloning of sgRNA sequence. SpCas9 requires Protospacer Adjacent Motif (PAM), namely NGG nucleotide sequence (or CCN on complementary strand), to correctly pair sgRNA in the genome and create dsDNA break (*190*). Ideally, sequence to be modified should be within 10 base-pairs of PAM for sufficient discrimination between non-perfect-match (i.e. edited) and perfect-match (i.e. unedited) sequences by Cas9 (*168*). However, this was not feasible in many cases, especially when targeted sequence was in an AT-rich region. To increase editing efficiency

in these circumstances, I introduced synonymous co-mutation of nearest PAM together with targeted sequence, and used sgRNA specific for that PAM to guide Cas9 (**Fig. 2.6A**). As an example, I set out to mutate K90 in HilA, and synonymous PAM co-mutation strategy increased editing efficiency from 0/11 to 1/15 (**Fig. 2.6B**).

After screening the colonies that were selected for pWJ297 and pCRISPR, I found that most of them were false-positive, suggesting that CRISPR-Cas9 system did not impose selection pressure high enough against unedited clones. Interestingly, many of these false-positive colonies could not grow when purified on a new agar plate selecting for pWJ297 and pCRISPR. I surmised that Stm has an intrinsic suppressive mechanism to inactivate exogenous CRISPR-Cas9 system, and this inactivation could be abrogated when Stm is in rapid growth state. Therefore, I collected all colonies from the plate of first-round selection, suspended them in liquid medium LB with corresponding antibiotics, and grew them to early stationary phase. The resulting culture was plated again on selection agar plate. This liquid selection protocol further increased frequency of successfully edited clones (**Fig. 2.6C**).

After successful genome editing, pWJ297 and pCRISPR need to be cured to eliminate any unnecessary complications. While pWJ297 could be efficiently cured by one round of purification on plain agar plate, pCRISPR was very resistant to curing, taking up to one month of serial purification, which is presumably due to its small size and high copy number. To facilitate the entire editing process, I incorporated sgRNA sequence directly at pWJ297 BsaI cloning site, sparing the usage of pCRISPR. The editing efficiency of pWJ297-sgRNA is similar to pWJ297 + pCRISPR combination (data not shown), while the former was cured in about 1 day. This modification greatly speeds up the whole CRISPR-Cas9 genome editing process.

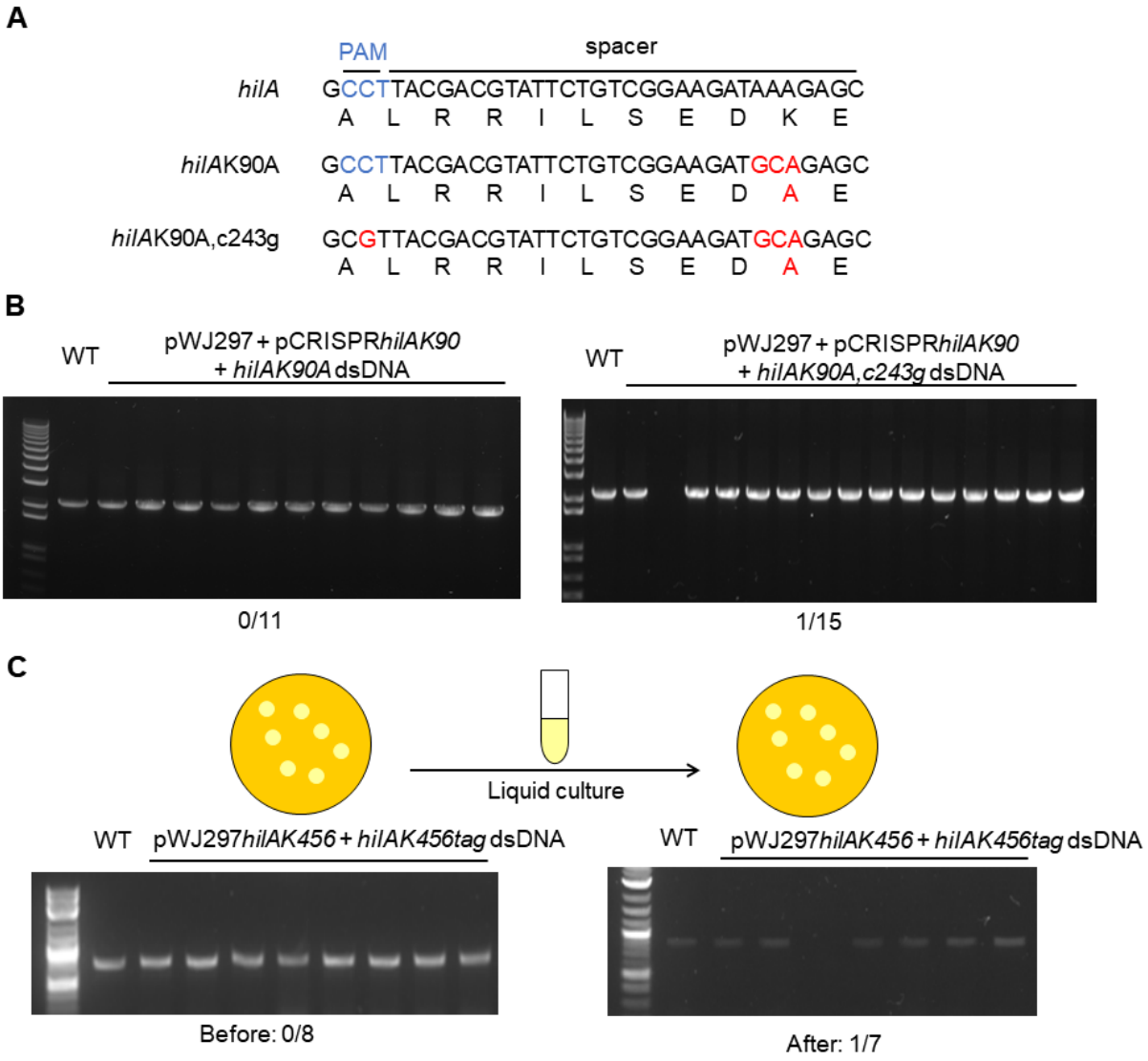


Figure 2.6. Optimizing CRISPR-Cas9 Genome Editing in *Salmonella* with co-mutation and liquid selection. (**A**) Scheme of synonymous co-mutation of nearest PAM. In addition to K90A (AAA to GCA) mutation, c243g synonymous mutation eliminates PAM sequence in blue. (**B**) PCR screening of Stm genome editing of *hilA*K90A with (right) or without (left) PAM co-mutation. Absence of PCR amplification band indicates successful editing. (**C**) PCR screening of Stm genome editing of *hilA*K456tag before (left) or after (right) additional liquid selection step. Absence of PCR amplification band indicates successful editing.

The Lambda Red Recombination system encodes three proteins: Exo, Gam, and Beta. Exo and Gam, required for dsDNA recombination, are not required for single-stranded DNA (ssDNA) reombination, in which case only Beta is required (*191*). SsDNA recombination efficiency is much

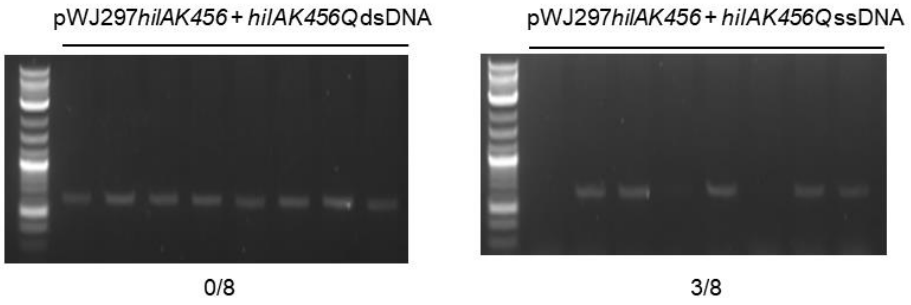
higher than dsDNA in *E. coli* because ssDNA has higher transformation efficiency (*191*). Therefore, I decided to use ssDNA template when editing small pieces of genome, for example, site-specific mutation of single codons on endogenous genes. CRISPR-Cas9 genome editing with ssDNA provided higher editing efficiency compared to that with long dsDNA template that has the same mutation site (**Fig. 2.7A**). I observed no significant difference of editing efficiency among ssDNA templates that matches either the same strand or the complementary strand of sgRNA, nor between templates matching either the leading or the lagging strand of genome (**Fig. 2.7B**).

To confirm that CRISPR-Cas9 genome editing in *Salmonella* did not introduce additional off-target mutations, I performed Next-Generation whole-genome sequencing on Stm 14028S WT, Stm 14028S HilA-HA, and Stm 14028S HilAK90tag-HA strains (**Table 2.1**). Compared to WT, HilA-HA and HilAK90tag-HA did not contain additional mutations other than intended insertion and substitution that could seemingly have functional impact. Therefore, CRISPR-Cas9 genome editing was specific in introducing intended mutations in *Salmonella*.

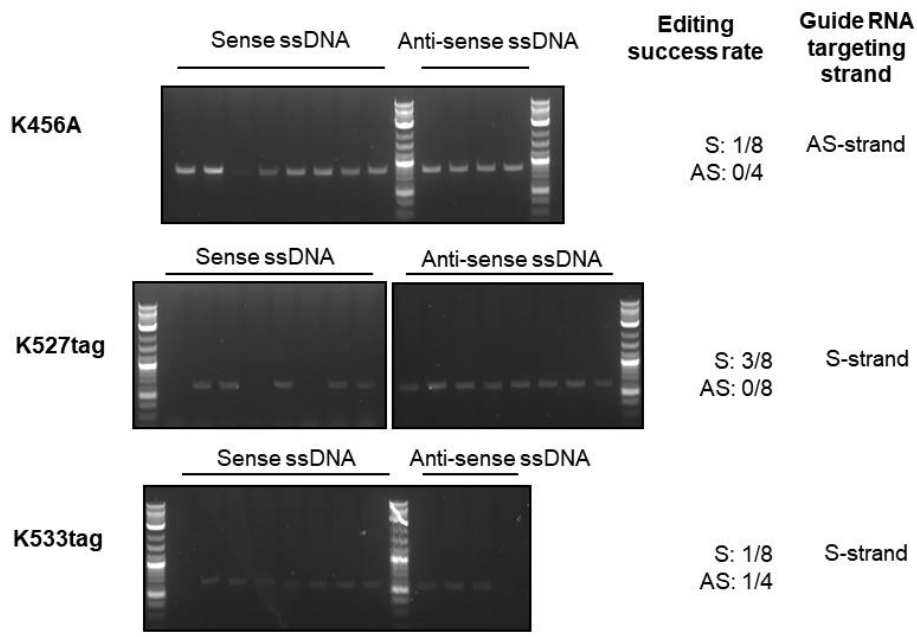
To summarize, I have optimized CRISPR-Cas9 genome editing protocol to facilitate the editing process in *Salmonella* (**Fig. 2.7C**). The entire procedure takes 3–4 days for completion, significantly shorter than previous protocol, and creates site-specific and scarless edited *Salmonella* mutants.

Figure 2.7. Optimizing CRISPR-Cas9 Genome Editing in *Salmonella*. (**A**) PCR screening of Stm genome editing of *hilA*K456Q with dsDNA (left) or ssDNA (right) as editing template. Absence of PCR amplification band indicates successful editing. (**B**) PCR screening of Stm genome editing of *hilA*K456A (top), K527tag (middle), or K533tag (bottom) with sense or anti-sense ssDNA as editing template. Absence of PCR amplification band indicates successful editing. Editing success rates as well as targeting strands of gRNA are listed on the right. (**C**) Scheme of optimized CRISPR-Cas9 genome editing in *Salmonella*. *Salmonella* is first transformed with pKD46, then is transformed with pWJ297-spacer and ssDNA template. Colonies from the plates undergo liquid selection before being plated again. The resulting edited colonies are purified on non-selection agar plates to cure pWJ297 plasmid.

A


pWJ297hilAK456 + hilAK456Q dsDNA
pWJ297hilAK456 + hilAK456Q ssDNA
0/8
3/8

B


Sense ssDNA
Anti-sense ssDNA
Editing success rate
Guide RNA targeting strand
K456A
S: 1/8
AS: 0/4
AS-strand
Sense ssDNA
Anti-sense ssDNA
K527tag
S: 3/8
AS: 0/8
S-strand
Sense ssDNA
Anti-sense ssDNA
K533tag
S: 1/8
AS: 1/4
S-strand

C

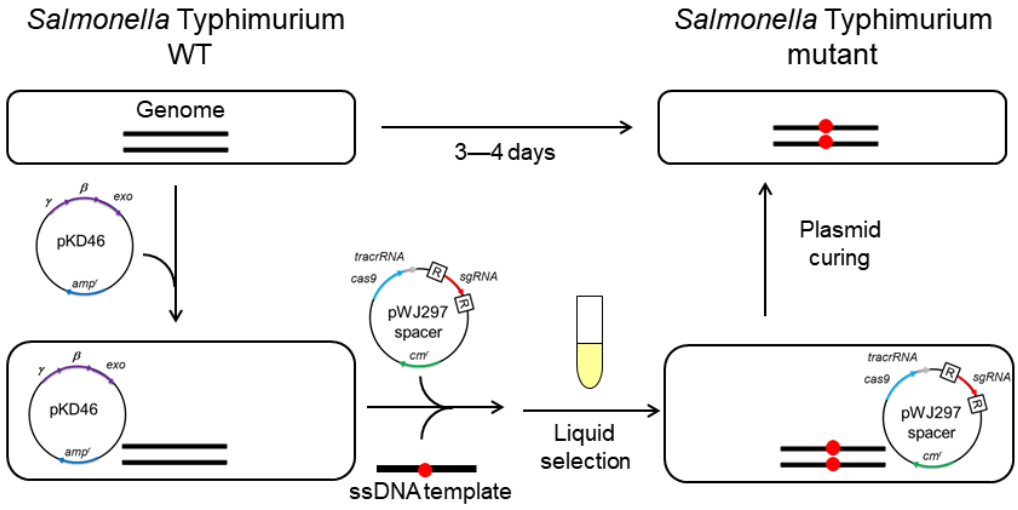

Salmonella Typhimurium WT
Salmonella Typhimurium mutant
Genome
3—4 days
pKD46
γ
β
exo
amp^r
tracrRNA
cas9
R
sgRNA
pWJ297 spacer
cm^r
Plasmid curing
Liquid selection
ssDNA template

Table 2.1. Genome-wide non-synonymous mutations in protein coding sequence of Stm strains compared to their parent strains, identified by Next-Generation whole-genome sequencing.

Stm 14028S HilA-HA

CDS	*CDS Position*	*Change*	*Codon Change*	*Amino Acid Change*	*Polymorphism Type*	*Protein Effect*	*Variant Frequency*	*Comments*
transcriptional regulator HilA	1659	=+tatccatatgatgttccagattatgct			Insertion	Extension	100.00%	C-terminus HA tagging
hypothetical protein	*233*	*A -> C*	*TTG -> TGG*	*L -> W*	*SNP (transversion)*	*Substitution*	*100.00%*	

Stm 14028S HilAK90tag-HA

CDS	*CDS Position*	*Change*	*Codon Change*	*Amino Acid Change*	*Polymorphism Type*	*Protein Effect*	*Variant Frequency*	*Comments*
transcriptional regulator HilA	268	AAA -> TAG	AAA -> TAG		Substitution	Truncation	99,70%	HilA K90 amber mutation
8-amino-7-oxononanoate synthase	*1070*	*CGGC -> TCAT*	*ACG,GCG -> ATC,ATG*	*TA -> IM*	*Substitution*	*Substitution*	*25.2% -> 27.8%*	*Spontaneous mutation*

Effects of acylation on HilA are site-specific

With optimized CRISPR-Cas9 genome editing protocol in *Salmonella*, I first generated Stm strain with HA epitag at C-terminus of endogenous *hilA* gene. HA epitagging did not alter Stm growth (**Fig. 2.8A**), and gentamicin protection assay showed that Stm HilA-HA strain was similarly infective to HeLa cells compared to wild-type (**Fig. 2.3B**). Based on this Stm HilA-HA strain, I further edited amber codon (TAG) in place of individual lysine codons in the coding sequence of *hilA* in Stm genome. Stm HilA-K90TAG, HilA-K231TAG, HilA-K324TAG, HilA-K456TAG, and HilA-K533TAG mutants are to address the acylation effect on protein function, while Stm HilA-K57TAG and Stm HilA-K527TAG mutants serve as loss-of-function and neutral control, respectively. I demonstrated that KBM could be efficiently incorporated into aforementioned amber codons at variable levels, respectively, as detected by both immunoblotting and in-gel fluorescence (**Fig. 2.8B, 2.8C**).

To characterize the phenotype of endogenous, dominantly acylated HilA-KBM mutants, I first measured the expression level of a panel of SPI-1 genes (*invF*, *prgH*, *sipA*, *orgB*, *spaO*), which are directly or indirectly activated by HilA, through qRT-PCR (**Fig. 2.8D**). The lost-of-function mutation K57KBM on HilA caused decreased expression of SPI-1 genes as expected. The positive control HilA K527KBM had similar expression of SPI-1 genes compared to HilA-HA strain (wild type, WT). Interestingly, HilA-K90KBM, -K324KBM, and -K456KBM mutants had impaired expression of SPI-1 genes compared to WT, but not HilA-K231KBM or -K533KBM mutants. These differences were not due to mere different expression levels of each HilA mutant, as shown in **Fig. 2.8C**. These results indicate that effects of acylation mimic on HilA are site-specific.

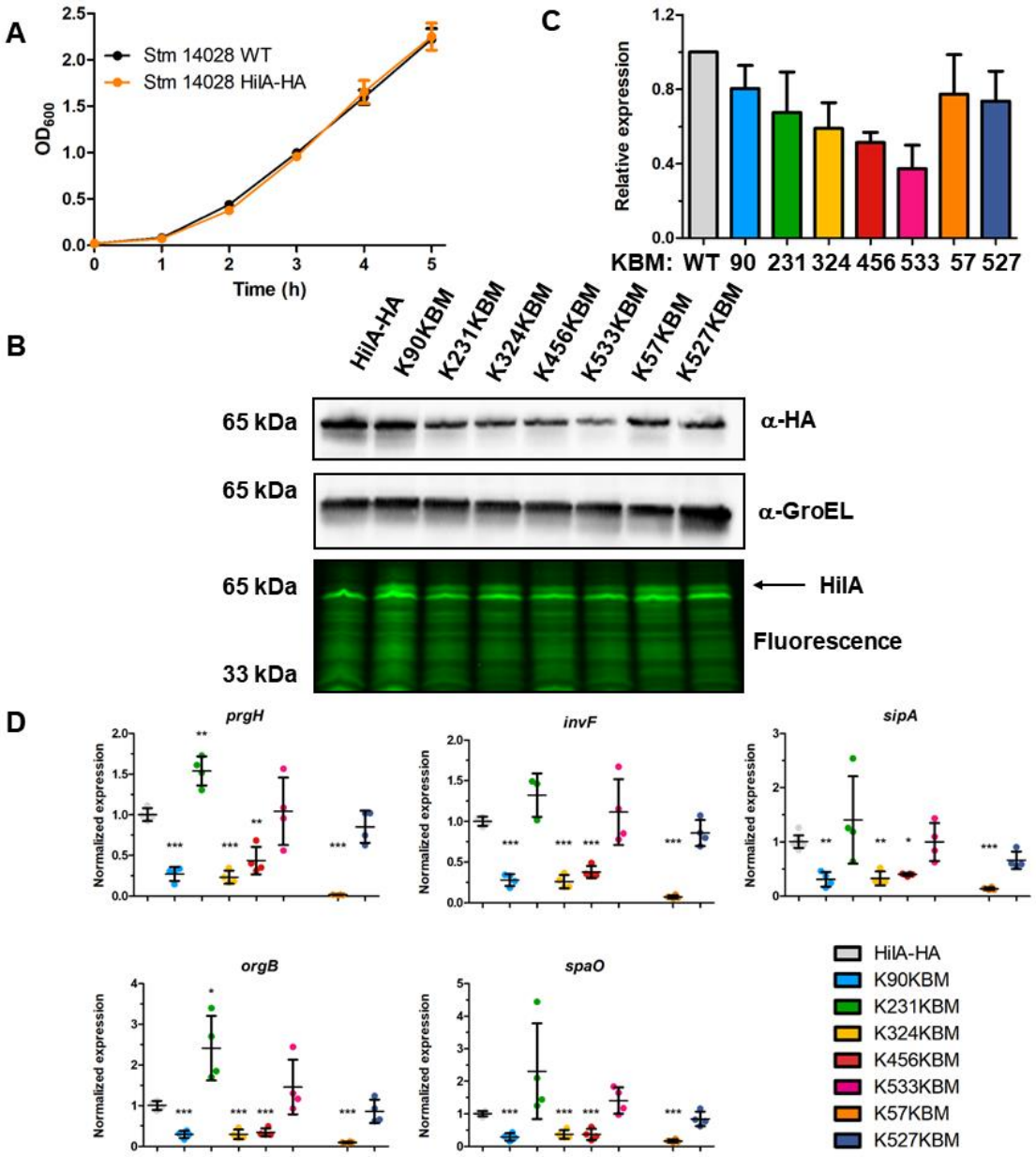


Figure 2.8. Effects of acylation on HilA are site-specific. (**A**) Growth curve of Stm 14028 WT and Stm 14028 HilA-HA in SPI-1 inducing LB. (**B**) Total cell lysates of Stm HilA-HA and Stm HilA-KBM mutants were analyzed with anti-HA or anti-GroEL immunoblotting, and SDS-PAGE in-gel fluorescence scanning. (**C**) Relative expression level of HilA-KBM mutants compared to HilA-HA, quantified over 3 independent experiments of immunoblotting, with anti-GroEL as normalization control. (**D**) Expression of SPI-1 genes *invF, prgH, sipA, orgB, and spaO* was measured by qRT-PCR from Stm endogenous HilA K-to-KBM mutants. All mutants were compared to HilA-HA with one-way ANOVA and Dunnett post-test. *, *P-value* < 0.05; **, *P-value* < 0.01; ***, *P-value* < 0.001.

Acylation at HilA K90 impacts HilA protein function

Because K90 of HilA was predicted to be in its N-terminus DNA-binding domain (**Fig. 2.3C, 2.3D**), I hypothesized that the defect in expression of SPI-1 genes observed in K90KBM mutant was a result of impaired DNA-binding ability. To test this hypothesis, I performed ChIP-Seq on Stm HilA-HA strain as well as Stm HilA-K90KBM-HA strain. Compared to HilA-HA, HilA-K90KBM-HA had decreased occupancy on promoter regions reported to be bound by HilA, including *invF* and *prgH* promoter regions (*192*) (**Fig. 2.9A**). K90KBM did not alter DNA binding specificity of HilA. This result was verified by ChIP-qPCR (**Fig. 2.9B**). HilA-K57KBM mutant, the negative control, occupied less at these promoter regions as well. Notably, HilA-K324KBM and -K456KBM mutants, which were defective in SPI-1 expression (**Fig. 2.8D**), had similar occupancy at these regions (**Fig. 2.9B**), suggesting that their defect in inducing SPI-1 expression are through alternative mechanism(s). Interestingly, HilA-K231KBM mutant has increased occupancy at *invF* and *prgH* promoter regions (**Fig. 2.9B**). These results suggest that decreased expression of SPI-1 genes in HilA-K90KBM mutant is the consequence of its defective DNA-binding activity.

Acylation at specific sites of HilA impairs *Salmonella* virulence *in vivo*

To further characterize virulence phenotype HilA K-to-KBM mutants, I assayed the infectivity of these mutants to HeLa cells. HilA-HA and K-to-KBM mutants were added to HeLa cell culture, and at 0.5 hour post-infection (hpi), gentamicin was added to kill extracellular *Salmonella*. Intracellular Stm were harvested 6 hpi and plated for CFU counting. HilA-K90KBM, -K324KBM, and -K456KBM mutants had impaired infectivity to HeLa cells compared to HilA-HA, while HilA-K231KBM had slightly enhanced infectivity, and K533KBM mutants had similar infectivity to wild-type. HilA-K57KBM and -K527KBM mutants served as loss-of-function control and positive control, respectively (**Fig. 2.10A**). These data corroborate with qRT-PCR data from these mutants, suggesting that effect of acylation on HilA is site-specific.

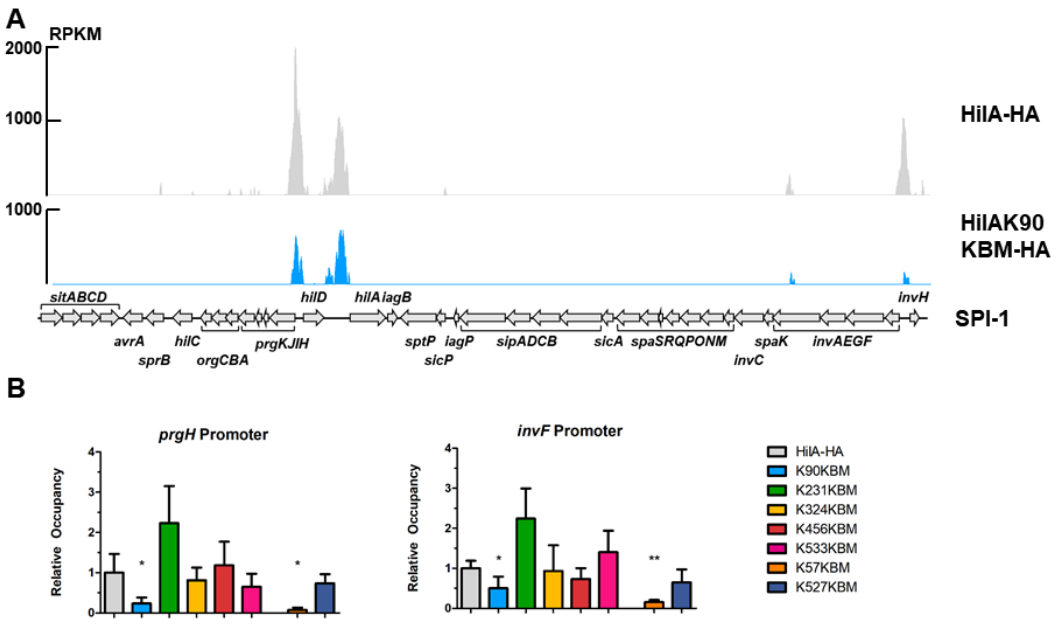


Figure 2.9. Acylation at HilA K90 impacts HilA DNA-binding activity. (**A**) HilA ChIP-Seq on Stm HilA-HA and Stm HilA-K90KBM-HA. Reads Per Kilobase Million (RPKM) were shown in SPI-1 region. (**B**) ChIP-qPCR on *prgH* promoter region (left) or *invF* promoter region (right) of Stm endogenous HilA K-to-KBM mutants. All mutants were compared to HilA-HA with one-way ANOVA and Dunnett post-test. *, *P-value* < 0.05; **, *P-value* < 0.01.

To further characterize virulence of HilA-K90KBM and -K231KBM mutants *in vivo*, we infected streptomycin-treated mice with $1x10^7$ CFU of different Stm strains, respectively. Stm in both livers and mesenteric lymph nodes (mLN) of mice were harvested and plated at 48 hpi, and Stm bacterial loads in these two organs were measured by CFU counting. Stm HilA-HA strain disseminated to livers and mLN at a level similar to its parent strain Stm 14028S wild-type (**Fig. 2.10B**). Stm HilA-K90KBM mutant was defective in systemic invasion, while K231KBM mutant behaved similarly to HilA-HA (**Fig. 2.10C**). These data suggest that HilA K90 acylated mutant has attenuated virulence *in vivo*, because of its defective expression of SPI-1 genes that are required for systemic infection.

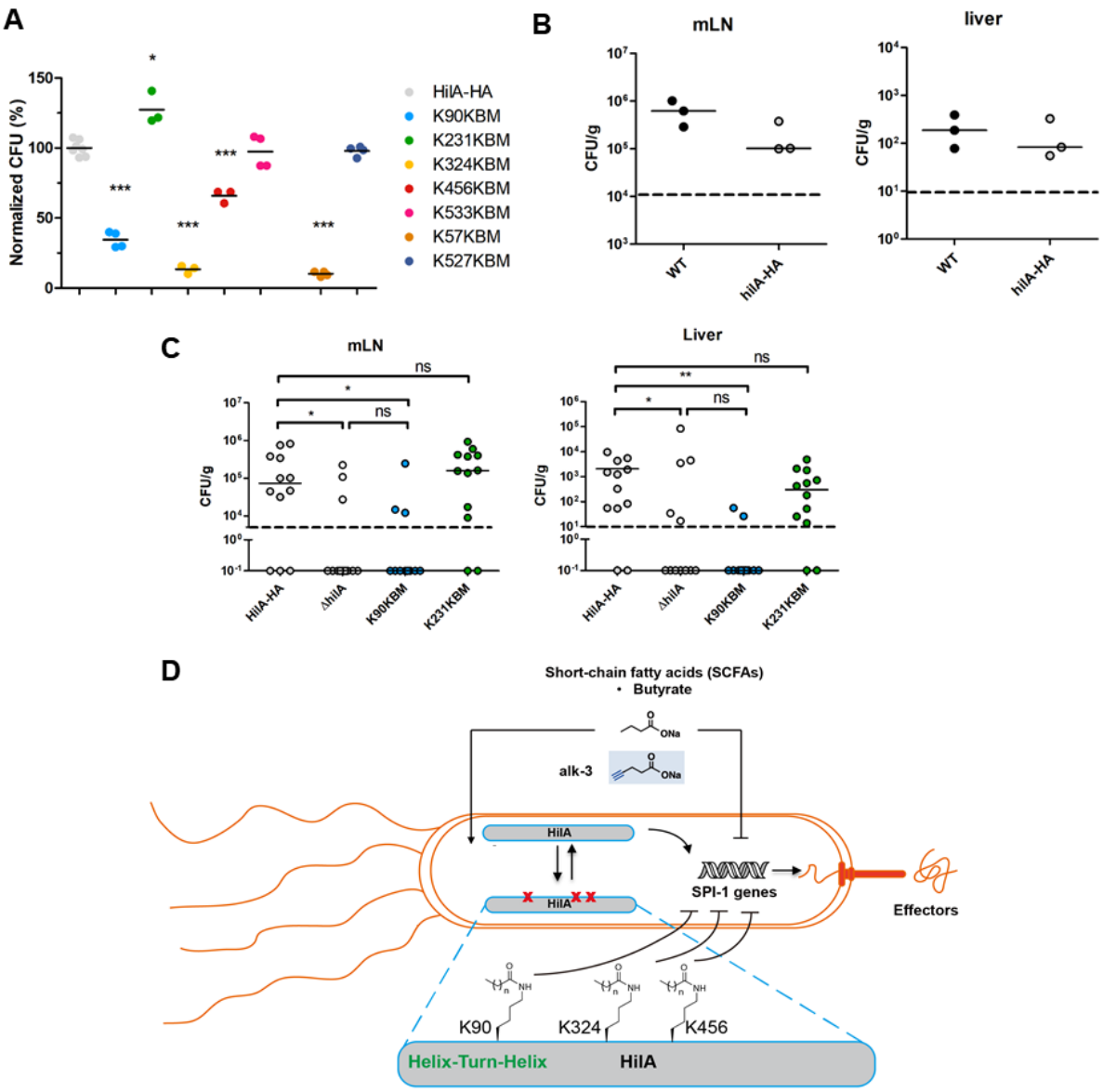


Figure 2.10. Acylation at specific sites of HilA impairs *Salmonella* virulence *in vivo*. (**A**) Gentamicin protection assay of Stm endogenous HilA K-to-KBM mutants infecting HeLa cells at MOI=10. All mutants were compared to HilA-HA with one-way ANOVA and Dunnett post-test. *, *P-value* < 0.05; **, *P-value* < 0.01; ***, *P-value* < 0.001. (**B**) Stm bacterial CFU counted from mesenteric lymph nodes (left) or liver (right) of mice 48 hpi infected with Stm WT or Stm HilA-HA. (**C**) Stm bacterial CFU counted from mesenteric lymph nodes (left) or liver (right) of mice 48 hpi infected with different Stm strains. K90KBM and K231KBM mutants were compared to HilA-HA with Kruscal-Wallis test and Dunns post-test. *, *P-value* < 0.05. (**D**) Scheme of SCFA inhibition on Stm virulence. Microbiota-derived SCFA could acylate Stm virulence regulator HilA at several lysine sites. K90, K324, and K456 acylation impact HilA function and reduce SPI-1 gene expression, leading to attenuation of Stm virulence.

Discussion

While recent microbiome researches have demonstrated strong effects of microbiome on both host immunity and pathogen infection, underlying molecular mechanisms have just begun to be unveiled. In particular, SCFAs fermented by mammalian gut microbiota accumulate abundantly in the mammalian intestinal tract, and their effects on invading enteric pathogens and the underlying mechanisms have remained largely uncharacterized.

The application of bio-orthogonal SCFA chemical reporter Alk-3 allows rapid and specific identification of molecular targets in pathogen proteome that are covalently modified by SCFAs. I demonstrated that Alk-3 behaves similarly to its natural counterpart SCFA in inhibiting virulence effector gene expression in *Salmonella* and its invasion in HeLa cells. Through CuAAC-mediated enrichment and mass spectrometry-based proteomics, I identified HilA, a key virulence transcriptional regulator, as a fatty-acylation target in *Salmonella*. I confirmed that HilA is acylated by both in-gel fluorescence scanning as well as Liquid Chromatography-tandem Mass Spectrometry (LC-MS/MS). While other SPI-1 related proteins were identified to be acylated by Alk-3 (**Appendix 2.1**), the pivotal role of HilA in SPI-1 leads us to investigate more on functional consequences of acylation on HilA. Interestingly, HilD, another key transcriptional regulator of SPI-1, which was suggested to be propionylated upon propionate incubation (*71*), was not found in our proteomics set.

I found that acylation level on HilA was not affected in the presence of acyltransferase Pat or deacylase CobB. In fact, it is reported that CobB is less efficient in deacylation of propionyllysine and butyryllysine compared to that of acetyllysine (*193*). We surmised that short-chain acyl-CoA other than acetyl-CoA may not be optimal substrates for Pat, but short-chain acyl-phosphates are

similarly reactive to acetyl-phosphate, thus Pat cannot significantly change short-chain acylation level on protein. Similarly, acetylation, but not other short-chain acylation, can be readily removed by CobB. Besides, more deacylases yet to be discovered are suggested to be present in *Salmonella* (*194–196*), and HilA may not be the substrate for CobB, therefore I could not observe a drastic change in short-chain acylation level in the presence of CobB.

After I confirmed HilA is acylated *in vivo* with both in-gel fluorescence scanning and LC-MS/MS, I seek to pinpoint the effects of acylation on HilA. CRISPR-Cas9 genome editing technique allows us to site-specifically edit codons of *hilA* gene in the *Salmonella* genome. The edited amber codon TAG enabled us to incorporate UAA to endogenously expressed HilA protein via amber suppression technology. Next-Generation Sequencing showed this genome editing approach had minimal off-target effect on other protein coding sequences. In fact, these strains allow us not only to incorporate UAA bearing bio-orthogonal chemical group such as alkynyl group, but also to incorporate photo-cross-linking diazirine group to capture protein complex *in vivo* (Chapter 3).

With Amber Suppression Technology, I could incorporate native or mimetic acyllysine to different sites of HilA. However, native acyllysines (including acetyllysine, propionyllysine, and butyryllysine) suffer from not only potential removal by endogenous deacylases, but also low incorporation efficiency. It is hard to interpret the defect in virulence when HilA native acyllysine mutants were expressed at significantly lower level than wild-type. Therefore, I incorporated KBM to study site-specific acylation effect on HilA. I found that acylation at K90, K324, and K456 affects expression of SPI-1 genes and impaired infectivity of *Salmonella* in HeLa cells. These different phenotypes cannot be explained by different expression level of these mutants. For example, HilA-K324KBM mutant was expressed at higher level than HilA-K533KBM, yet K324KBM was attenuated in virulence while K533KBM mutant was not.

The DNA-binding domain of HilA is essential for its function, as deletion of this domain renders it not active in inducing SPI-1 expression (*197*). I also demonstrated that mutation to a key DNA-interacting residue K57 also made the protein non-functional. Interestingly, K90 locates in the DBD of HilA. I showed that acylation at K90 affects the DNA-binding activity of HilA, yet K324KBM and K456KBM mutants, although attenuated in virulence, did not have defective DNA-binding activity. This suggest that acylation on HilA at different sites may alter function of the protein through various mechanisms. K324 and K456 locate in or near the predicted TPR domain of HilA, so it is possible that modification at these sites may affect interaction between HilA and its binding partners, for example, *Salmonella* RNA Polymerase (RNAP), and downregulate the transcription of SPI-1 genes. Another interesting observation is that HilAK231KBM mutant had slightly enhanced expression of SPI-1 genes and invasion in HeLa cells. Our *in vivo* photo-cross-linking data and protein size-exclusion chromatography data suggests that HilA may form homo-oligomer *in vitro* and *in vivo*, and K231 may be at the protein interface (Chapter 3).

SCFA accumulate to very high concentration (more than 10 mM) in mammalian gut, but their concentration drops significantly when the gut undergoes inflammation, which causes microbiota dysbiosis (*67*). Our work reveals that Stm exploits SCFA as an important environmental cue for its opportunistic lifestyle. When gut microbiota is normal, Stm could settle in the gut without activating its virulence, utilizing SCFA as its energy source (*198*). However, when SCFA concentration drops greatly, indicating dysbiosis of microflora and removal of “colonization resistance”, Stm could remove PTM on HilA and activate its virulence, eliciting inflammation in the gut, which further benefit Stm survival in the gut (*199*), effectively causing a vicious cycle of infection.

In summary, I applied bio-orthogonal SCFA reporter alk-3 to directly identify the biochemical targets of SCFA in *Salmonella* Typhimurium, and I found that exogenous SCFA can inhibit T3SS

of Stm and covalently modify key virulence transcriptional regulator HilA. By incorporation of bio-orthogonal stable lysine acylation mimic site-specifically to endogenously expressed HilA in *Salmonella*, I revealed that fatty-acylation on K90, K324, and K456 of HilA impaired its function to activate virulence gene expression, and decreased *Salmonella* infectivity to HeLa cells (**Fig. 2.10D**). In particular, acylation at K90 decreased HilA DNA-binding ability, making Stm less infective in mice. Our studies help elucidate fundamental mechanisms of microbiota-mediated resistance on bacterial virulence and should facilitate the development of new antibiotics and probiotics to treat bacterial infections.

Contributions and Acknowledgements

We thank Dr. Tao Peng for synthesizing KBM. Dr. Virginia Pedicord has contributed to experimental design, and data acquisition and analysis of *Salmonella* infection. We thank Dr. Shixian Lin and Prof. Peng R. Chen (Peking University) for providing pSupAR plasmid, Dr. Wenyan Jiang and Prof. Luciano Marraffini (Rockefeller University) for providing pWJ297 (pCas9) and pCRISPR plasmids and helpful discussion, Dr. Chien-Che Hung and Prof. Craig Altier (Cornell University) for providing *Salmonella* Typhimurium 14028S strain and helpful discussion, Dr. Milica Tesic Mark and Dr. Henrik Molina from The Rockefeller Proteomics Resource Center for technical assistance, and Dr. Connie Zhao from The Rockefeller Genomics Resource Center for technical assistance.

Materials and Methods

Microbial Strains and Growth Conditions

All strains used are listed in **Appendix Table 1**. All *Salmonella* Typhimurium strains used were derivatives of S. Typhimurium 14028S (*71*). *Salmonella* strains were cultured at 37°C in liquid Miller Luria-Bertani (LB) medium [10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl] (Becton Dickinson, Difco™), SPI-1 inducing LB medium [10 g/L tryptone, 5 g/L yeast extract, 300 mM NaCl], or on *Salmonella Shigella* agar (Becton Dickinson). Cultures were grown at 37°C in Multitron shaking incubator (INFORS HT) at 220 rpm. When required, antibiotics were added to the medium as follows: carbenicillin 100 µg/mL, kanamycin 50 µg/mL, and chloramphenicol 10 µg/mL.

Animal Experiments

C57BL/6J (000664) mice were purchased from the Jackson Laboratory and maintained at the Rockefeller University animal facilities under SPF conditions. Animal care and experimentation were consistent with the National Institutes of Health guidelines and approved by the Institutional Animal Care and Use Committee of the Rockefeller University.

Chemicals

Sodium butyrate was purchased from Sigma-Aldrich (303410). Alk-3 (4-pentynoic acid) was purchased from Sigma-Aldrich (232211). Az-Rho was synthesized in the lab as previously described (*164*). Az-biotin was purchased from Sigma-Aldrich (762024). Acetyllysine was purchased from Sigma-Aldrich (A4021). Propionyllysine and butyryllysine were synthesized according to previously described (*188*). KBM was synthesized by Tao Peng according to previously described (*189*).

Salmonella growth curve

Overnight cultures of *Salmonella* strains in Miller LB were diluted 1:100 to 5 mL fresh SPI-1 inducing LB medium in Falcon round-bottom 15 mL tubes. Cultures were taken out aliquots of 750 uL from 0 hour to 5 hours at 1-hour interval. Aliquots were added to 10 mm polystyrene cuvettes (Sarstedt) and OD_{600} was measured with Biophotometer plus (Eppendorf).

Preparation of *Salmonella* bacterial total cell lysates

1:50 dilutions of overnight Miller LB cultures of *Salmonella* Typhimurium strain 14028 WT overnight culture were grown in 4 mL SPI-1 inducing LB for 4 h at 37°C with 220rpm shaking. For alk-3 labeling experiments, cultures were incubated with or without 10 mM fresh alk-3 (in dH_2O). For incorporation of UAA into overexpressed HilA, cultures were added with 1 mM UAA and 0.2% arabinose. For incorporation of UAA into endogenous HilA, cultures were added with 100 uM UAA and 0.01% arabinose. *S.* Typhimurium cells were pelleted at 15000 g for 1 min, and pellets were lysed with 200 µL lysis buffer (phosphate-buffered saline (PBS) containing 0.5% Nonidet P-40, 1X EDTA-free protease inhibitor cocktail (Roche), 0.5 mg/mL lysozyme (in dH_2O) (Sigma), and 1:1,000 dilution of Benzonase (Millipore)). After re-suspension, pellets were sonicated for 10 sec for 3 times, then were incubated on ice for 30 min. Cell lysates were centrifuged at 15000 g for 1 min to remove cell debris and supernatants were collected. Protein concentration was estimated by BCA assay with BCA Protein Assay Kit (Thermo).

Salmonella protein immunoprecipitation and immunoblotting

From *Salmonella* total cell lysates prepared as described above, protein samples were boiled with 1X Laemmli buffer 95°C for 5 min. 20 uL of each sample was loaded onto a 4-20% Tris-HCl gel (Bio-Rad) for SDS-PAGE. Proteins were transferred onto 0.45 µm nitrocellulose membrane (Bio-Rad) with Trans-Blot Turbo Transfer System (Bio-Rad) at 25 V for 30 min. Membrane was blocked

with 5% non-fat milk in PBS with 0.1% Tween-20 (PBS-T) for 30 min, and primary antibody was added to solution before incubating membrane at 4°C overnight. Dilution of primary antibodies were as follows: for HA-tagged proteins, 1:2,000 anti-HA rabbit antibody H6908 (Sigma); for FLAG-tagged proteins, 1:2000 anti-FLAG rabbit antibody F7425 (Sigma); for HilA-K90Bu antigen, 1:200 anti-HilAK90Bu rabbit custom antibody (Thermo Fisher) or 1:200 anti-HilAK90 rabbit custom antibody (Thermo Fisher). Membrane was washed with PBS-T 3 times, and incubated with 1:10,000 goat polyclonal anti-rabbit HRP ab97051 (Abcam) in PBS-T with 5% non-fat milk at room temperature for 1 hour. Membrane was washed with PBS-T 3 times, and imaged with Clarity Western ECL substrate (Bio-Rad) and ChemiDoc XRS+ System (Bio-Rad).

For *Salmonella* protein immunoprecipitation, 250 µg of each total cell lysates were incubated with 20 µL PBS-T-washed EZview™ Red Anti-HA Affinity Gel (Sigma) at 4 °C for 1 hour with end-to-end rotation. Samples were washed with 200 µL PBS-T for 3 times, before being boiled with 1X Laemmli buffer 95°C for 5 min. 20 uL of each sample was loaded onto a 4-20% Tris-HCl gel (Bio-Rad) for SDS-PAGE and further immunoblotting.

In-gel fluorescence analysis of alk-3 labeling

For in-gel fluorescence analysis of alk-3 labeled *Salmonella* proteome, from the alk-3-treated or control total cell lysates prepared as described above, 45 µL of each total cell lysates (~50 µg) was added with 5 µL of click chemistry reagents as a 10X master mix (az-Rho: 0.1 mM, 10 mM stock solution in DMSO; tris(2-carboxyethyl)phosphine hydrochloride (TCEP): 1 mM, 50 mM freshly prepared stock solution in dH_2O; tris[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]amine (TBTA): (0.1 mM, 2 mM stock in 4:1 t-butanol: DMSO); $CuSO_4$ (1 mM, 50 mM freshly prepared stock in dH_2O). Samples were mixed well and incubated at room temperature for 1 h. After incubation, samples were mixed with 200 µL cold methanol, 150 µL cold water, and 75 µL cold chloroform. Sample proteins were precipitated at 18000 g for 1 min at 4 °C. After gently removing the aqueous layer, protein pellets were washed with 200 µL cold methanol, spinning down at 18000 g for 1 min

at 4 °C, and liquid was gently decanted. After washing twice, pellets were allowed air-dried before boiling with 1X Laemmli buffer.

For in-gel fluorescence analysis of alk-3 labeled HilA, from the alk-3-treated or control total cell lysates prepared as described above, 250 µg of each total cell lysates were immunoprecipitated with 20 µL PBS-T-washed EZview™ Red Anti-HA Affinity Gel (Sigma). After samples were washed with 200 µL PBS-T for 3 times, 36 µL of PBS was added to each sample. 4 µL of click chemistry reagents as a 10X master mix mentioned above were added to each sample. Samples were mixed well and incubated at room temperature for 1 h. After incubation, samples were washed with 200 µL PBS-T for 3 times.

Samples were boiled with 1X Laemmli buffer 95°C for 5 min before being loaded onto a 4-20% Tris-HCl gel (Bio-Rad) for SDS-PAGE. In-gel fluorescence scanning was performed using a Typhoon 9400 imager (Amersham Biosciences).

Alk-3 labeling Label-Free Quantitative proteomics

1:50 dilutions of overnight Miller LB cultures of *Salmonella* Typhimurium strain 14028 WT overnight culture were grown in 20 mL SPI-1 inducing LB for 4 h at 37°C with 220rpm shaking, each sample growing in 4 mL aliquots. Cultures were incubated with or without 10 mM fresh alk-3 (in dH_2O). Cultures were pooled back to 20 mL per sample in Falcon tubes, and lysed in lysis buffer described above. After re-suspension in 1 mL lysis buffer, bacteria were sonicated for 15 sec with Sonic Dismembrator Model 500 (Fisher Scientific) with 5 sec on and 10 sec off per cycle. Cell lysates were centrifuged at 15000 g for 1 min to remove cell debris and supernatants were collected. Each total cell lysates (~2 mg) was added with 100 µL of click chemistry reagents as a 10X master mix (az-Biotin: 0.1 mM, 10 mM stock solution in DMSO; tris(2-carboxyethyl)phosphine hydrochloride (TCEP): 1 mM, 50 mM freshly prepared stock solution in dH_2O; tris[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]amine (TBTA): (0.1 mM, 2 mM stock in 4:1 t-butanol: DMSO); $CuSO_4$ (1 mM, 50 mM freshly prepared stock in dH_2O). Samples were mixed well and incubated at room

temperature for 1 h. After incubation, samples were mixed with 4 mL cold methanol and incubated at -20°C overnight. Protein pellets were centrifuged at 5000 g for 30 min at 4°C, and were washed with 1 mL cold methanol 3 times. After last wash, pellets were let air dried before being re-solubilized in 250 uL 4% SDS PBS with bath sonication. Solutions were diluted with 750 uL PBS, and incubated with 60 uL PBS-T-washed High Capacity NeutrAvidin agarose (Pierce) at room temperature for 1 h with end-to-end rotation. Agarose were washed with 500 uL 1% SDS PBS 3 times, 500 uL 2M Urea PBS 3 times, and 500 uL PBS 3 times. Agarose were then reduced with 100 uL 10 mM DTT (Sigma) in PBS for 30 min at 37°C, and alkylated with 100 uL 50 mM iodoacetamide (Sigma) in PBS for 20 min in dark. On-bead proteins were digested with 400 ng Trypsin/Lys-C mix (Promega) at 37°C overnight with shaking. Digested peptides were collected and lyophilized before being desalted with custom-made stage-tip containing Empore SPE Extraction Disk (3M). Peptides were eluted with 2% acetonitrile, 2% formic acid in dH_2O.

Peptide LC-MS analysis was performed with a Dionex 3000 nano-HPLC coupled to an Orbitrap XL mass spectrometer (Thermo Fisher). Peptide samples were pressure-loaded onto a home-made C18 reverse-phase column (75 µm diameter, 15 cm length). A 180-minute gradient increasing from 95% buffer A (HPLC grade water with 0.1% formic acid) and 5% buffer B (HPLC grade acetonitrile with 0.1% formic acid) to 75% buffer B in 133 minutes was used at 200 nL/min. The Orbitrap XL was operated in top-8-CID-mode with MS spectra measured at a resolution of 60,000@m/z 400. One full MS scan (300–2000 MW) was followed by three data-dependent scans of the nth most intense ions with dynamic exclusion enabled. Peptides fulfilling a Percolator calculated 1% false discovery rate (FDR) threshold were reported.

Label-free quantification of alk-3 labeled proteins was performed with the label-free MaxLFQ algorithm in MaxQuant software as described (*200*). The search results from MaxQuant were analyzed by Perseus (http://www.perseusframework.org/). Briefly, the control replicates and alk-3 labeled sample replicates were grouped correspondingly. The results were cleaned to filter off reverse hits and contaminants. Only proteins that were identified in all alk-3 labeled sample

replicates and with more than two unique peptides were subjected to subsequent statistical analysis. LFQ intensities were used for measuring protein abundance and logarithmized. Signals that were originally zero were imputed with random numbers from a normal distribution, whose mean and standard deviation were chosen to best simulate low abundance values below the noise level (Replace missing values by normal distribution – Width = 0.3; Shift = 2.2). Significant proteins that were more enriched in alk-3 labeled sample group versus control group were determined by a threshold strategy, which combined t test p-values with ratio information. Proteins with ratio larger than or equals to 2 and *P*-value smaller than 0.05 were categorized as hits. The resulting table was exported as **Appendix 2.1**.

MS/MS detection of protein PTM

1:50 dilutions of overnight Miller LB cultures of *Salmonella* Typhimurium strain 14028 Δ*hilA* pBAD-HilA-HA-His overnight culture were grown in 500 mL SPI-1 inducing LB with 10 mM sodium propionate or 10 mM sodium butyrate for 2 h at 37°C with 220rpm shaking, before 0.2% arabinose was added to induce HilA-HA-His expression for 3 h. Bacteria was harvested with 5000 g for 10 min at 4°C, and lysed in 25 mL lysis buffer described above with 1 mM EDTA and 50 mM nicotinamide (NAM). bacteria were sonicated for 5 min with Sonic Dismembrator Model 500 (Fisher Scientific) with 5 sec on and 10 sec off per cycle. Lysates were centrifuged at 5000 g for 10 min at 4°C, and supernatants were filtered with 0.22 um filter before being loaded onto HisTrap FF 5mL column (GE Healthcare). HilA-HA-His protein was purified with wash with 3 Column Volume (CV) 100% buffer A (PBS-T, 50 mM NAM), wash with 5 CV 10% buffer B (PBS-T, 50 mM NAM, 300 mM imidazole), and 5 CV elution with gradient from 10% to 100% buffer B. Fractions containing HilA was pooled and dialyzed to PBS-T with 3 spin-dilution cycles in 10,000 MWCO filter (Amicon).

Purified HilA-HA-His was immunoprecipitated with anti-butyryllysine antibody (PTM Biolab)- or anti-propionyllysine antibody (PTM Biolab)-conjugated protein A/G magnetic beads (Pierce), and

samples were boiled and run on SDS-PAGE. Coomassie Blue-stained bands corresponding to HilA was cut out, reduced with 10 mM DTT (Sigma) in fresh 100 mM ammonium bicarbonate (ABC) for 30 min at 37°C, and alkylated with 50 mM iodoacetamide (Sigma) in fresh 100 mM ABC for 20 min in dark. Gel pieces were digested with 200 ng Trypsin/Lys-C mix (Promega) at 37°C overnight with shaking.

Salmonella CRISPR-Cas9 genome editing

To make electrocompetent *Salmonella* Typhimurium, overnight culture of *Salmonella* Typhimurium strains were diluted 1:50 to 100 mL fresh LB, and were grown at 37°C in shaking incubator at 220 rpm for 2 hours, until OD_{600} reached 0.5 to 0.7. Cells were pelleted at 5000 g for 10 min at 4°C, and washed with 50 mL ice-cold 10% glycerol twice. Cell pellets were resuspended in 500 uL 10% glycerol, and aliquoted 50 uL per tube.

Electrocompetent parent *Salmonella* Typhimurium strains were transformed with pKD46 via electroporation with Gene Pulser II (Bio-Rad) at 2.5 kV and 25 uF in 2 mm cuvette, and selected on Ampicillin agar plates at 30 °C overnight. The resulting *Salmonella* Typhimurium pKD46 strains were made into electrocompetent cells after grown at 30°C with 0.2% arabinose and ampicillin. *Salmonella* Typhimurium pKD46 electrocompetent cells were transformed with 2 uL pWJ297-sgRNA (~100 ng) and 10 uL 10 uM ssDNA editing template and selected on Chloramphenicol agar plates at 37 °C overnight. All colonies on the plate were collected with cell scraper and resuspended in 4 mL LB with Chloramphenicol. The bacterial suspension were diluted 1:50 to 4 mL fresh LB with Chloramphenicol, and were grown at 37°C in Multitron shaking incubator (INFORS HT) at 220 rpm for 2 hours. Culture was streaked onto Chloramphenicol agar plates, and colonies from the plates were randomly picked for colony PCR to confirm successful editing. Successfully edited colonies were streaked onto plain agar plates to cure pWJ297-sgRNA, and curing was confirmed by streaking on Chloramphenicol agar plates.

Salmonella whole-genome sequencing

1 mL of *Salmonella* cultures were processed with Quick-DNA Fungal/Bacterial kit (Zymo Research) per manufacturer's manual. Purified *Salmonella* genomes were sent to Rockefeller University Genomics Center for processing with Nextera XT gDNA library preparation and sequencing with MiSeq 75 Pair-End sequencing. Sequencing results were analyzed with Geneious software.

Salmonella Quantitative Reverse-Transcription PCR

500 uL of *Salmonella* cultures were processed with RNeasy Mini Kit (Qiagen) per manufacturer's manual. Concentrations of purified RNA were normalized to 100 ng/uL with RNase-free water. Quantitative Reverse-Transcription PCR (qRT-PCR) were performed with Power SYBR Green RNA-to-C_T 1-Step Kit (Applied Biosystems) per manufacturer's manual and primers listed in **Appendix Table 2**.

Salmonella HilA ChIP-qPCR and ChIP-seq

4 mL *Salmonella* cultures were crosslinked with 1% methanol-free PFA (Thermo) for 20 min at room temperature. Crosslinking were quenched with 125 mM Glycine (Fisher). Bacteria were centrifuged 16000 g at 4°C for 1 min and washed with 1 mL PBS twice. Pellets were resuspended in 500 uL ChIP Lysis Buffer (10 mM Tris-HCl, pH 8.0, 20% sucrose, 50 mM NaCl, 10 mM EDTA, 10 mg/mL lysozyme), and incubated at 37°C for 30 min. Lysates were added with 500 uL 2X RIPA buffer (100 mM Tris-HCl, pH 8.0, 300 mM NaCl, 2% Nonidet P-40 (NP-40), 1% sodium deoxycholate, 0.2% SDS) and sonicated with Sonic Dismembrator Model 500 (Fisher Scientific) for 10 sec. Resulting solutions were centrifuged at 16000 g for 1 min at room temperature. 100 uL supernatants were saved as total inputs. 750 uL of the remaining supernatants of each were incubated with 2 uL anti-HA ChIP-grade polyclonal antibody (ab9110, Abcam) at 4°C for 1 h with end-to-end rotation. The solutions were then added to 30 uL PBS-T-washed protein A/G magnetic

beads (Pierce) and incubated at 4°C for 1 h with end-to-end rotation. Beads were washed with 500 uL 1X RIPA buffer twice, 500 uL LiCl Wash buffer (10 mM Tris-HCl, pH 8.0, 250 mM LiCl, 1 mM EDTA, 0.5% NP-40, 0.5% sodium deoxycholate) twice, and 500 uL Tris-EDTA buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA) once. Samples were eluted with 100 uL SDS Elution buffer (50 mM Tris-HCl, pH 8.0, 10 mM EDTA, 1% SDS) at 65°C for 10 min. Each total input sample and ChIP sample were added with 5 uL 20 mg/mL proteinase K (Qiagen) and de-crosslinked at 65°C overnight. All de-crosslinked samples were purified with E.Z.N.A. Cycle Pure kit (Omega Bio-tek) and eluted with 100 uL elution buffer in the kit. ChIP-qPCR were performed with PowerUp SYBR Green Master Mix (Applied Biosystems) per manufacturer's manual. Same samples were sent to Rockefeller University Genomics Center for library preparation and sequenced with NextSeq High Output 75 Single-Read sequencing.

In vitro invasion assay and intracellular survival assay

HeLa cells were cultured in 12-well tissue culture plates at 80-90% confluency. Wells were added with *Salmonella* cells at an MOI = 10:1 and centrifuged at 1000 g for 5 min. Cells were incubated at 37°C with 5% CO_2 for 30 min to allow invasion. The media was then replaced with medium containing 100 μg/mL gentamicin and incubated for an additional hour to kill extracellular *Salmonella*. Wells were then washed 3 times with PBS, and cells were lysed with 500 uL 1% Triton X-100 PBS. Lysates were serially diluted and drip-dropped on *Salmonella Shigella* agar plates (BD 211597) to determine the number of invaded bacteria.

For intracellular survival assay of *Salmonella*, after incubation with medium containing 100 μg/mL gentamicin for 1 h, media was replaced with medium containing 10 μg/mL gentamicin and incubated for additional 4.5 hours at 37°C, 5% CO_2. Intracellular bacterial counts were obtained by lysing cells and drip-dropping serial dilutions on *Salmonella Shigella* agar plates.

S. Typhimurium infection of mice

To ensure effective colonization and induce infection susceptibility, SPF mice were gavaged with a single dose of 20 mg of streptomycin 24 hours before infection. Bacterial cultures of different *S.* Typhimurium strains were washed and re-suspended in sterile phosphate-buffered saline (PBS) at 10^7 CFU/mL. Mice were gavaged with 100 µL of the bacterial suspension. Leftover inocula were serially diluted and plated to confirm the number of CFU administered.

For 48 h infection experiments, mice were euthanized 48 hours after *S.* Typhimurium gavage. Colony-forming units (CFU) in the livers and mesenteric lymph nodes (mLN) were determined by plating five serial dilutions of livers or mLN suspended in sterile 0.1% Triton X-100 PBS on *Salmonella Shigella* agar (BD 211597). Resulting quantities were normalized to liver or mLN weight.

For *S.* Typhimurium infection survival assay, mice weight loss was monitored just before infection, and mice were euthanized when they reached 80% baseline weight, appeared hunched or moribund, or exhibited a visibly distended abdomen (indicative of peritoneal effusion), whichever occurred first. Death was not used as an end point. Colony-forming units (CFU) in the feces were determined by plating five serial dilutions of feces suspended in sterile PBS on *Salmonella Shigella* agar (BD 211597). Resulting quantities were normalized to fecal weight.

Quantification and Statistical Analysis

Comparisons and statistical tests were performed as indicated in each figure legend. Briefly, Pairwise comparisons were generated with two-tailed t tests. For comparisons of multiple groups over time or with two variables, a two-way analysis of variance (ANOVA) was used with an appropriate Bonferroni posttest comparing all groups to each other, all groups to a control, or selected groups to each other. Survival data were analyzed using a log-rank (Mantel-Cox) test with a Bonferroni correction for the degrees of freedom based on the number of comparisons made. To compare two groups with non-normal distribution or low sample size, the medians of the two groups were compared using a Mann-Whitney test, unless one data set contained only

zero values. In these cases, a Mann-Whitney test could not be performed because all values in the group were identical; a Wilcoxon test was performed instead. For comparisons of multiple groups with only one variable, a one-way ANOVA or Kruskal-Wallis test was performed for data with underlying normal or non-normal distribution, respectively, with Bonferroni, Dunnett's, or Dunn's posttests where appropriate. Statistical analyses were performed in GraphPad Prism software. A P value of less than 0.05 was considered significant, denoted as $*P \leq 0.05$, $**P \leq 0.01$, and $***P \leq 0.001$ for all analyses.

Chapter 3

Effects of diet-derived long-chain fatty acids on *Salmonella*

and

biochemical characterization of *Salmonella* virulence regulator HilA

Abstract

Fatty acids, especially long-chain fatty acids (LCFA) derived from the diet, have been shown to inhibit key bacterial virulence pathways, such as type III secretion system (T3SS) in Gram-negative enteric bacterial pathogens, but the molecular mechanism(s) are still under investigation. I found that fatty acids of different carbon chain length had different inhibitory efficacy on *Salmonella* Pathogenicity Island-1 T3SS, and fatty acids with longer chain were more potent. We applied bio-orthogonal alkyne-fatty acid reporters to directly identify the biochemical targets of LCFA in *Salmonella* Typhimurium. With in-gel fluorescence profiling, click chemistry-mediated enrichment and mass spectrometry-based proteomics, I found that exogenous LCFA could covalently modify key virulence transcriptional regulator HilA. With a different LCFA chemical reporter and hydroxylamine treatment, I demonstrated that the modification is a *bona fide* post-translational *N*-long-chain fatty acylation. The modification on HilA was not susceptible to enzymatic acylation or deacylation mediated by Pat and CobB. Mutating individual potential acylated lysine residues to alanine in HilA could not abolish LCFA labeling on HilA. Moreover, via Amber Suppression Technology, a photo-crosslinking unnatural amino acid was incorporated in place of key lysine residues in endogenous HilA *in vivo*, and it revealed that HilA forms oligomers in *Salmonella*, with K231 at potential protein interaction interface. These studies elucidate fundamental mechanisms of diet-mediated resistance against *Salmonella* infection and should facilitate the development of new approaches to treat bacterial infections.

Introduction

S. enterica is a Gram-negative intracellular pathogen that causes gastroenteritis and typhoid fever worldwide (*159*). Once ingested, *Salmonella* traverses the gut to the small intestine, where a set of virulence genes are activated to promote gut inflammation as well as invasion of the intestinal epithelia, allowing *Salmonella* to replicate and disseminate throughout the host (*160*). Systemic infection associated with typhoid fever is mediated by two *Salmonella* pathogenicity islands (SPI) that encode Type 3 Secretion Systems (T3SS) for bacterial invasion, dissemination, and replication inside host cells (*159*, *160*). Specifically, SPI-1 is important for *Salmonella* invasion, while SPI-2 is crucial for *Salmonella* replication in host cells (*161*). Genetic and biochemical studies have demonstrated that T3SS form multi-protein complexes to inject a variety of bacterial protein effectors into host cells for *Salmonella* pathogenesis (*159*, *160*).

Of note, medium-chain fatty acids (MCFA) and long-chain fatty acids (LCFA) from the diet have been shown to inhibit *Salmonella enterica* serovar Typhimurium (Stm) virulence and colonization (*201*, *202*). Unsaturated LCFA are shown to act as input signals for PhoP/PhoQ two-component regulatory system, and inhibit PhoP-dependent regulon, including SPI-1 (*203*). LCFA have been implicated in inhibiting *Salmonella* virulence through non-covalent interaction with SPI-1 transcription regulator HilD (*204*), but whether LCFA may affect *Salmonella* virulence through other molecular mechanism(s) remains to be studied. Understanding the effects of long-chain fatty acids on Stm will help elucidate the interactions between host, commensal bacteria, and enteric pathogen, and guide new treatment and prevention to enteric bacterial pathogens.

In this chapter, I demonstrate that HilA, a key virulence regulator of SPI-1 T3SS in Stm, is long-chain fatty-acylated using long-chain fatty acid chemical reporter alk-16. The modification is post-translational and attached to amine group on the protein. Moreover, using site-specific incorporation of photo-crosslinking unnatural amino acid, I provided evidence that HilA forms homo-oligomers *in vivo*, and K231 is at the interacting interface.

Results

Long-chain fatty acids inhibit *Salmonella* virulence

Beyond my studies with butyrate in Chapter 2, I observed that fatty acids of different carbon chain length could inhibit *Salmonella* secretion *in vitro* (**Fig. 3.1A**). For example, SCFA butyrate (C-4) inhibited *Salmonella* SPI-1 secretion at 10 mM; medium-chain fatty acids, caprylic acid (C-8) and lauric acid (C-12), inhibited secretion at 500 uM; LCFA palmitic acid (C-16) could significantly inhibit *Salmonella* secretion at 100 uM. Of note, in mammalian gut, C-16 concentrations could reach to about 400 uM (*205*). The half maximal inhibitory concentrations (IC_{50}) of each tested fatty acids were estimated based on secretion assay, and were plotted on **Fig. 3.1B**. Longer chain length of fatty acids was correlated with stronger potency in Stm T3SS inhibition. This inhibitory effect was not due to inhibition of bacterial growth, as measured by OD_{600} (data not shown). C-16 at 100 uM decreased Stm invasion ability to HeLa cells (**Fig. 3.1C**), and inhibited expression level of Stm SPI-1 effector genes such as *sipA*, as shown by quantitative reverse–transcription PCR (qRT-PCR) (**Fig. 3.1D**). This suggest that C-16 inhibits Stm virulence through antagonizing transcription and expression of Stm SPI-1 effector proteins.

Proteomic analysis of long-chain fatty acylated proteins in *Salmonella*

To identify long-chain fatty acylated proteins in *Salmonella*, I employed LCFA chemical reporter alk-16 (17-octadecynoic acid) (**Fig. 3.2A**). Alk-16 behaved similarly to its natural counterpart C-16, as it retained the ability to inhibit Stm invasion to HeLa cells (**Fig. 3.2B**). To visualize long-chain fatty acylated proteins with alk-16, I incubated Stm culture with 100 uM alk-16, and harvested total cell lysates for CuAAC reaction with azide-functionalized Rhodamine (az-Rho, **Fig.**

2.2A). SDS-PAGE followed by in-gel fluorescence scanning demonstrated that alk-16 metabolically labeled a diverse repertoire of proteins in Stm (**Fig. 3.2C**). To identify these long-chain fatty acylated proteins, I performed Label-Free Quantitative (LFQ) Proteomics analysis on Stm proteome with or without alk-16 labeling. Stm cell lysates were reacted with an azido-biotin affinity tag (az-biotin, **Fig. 2.2B**). Alk-16 labeled proteins were enriched by streptavidin beads, and digested by Trypsin/LysC mix on-bead. Digested peptides were processed and identified by Liquid Chromatography–tandem Mass Spectrometry (LC-MS/MS). The resulting spectrum were searched with MaxQuant (*176*) and quantified with Perseus (*177*).

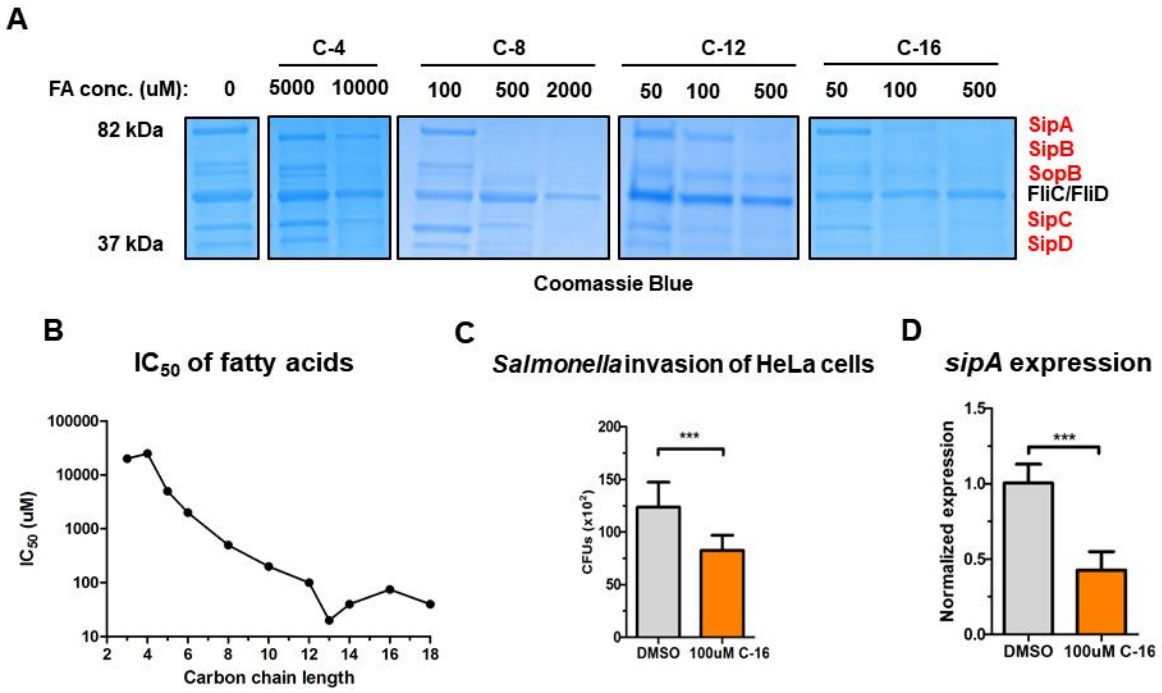


Figure 3.1. Long-chain fatty acids inhibit *Salmonella* virulence. (**A**) *Salmonella* secretion assay of different fatty acids. *Salmonella* was grown in LB in the presence of indicated concentration of different fatty acids, and secreted proteins in supernatants were precipitated with TCA and analyzed on SDS-PAGE. Identity of secreted protein bands are labeled on the right, with SPI-1 proteins colored in red. (**B**) IC_{50} of different saturated fatty acids on *Salmonella* SPI-1 secretion, estimated by secretion assay. (**C**) Gentamicin protection assay of Stm incubated with or without 100 uM C-16 infecting HeLa cells at MOI=10. (**D**) Expression of SPI-1 gene *sipA* was measured by qRT-PCR from Stm with or without 100 uM C-16 incubation.

I identified 'hits' as proteins qualified for criteria of FDR = 0.05 and S0=2 in Perseus. With this approach, I identified 236 proteins labeled by alk-16 compared to control samples (**Fig. 3.2D**, **Appendix 3.1**). Of all the hits identified, 59% were categorized as metabolism-related proteins (**Fig. 3.2E**). Moreover, 9 proteins in these hits (3.8%) were directly related to Stm virulence (**Fig. 3.2D**, **Appendix 3.1**). Notably, HilA, a master transcriptional activator of Stm SPI-1 virulence (*178*, *179*), was present in the hit set.

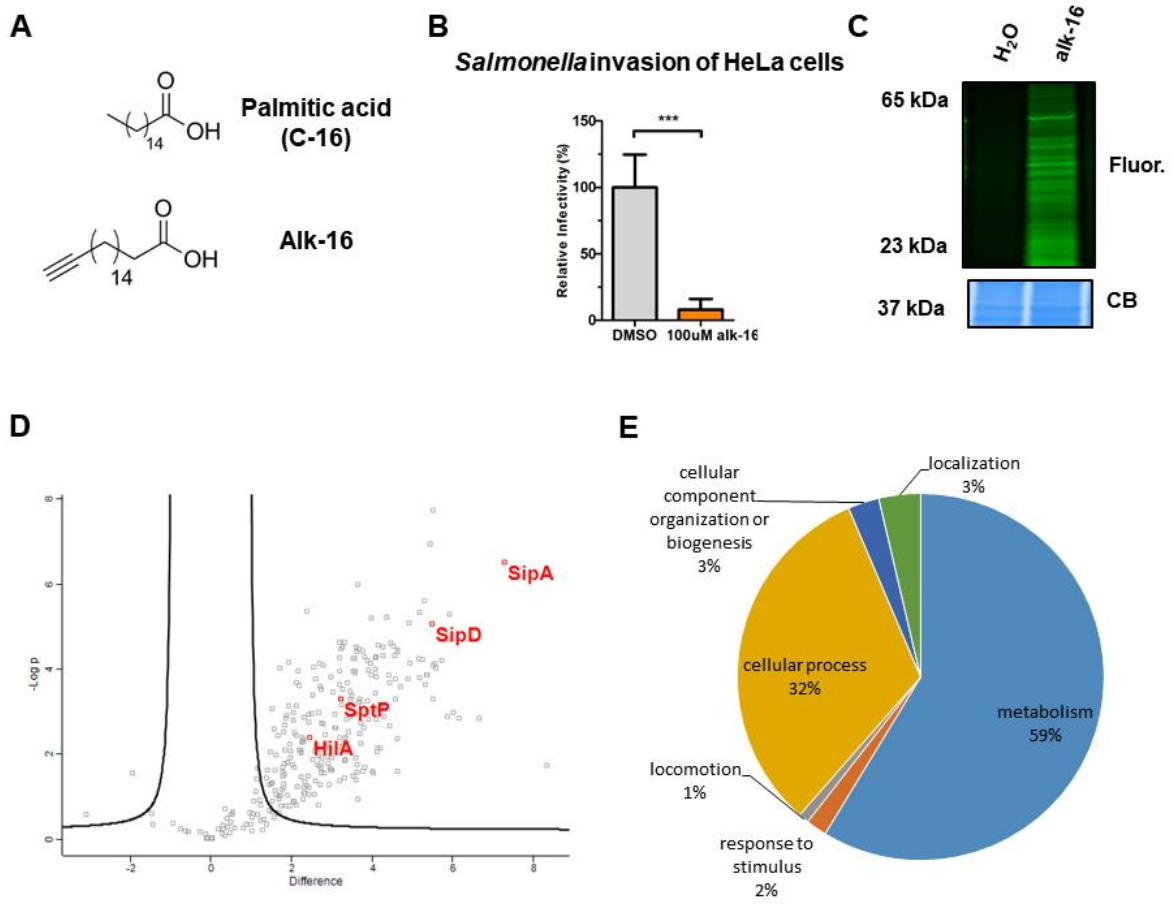


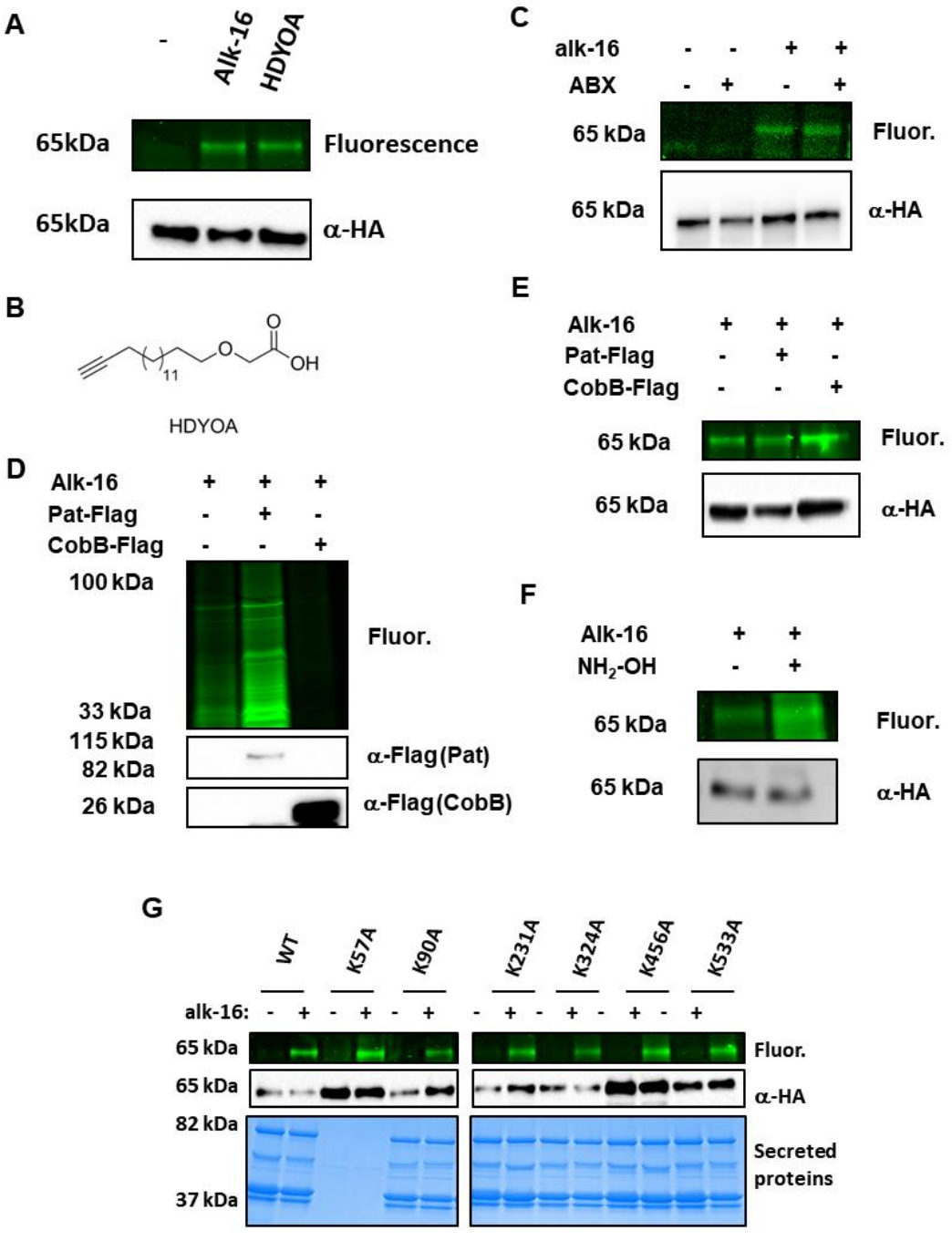
Figure 3.2. Proteomic analysis of long-chain fatty acylated proteins in *Salmonella*. (**A**) Molecular structure of palmitic acid (C-16) and Alk-16. (**B**) Gentamicin protection assay of Stm incubated with or without 100 uM alk-16 infecting HeLa cells at MOI=10. (**C**) *Salmonella* cell lysates were reacted with az-Rho by CuACC, and proteins were separated by SDS-PAGE for visualization by fluorescence gel scanning (top). Coomassie blue staining demonstrates comparable loading (bottom). (**D**) *Salmonella* cell lysates were reacted with az-biotin by CuAAC for the enrichment of alk-16–labeled proteins with streptavidin beads and identification by mass spectrometry. LFQ proteomic analysis Identified proteins that were enriched by alk-16 (top right corner above threshold line). Four selected SPI-1 proteins are labeled and colored in red. (**E**) Pie chart of annotated functions of alk-16 enriched proteins.

HilA is *N*-long-chain fatty acylated in *Salmonella*

As shown in Chapter 2, HilA is a master transcription regulator of SPI-1 that is essential for *Salmonella* virulence. To confirm that HilA was indeed long-chain fatty acylated, Stm overexpressing HilA-HA-His6 was grown in medium with alk-16, and CuAAC in-gel fluorescence scanning of cell lysates demonstrated that HilA was labeled by alk-16 (**Fig. 3.3A**). LCFA could be broken down to shorter-chain fatty acids in *Salmonella* through β-oxidation pathway (*206*), thus the fluorescence signal detected on HilA with alk-16 does not necessarily demonstrate long-chain fatty acylation on HilA. Therefore, we decided to employ another LCFA chemical reporter, HDYOA (15-hexadecynyloxyacetic acid, **Fig. 3.3B**) (*207*). HDYOA is structurally similar to alk-16, except that oxygen substitutes for carbon at beta position of the carboxyl group. This renders HDYOA resistant to β-oxidation, yet HDYOA could still label similar sets of palmitoylated proteins compared to alk-16 *in vivo* (*207*). HDYOA also readily labeled HilA (**Fig. 3.3A**), suggesting that HilA was indeed long-chain fatty acylated.

To confirm that HilA is modified post-translationally, we used a combination of antibiotics (rifampin, streptomycin, and spectinomycin) to stop the protein translation in Stm before adding alk-16 for protein labeling. HilA was still readily labeled by alk-16 (**Fig. 3.3C**), suggesting that long-chain fatty acylation on HilA is a *bona fide* post-translational modification. To investigate whether acylation on HilA is regulated by acyltransferase and deacylase in *Salmonella*, HilA-HA-His6 was co-expressed with Gcn5-like Protein Acyltransferase (Pat), or the only known protein deacylase in *Salmonella*, CobB. CuAAC in-gel fluorescence scanning showed that while Pat increased and CobB decreased alk-16 labeling level on *Salmonella* proteome (**Fig. 3.3D**), alk-16 labeling level on HilA was not affected by either Pat or CobB (**Fig. 3.3E**).

Figure 3.3. HilA is *N*-long-chain fatty acylated in *Salmonella*. (**A**) Stm overexpressing HilA-HA-His were incubated with DMSO, 100 uM alk-16, or 100 uM HDYOA before CuAAC with az-Rho, SDS-PAGE in-gel fluorescence scanning (top), and immunoblotting (bottom). (**B**) Molecular structure of HDYOA. (**C**) Stm overexpressing HilA-HA-His were incubated with or with antibiotic cocktail (ABX), before DMSO or 100 uM alk-16 were added. Cell lysates were reacted through CuAAC with az-Rho, followed by SDS-PAGE in-gel fluorescence scanning (top), and immunoblotting (bottom). (**D**)(**E**) *Salmonella* overexpressing HilA-HA-His, as well as Pat-Flag or CobB-Flag, were incubated with or without alk-16 during overexpression. Total cell lysates (**D**) and anti-HA immunoprecipitated samples (**E**) were analyzed with SDS-PAGE in-gel fluorescence scanning (top), and anti-Flag (**D**) or anti-HA (**E**) immunoblotting (bottom). (**F**) Stm overexpressing HilA-HA-His were incubated with 100 uM alk-16 and treated with or without NH_2OH, before CuAAC with az-Rho, SDS-PAGE in-gel fluorescence scanning (top), and immunoblotting (bottom). (**G**) Stm overexpressing different HilA K-toA mutants were incubated with DMSO or 100 uM alk-16, before CuAAC with az-Rho, SDS-PAGE in-gel fluorescence scanning (top), and immunoblotting (middle). The secreted SPI-1 effector proteins were precipitated from supernatant with trichloroacetic acid and run on SDS-PAGE, followed by Coomassie Blue staining (bottom).


A
-
Alk-16
HDYOA
65kDa
Fluorescence
65kDa
α-HA
B
O
OH
11
HDYOA
C
alk-16 - - + +
ABX - + - +
65 kDa
Fluor.
65 kDa
α-HA
D
Alk-16 + + +
Pat-Flag - + -
CobB-Flag - - +
100 kDa
Fluor.
33 kDa
115 kDa
82 kDa
α-Flag(Pat)
26 kDa
α-Flag(CobB)
E
Alk-16 + + +
Pat-Flag - + -
CobB-Flag - - +
65 kDa
Fluor.
65 kDa
α-HA
F
Alk-16 + +
NH2-OH - +
65 kDa
Fluor.
65 kDa
α-HA
G
WT
K57A
K90A
K231A
K324A
K456A
K533A
alk-16: - + - + - + - + - + - + - +
65 kDa
Fluor.
65 kDa
α-HA
82 kDa
37 kDa
Secreted proteins

Long-chain fatty acylation could happen on various residues on proteins. For example, *S*-palmitoylation is a post-translational modification on cysteine residues, and dynamic modification on different cysteines could affect protein localization and function in mammalian cells (*163*, *208*). *S*-stearoylation of TFR1 regulates its activation of JNK signaling and mitochondrial function in *Drosophila* (*209*). Long-chain fatty acylation may also happen on lysine residues, which could regulate protein secretion (*210*). *S*-long-chain fatty acylation is sensitive to hydroxylamine treatment, while *N*-long-chain fatty acylation would be resistant. I showed that alk-16 labeling on HilA was not sensitive to hydroxylamine treatment (**Fig. 3.3F**), which indicates that alk-16 modification on HilA is *N*-long-chain fatty acylation.

In Chapter 2, I have identified 5 lysine residues in HilA that could be short-chain fatty acylated. We hypothesized that these same lysine residues may also be residues for long-chain fatty acylation. Alanine mutation would abolish post-translational modification on the lysine residue, therefore I made individual K-to-A HilA mutants, namely K90A, K231A, K324A, K456A, and K533A, as well as K57A as a control. All mutants except K57A mutant were functional when over-expressed as they could rescue SPI-1 secretion in Stm Δ*hilA* mutant (**Fig. 3.3G**). However, all mutants were still labeled by alk-16 (**Fig. 3.3G**), suggesting that long-chain fatty acylation may be on multiple lysine residues of HilA. These results suggest that LCFA may attenuate *Salmonella* virulence also through direct fatty acylation of proteins, including HilA and other SPI-1 factors, which remains to be further characterized.

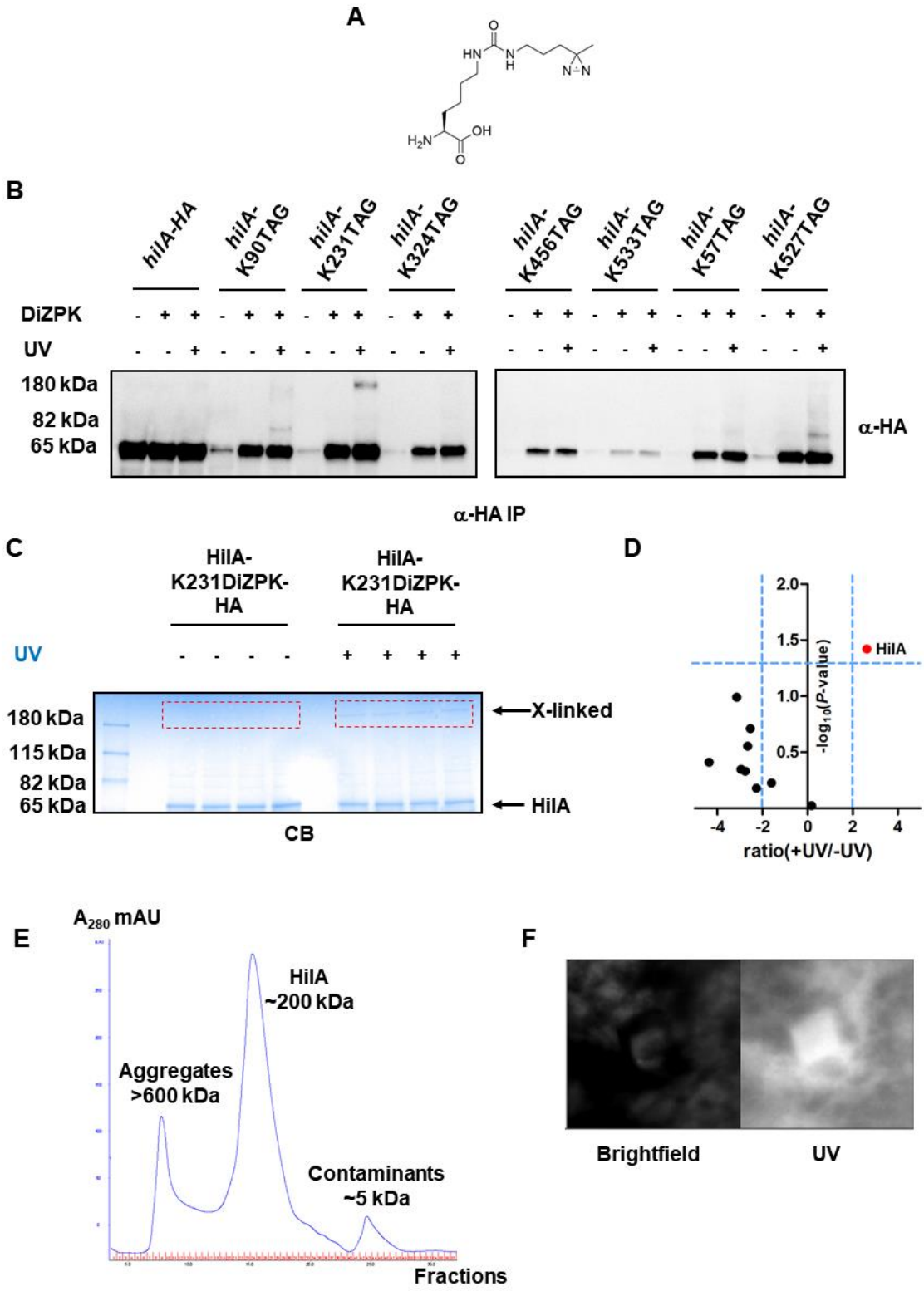
HilA forms oligomers in *Salmonella*

In Chapter 2, I have shown that when K231 is dominantly acylated through incorporation of KBM, Stm has elevated SPI-1 expression and enhanced invasion ability to HeLa cells. Since HilA is a transcription activator, which presumably recruits RNA polymerase complex during transcription

initiation, we hypothesized that K231 of HilA may be important for protein-protein interaction. To explore whether HilA has interacting protein partners in vivo, we decided to use Amber Suppression Technology to incorporate an unnatural amino acid with photo-crosslinking functional group. DiZPK (3-(3-methyl-3H-diazirin-3-yl)-propamino-carbonyl-*Nε-L*-Lysine) UAA (**Fig. 3.4A**) is developed to be site-specifically incorporated in proteins-of-interest to capture their interacting protein partners *in vivo* (*211*), including in *Salmonella* (*212*). The diazirine group at the side chain of DiZPK, when irradiated with 365 nm UV light, would form a highly reactive carbene group, crosslinking any nearby interacting molecules. This enables covalently capturing interacting partners with transient interactions at different conditions *in vivo*.

With endogenous amber mutants in hand, I incorporated DiZPK into K90, K231, K324, K456, K533, K57, and K527 of endogenous HilA, and photo-cross-linked *in vivo* with 365 nm UV light for 5 min. Interestingly, HilA-K231DiZPK crosslinked unknown partner(s) and formed a complex with apparent molecular weight about 200 kDa (**Fig. 3.4B**). Interestingly, K90DiZPK and K527DiZPK mutant also crosslinked unknown partner(s) and formed higher-molecular-weight complex. This photo-crosslinking was both site-specific and UV-dependent, as other HilA K-to-DiZPK mutants did not cross-link similar complex, nor did HilA-K231DiZPK, -K90DiZPK, or -K527DiZPK without UV treatment (**Fig. 3.4B**). I then set out to identify what interacting protein(s) HilA-K231DiZPK captured. We performed a large-scale photo-cross-linking with immunoprecipitation enrichment for HilA and its crosslinked complex, and cut out protein-complex gel bands and corresponding regions in no-UV-treatment samples in SDS-PAGE (**Fig. 3.4C**). These gel bands were digested with Trypsin/LysC mix and processed for LFQ proteomic analysis. Surprisingly, HilA was the only protein identified in the protein complex (**Fig. 3.4D**). Moreover, size-exclusion chromatography (SEC) showed that purified HilA eluted at fractions that corresponds to ~200 kDa molecular weight proteins, indicating HilA formed oligomers *in vitro* (**Fig. 3.4E**). These data suggest that HilA forms homo-oligomer *in vivo*, and K231 is at the interface between monomers. We are actively pursuing to elucidate atomic structure of HilA through X-ray crystallography. We have obtained some protein crystals of HilA through screening (**Fig. 3.4F**), and X-ray diffraction data acquisition as well as data analysis will be performed in collaboration with The Rockefeller Structural Biology Resource Center.

Figure 3.4. HilA forms oligomers in *Salmonella*. (A) Molecular structure of DiZPK. (B) Anti-HA immunoblotting of different HilA-HA amber codon mutants with or without DiZPK, and with or without UV cross-linking treatment. (**C**) Coomassie Blue stained SDS-PAGE gel of large-scale anti-HA immunoprecipitated HilA-K231DiZPK samples with or without UV treatment. Red dashed boxes indicate gel regions cut out for protease digestion and LFQ proteomics. (**D**) Volcano plot of HilAK231DiZPK photo-cross-linking LFQ proteomics. Vertical blue dashed lines indicate 2-fold difference between two sets of samples; horizontal blue dashed line indicates *P*-value=0.05. (**E**) SEC chromatogram of affinity purified HilA. Estimated molecular weight of each peak and its identity was labeled. (**F**) Brightfield and UV channel photos of a HilA crystal.


A
B
hilA-HA
hilA-K90TAG
hilA-K231TAG
hilA-K324TAG
hilA-K456TAG
hilA-K533TAG
hilA-K57TAG
hilA-K527TAG
DiZPK
UV
180 kDa
82 kDa
65 kDa
α-HA
α-HA IP
C
HilA-K231DiZPK-HA
HilA-K231DiZPK-HA
UV
180 kDa
115 kDa
82 kDa
65 kDa
X-linked
HilA
CB
D
HilA
-log10(P-value)
ratio(+UV/-UV)
E
A280 mAU
HilA ~200 kDa
Aggregates >600 kDa
Contaminants ~5 kDa
Fractions
F
Brightfield
UV

Discussion

Recent studies on host-microbe interactions have elucidated that environmental factors have significant impact on both host immunity and pathogen infection, but underlying molecular mechanisms have just begun to be unveiled. In particular, LCFAs derived from diet accumulate abundantly in the mammalian intestinal tract, and their effects on invading enteric pathogens and the underlying mechanisms have remained largely uncharacterized.

I have shown that fatty acids of different chain length have different potency in inhibiting *Salmonella* T3SS, and long-chain fatty acids are most effective. Of note, palmitic acid inhibit *Salmonella* SPI-1 secretion efficiently at 100 uM. While medium-chain fatty acids (7–12 carbon fatty acids) are present naturally in food, their concentrations in mammalian gut remain elusive. Long-chain fatty acids, e.g. palmitic acid, are abundant in both animal and vegetarian fat, and their concentrations accumulates to hundreds micromolar range in the gut (*205*). This raises the question how *Salmonella* cope with the combination of tens millimolar SCFA and hundreds micromolar LCFA, both of which are inhibitory on *Salmonella* virulence. A coordinate and centralized response to fluctuation of both SCFA and LCFA concentrations would be an efficient approach for *Salmonella* to switch itself between virulent and avirulent phenotype.

The application of bio-orthogonal LCFA chemical reporter Alk-16 allows rapid and specific identification of molecular targets in pathogen proteome that are covalently modified by LCFAs (*164*). I demonstrated that Alk-16 behaves similarly to its natural counterpart LCFA in inhibiting *Salmonella* invasion in HeLa cells. Through CuAAC-mediated enrichment and mass spectrometry-based proteomics, I identified HilA, a key virulence transcriptional regulator, as a fatty-acylation target in *Salmonella*. I confirmed that HilA is long-chain fatty acylated by in-gel

fluorescence scanning. Direct detection of long-chain fatty acylation on proteins by Liquid Chromatography-tandem Mass Spectrometry (LC-MS/MS) has been a technical challenge, requiring extensive optimization on sample preparation, LC condition, and data acquisition (Thinon E. *et al*., unpublished). Interestingly, HilD, another key transcriptional regulator of SPI-1, which was suggested to be inhibited by LCFA through non-covalent interaction (*204*), was not found in our proteomics set. Indeed, our method could only detect proteins covalently labeled by alk-16. A bi-functional LCFA, containing both photo-crosslinking and alkyne groups, would enable capture of non-covalent interacting proteins with LCFA (*213*).

I found that long-chain fatty acylation on HilA was a *bona fide* post-translational *N*-acylation. While long-chain fatty acylations, both *S*- and *N*-linked, have been studied extensively in eukaryotes (*163*, *210*, *214*), their presence on bacterial proteins has been less studied. Previous report suggests the presence of *S*-long-chain acylation on *E. coli* protein YjgF, as the modification is abolished by cysteine-to-serine mutation (*164*). Lysine long-chain fatty acylation is also reported on *E. coli* hemolysin, which is required for its toxic activity (*215*). *N*-long-chain fatty acylation may occur on multiple lysine residues of HilA, and identification of exact modification sites would facilitate downstream studies on effects of long-chain fatty acylation on HilA function. Hang lab is currently developing a new isotopic cleavable affinity tag for direct identification of PTM sites (Tsukidate T *et al*., unpublished), which may help identification of long-chain fatty acylation sites on HilA.

To study protein-protein interaction between HilA and its partners, I have incorporated photo-crosslinking DiZPK to endogenous HilA. I found that K231 and K527 could photo-crosslink interacting protein partner(s). K231 is identified to be short-chain fatty acylated (Chapter 2), while K527 is not reported to have significant functional role in HilA. Structural studies might help explain where these two lysine residues reside spatially in the protein, and how they might contribute to

protein-protein interaction. Surprisingly, through LFQ proteomics analysis, we found that HilA-K231DiZPK photo-crosslinked itself but not other proteins *in vivo*. SEC of purified HilA also demonstrated oligomerization of HilA *in vitro*. It would be interesting to see if HilA crystal is packed in oligomer form, which may reveal information about protein-protein interaction interface.

In summary, I applied bio-orthogonal LCFA reporter alk-16 to directly identify long-chain fatty acylated proteins in *Salmonella* Typhimurium, and I found that exogenous LCFA can inhibit T3SS of Stm and covalently modify key virulence transcriptional regulator HilA. The post-translational *N*-long-chain fatty acylation may contribute to inhibitory effect of LCFA on *Salmonella* virulence. Moreover, photo-crosslinking studies reveal that K231 and K527 may contribute to HilA interaction with other proteins, and K231 may be at the interface of HilA homo-oligomers. Our studies help elucidate fundamental mechanisms of diet-mediated resistance on bacterial virulence and should facilitate the development of new approaches to treat bacterial infections.

Contributions and Acknowledgements

We thank Dr. Tao Peng for synthesizing alk-16 and DiZPK. Rotation student Nicholas Prescott purified HilA in large scale, and is collaborating with Dr. Deena Oren from The Rockefeller Structural Biology Resource Center for X-ray crystallography of HilA. We thank Dr. Shixian Lin and Prof. Peng R. Chen (Peking University) for providing pSupAR plasmid, Wenyan Jiang and Prof. Luciano Marraffini (Rockefeller University) for providing pCas9 and pCRISPR plasmids and helpful discussion, Dr. Chien-Che Hung and Prof. Craig Altier (Cornell University) for providing *Salmonella* Typhimurium 14028S strain and helpful discussion. We thank Dr. Milica Tesic Mark and Dr. Henrik Molina from The Rockefeller Proteomics Resource Center for technical assistance. We thank Corey Model (SSRP student) and Michael Kuckyr (SURF student) for experimental assistance.

Materials and Methods

Microbial Strains and Growth Conditions

All strains used are listed in **Appendix Table 1**. All *Salmonella* Typhimurium strains used were derivatives of *S.* Typhimurium 14028S (14). *Salmonella* strains were cultured at 37°C in liquid Miller Luria-Bertani (LB) medium [10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl] (Becton Dickinson, DifcoTM), SPI-1 inducing LB medium [10 g/L tryptone, 5 g/L yeast extract, 300 mM NaCl], or on *Salmonella Shigella* agar (Becton Dickinson). Cultures were grown at 37°C in Multitron shaking incubator (INFORS HT) at 220 rpm. When required, antibiotics were added to the medium as follows: carbenicillin 100 µg/mL, kanamycin 50 µg/mL, and chloramphenicol 10 µg/mL.

Chemicals

Alk-16 (17-Octadecynoic Acid) was synthesized according to previously described (*216*). Az-Rho was synthesized in the lab as previously described (17). Az-biotin was purchased from Sigma-Aldrich (762024). DiZPK was synthesized according to previously described (*217*).

Preparation of *Salmonella* bacterial total cell lysates

1:50 dilutions of overnight Miller LB cultures of *Salmonella* Typhimurium strain 14028 WT overnight culture were grown in 4 mL SPI-1 inducing LB for 4 h at 37°C with 220rpm shaking. For alk-16 labeling experiments, cultures were incubated with or without 10 mM fresh alk-16 (in DMSO). For incorporation of UAA into overexpressed HilA, cultures were added with 1 mM UAA and 0.2% arabinose. For incorporation of UAA into endogenous HilA, cultures were added with

100 uM UAA and 0.01% arabinose. S. Typhimurium cells were pelleted at 15000 g for 1 min, and pellets were lysed with 200 µL lysis buffer (phosphate-buffered saline (PBS) containing 0.5% Nonidet P-40, 1X EDTA-free protease inhibitor cocktail (Roche), 0.5 mg/mL lysozyme (in dH_2O) (Sigma), and 1:1,000 dilution of Benzonase (Millipore)). After re-suspension, pellets were sonicated for 10 sec for 3 times, then were incubated on ice for 30 min. Cell lysates were centrifuged at 15000 g for 1 min to remove cell debris and supernatants were collected. Protein concentration was estimated by BCA assay with BCA Protein Assay Kit (Thermo).

***Salmonella* protein immunoprecipitation and immunoblotting**

From *Salmonella* total cell lysates prepared as described above, protein samples were boiled with 1X Laemmli buffer 95°C for 5 min. 20 uL of each sample was loaded onto a 4-20% Tris-HCl gel (Bio-Rad) for SDS-PAGE. Proteins were transferred onto 0.45 µm nitrocellulose membrane (Bio-Rad) with Trans-Blot Turbo Transfer System (Bio-Rad) at 25 V for 30 min. Membrane was blocked with 5% non-fat milk in PBS with 0.1% Tween-20 (PBS-T) for 30 min, and primary antibody was added to solution before incubating membrane at 4°C overnight. Dilution of primary antibodies were as follows: for HA-tagged proteins, 1:2,000 anti-HA rabbit antibody H6908 (Sigma); for FLAG-tagged proteins, 1:2000 anti-FLAG rabbit antibody F7425 (Sigma); for HilA-K90Bu antigen, 1:200 anti-HilAK90Bu rabbit custom antibody (Thermo Fisher) or 1:200 anti-HilAK90 rabbit custom antibody (Thermo Fisher). Membrane was washed with PBS-T 3 times, and incubated with 1:10,000 goat polyclonal anti-rabbit HRP ab97051 (Abcam) in PBS-T with 5% non-fat milk at room temperature for 1 hour. Membrane was washed with PBS-T 3 times, and imaged with Clarity Western ECL substrate (Bio-Rad) and ChemiDoc XRS+ System (Bio-Rad).

For *Salmonella* protein immunoprecipitation, 250 µg of each total cell lysates were incubated with 20 µL PBS-T-washed EZview™ Red Anti-HA Affinity Gel (Sigma) at 4 °C for 1 hour with end-to-

end rotation. Samples were washed with 200 µL PBS-T for 3 times, before being boiled with 1X Laemmli buffer 95°C for 5 min. 20 uL of each sample was loaded onto a 4-20% Tris-HCl gel (Bio-Rad) for SDS-PAGE and further immunoblotting.

In-gel fluorescence analysis of LCFA reporter labeling

For in-gel fluorescence analysis of alk-16 labeled *Salmonella* proteome, from the alk-16-treated or control total cell lysates prepared as described above, 45 µL of each total cell lysates (~50 µg) was added with 5 µL of click chemistry reagents as a 10X master mix (az-Rho: 0.1 mM, 10 mM stock solution in DMSO; tris(2-carboxyethyl)phosphine hydrochloride (TCEP): 1 mM, 50 mM freshly prepared stock solution in dH2O; tris[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]amine (TBTA): (0.1 mM, 2 mM stock in 4:1 t-butanol: DMSO); CuSO4 (1 mM, 50 mM freshly prepared stock in dH2O). Samples were mixed well and incubated at room temperature for 1 h. After incubation, samples were mixed with 200 µL cold methanol, 150 µL cold water, and 75 µL cold chloroform. Sample proteins were precipitated at 18000 g for 1 min at 4 °C. After gently removing the aqueous layer, protein pellets were washed with 200 µL cold methanol, spinning down at 18000 g for 1 min at 4 °C, and liquid was gently decanted. After washing twice, pellets were allowed air-dried before boiling with 1X Laemmli buffer.

For in-gel fluorescence analysis of alk-16 or HDYOA labeled HilA, from the alk-16 or HDYOA-treated or control total cell lysates prepared as described above, 250 µg of each total cell lysates were immunoprecipitated with 20 µL PBS-T-washed EZview™ Red Anti-HA Affinity Gel (Sigma). After samples were washed with 200 µL PBS-T for 3 times, 36 µL of PBS was added to each sample. 4 µL of click chemistry reagents as a 10X master mix mentioned above were added to each sample. Samples were mixed well and incubated at room temperature for 1 h. After incubation, samples were washed with 200 µL PBS-T for 3 times.

Samples were boiled with 1X Laemmli buffer 95°C for 5 min before being loaded onto a 4-20% Tris-HCl gel (Bio-Rad) for SDS-PAGE. In-gel fluorescence scanning was performed using a Typhoon 9400 imager (Amersham Biosciences).

Alk-16 labeling Label-Free Quantitative proteomics

1:50 dilutions of overnight Miller LB cultures of *Salmonella* Typhimurium strain 14028 WT overnight culture were grown in 20 mL SPI-1 inducing LB for 4 h at 37°C with 220rpm shaking, each sample growing in 4 mL aliquots. Cultures were incubated with or without 100 uM fresh alk-16 (in DMSO). Cultures were pooled back to 20 mL per sample in Falcon tubes, and lysed in lysis buffer described above. After re-suspension in 1 mL lysis buffer, bacteria were sonicated for 15 sec with Sonic Dismembrator Model 500 (Fisher Scientific) with 5 sec on and 10 sec off per cycle. Cell lysates were centrifuged at 15000 g for 1 min to remove cell debris and supernatants were collected. Each total cell lysates (~2 mg) was added with 100 µL of click chemistry reagents as a 10X master mix (az-Biotin: 0.1 mM, 10 mM stock solution in DMSO; tris(2-carboxyethyl)phosphine hydrochloride (TCEP): 1 mM, 50 mM freshly prepared stock solution in dH2O; tris[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]amine (TBTA): (0.1 mM, 2 mM stock in 4:1 t-butanol: DMSO); CuSO4 (1 mM, 50 mM freshly prepared stock in dH2O). Samples were mixed well and incubated at room temperature for 1 h. After incubation, samples were mixed with 4 mL cold methanol and incubated at -20°C overnight. Protein pellets were centrifuged at 5000 g for 30 min at 4°C, and were washed with 1 mL cold methanol 3 times. After last wash, pellets were let air dried before being re-solubilized in 250 uL 4% SDS PBS with bath sonication. Solutions were diluted with 750 uL PBS, and incubated with 60 uL PBS-T-washed High Capacity NeutrAvidin agarose (Pierce) at room temperature for 1 h with end-to-end rotation. Agarose were washed with 500 uL 1% SDS PBS 3 times, 500 uL 2M Urea PBS 3 times, and 500 uL PBS 3 times. Agarose were then reduced with 100 uL 10 mM DTT (Sigma) in PBS for 30 min at 37°C, and alkylated with 100 uL 50 mM

iodoacetamide (Sigma) in PBS for 20 min in dark. On-bead proteins were digested with 400 ng Trypsin/Lys-C mix (Promega) at 37°C overnight with shaking. Digested peptides were collected and lyophilized before being desalted with custom-made stage-tip containing Empore SPE Extraction Disk (3M). Peptides were eluted with 2% acetonitrile, 2% formic acid in dH2O.

Peptide LC-MS analysis was performed with a Dionex 3000 nano-HPLC coupled to an Orbitrap XL mass spectrometer (Thermo Fisher). Peptide samples were pressure-loaded onto a home-made C18 reverse-phase column (75 µm diameter, 15 cm length). A 180-minute gradient increasing from 95% buffer A (HPLC grade water with 0.1% formic acid) and 5% buffer B (HPLC grade acetonitrile with 0.1% formic acid) to 75% buffer B in 133 minutes was used at 200 nL/min. The Orbitrap XL was operated in top-8-CID-mode with MS spectra measured at a resolution of 60,000@m/z 400. One full MS scan (300–2000 MW) was followed by three data-dependent scans of the nth most intense ions with dynamic exclusion enabled. Peptides fulfilling a Percolator calculated 1% false discovery rate (FDR) threshold were reported.

Label-free quantification of alk-16 labeled proteins was performed with the label-free MaxLFQ algorithm in MaxQuant software as described (54). The search results from MaxQuant were analyzed by Perseus (http://www.perseusframework.org/). Briefly, the control replicates and alk-16 labeled sample replicates were grouped correspondingly. The results were cleaned to filter off reverse hits and contaminants. Only proteins that were identified in all alk-16 labeled sample replicates and with more than two unique peptides were subjected to subsequent statistical analysis. LFQ intensities were used for measuring protein abundance and logarithmized. Signals that were originally zero were imputed with random numbers from a normal distribution, whose mean and standard deviation were chosen to best simulate low abundance values below the noise level (Replace missing values by normal distribution – Width = 0.3; Shift = 2.2). Significant proteins that were more enriched in alk-16 labeled sample group versus control group were determined by a volcano plot-based strategy, which combined t test p-values with ratio information.

A hyperbolic significance curve in the volcano plot corresponding to a given FDR (= 0.05) and S0 value (= 2) was determined by a permutation-based method. The resulting table was exported as **Appendix 3.1**.

Salmonella in vivo photo-crosslinking

For small-scale experiment, 4 mL of *Salmonella* amber mutants expressing PylRS-ASF mutant were grown in the presence of 0.01% arabinose and 100 uM DiZPK. After 4 hours, bacterial cells were pelleted at 15000 g for 1 min, and resuspended in 1 mL PBS. After cell suspensions were transferred to 6-well plate, samples were photo-crosslinked in XL-1000 UV Crosslinker (Spectronics Corporation) for 5 min, with samples about 3 cm from the lamp on ice. Cells were lysed, immunoprecipitated, and immunoblotted as described above.

For large-scale purification and identification, 20 mL of *Salmonella* amber mutants expressing PylRS-ASF mutant were grown in the presence of 0.01% arabinose and 100 uM DiZPK. Cells were resuspended in 5 mL PBS and photo-crosslinked as described above. Lysates were immunoprecipitated with 50 uL PBS-T washed anti-HA magnetic beads (Pierce). Samples were run on SDS-PAGE and stained with SafeStain Coomassie Blue. Crosslinked bands as well as control bands were cut out with clean blazers and diced into 1 mm x 1mm cubes. Gels were processed per in-gel digestion protocol described previously (*218*) and sent to The Rockefeller Proteomics Resource Center for protein identification.

Salmonella Quantitative Reverse-Transcription PCR

500 uL of *Salmonella* cultures were processed with RNeasy Mini Kit (Qiagen) per manufacturer's manual. Concentrations of purified RNA were normalized to 100 ng/uL with RNase-free water.

Quantitative Reverse-Transcription PCR (qRT-PCR) were performed with Power SYBR Green RNA-to-CT 1-Step Kit (Applied Biosystems) per manufacturer's manual and primers listed in **Appendix Table 2**.

In vitro invasion assay and intracellular survival assay

HeLa cells were cultured in 12-well tissue culture plates at 80-90% confluency. Wells were added with *Salmonella* cells at an MOI = 10:1 and centrifuged at 1000 g for 5 min. Cells were incubated at 37°C with 5% CO2 for 30 min to allow invasion. The media was then replaced with medium containing 100 µg/mL gentamicin and incubated for an additional hour to kill extracellular *Salmonella*. Wells were then washed 3 times with PBS, and cells were lysed with 500 uL 1% Triton X-100 PBS. Lysates were serially diluted and drip-dropped on *Salmonella* Shigella agar plates (BD 211597) to determine the number of invaded bacteria.

Quantification and Statistical Analysis

Comparisons and statistical tests were performed as indicated in each figure legend. Briefly, Pairwise comparisons were generated with two-tailed t tests. For comparisons of multiple groups over time or with two variables, a two-way analysis of variance (ANOVA) was used with an appropriate Bonferroni posttest comparing all groups to each other, all groups to a control, or selected groups to each other. Survival data were analyzed using a log-rank (Mantel-Cox) test with a Bonferroni correction for the degrees of freedom based on the number of comparisons made. To compare two groups with non-normal distribution or low sample size, the medians of the two groups were compared using a Mann-Whitney test, unless one data set contained only zero values. In these cases, a Mann-Whitney test could not be performed because all values in the group were identical; a Wilcoxon test was performed instead. For comparisons of multiple

groups with only one variable, a one-way ANOVA or Kruskal-Wallis test was performed for data with underlying normal or non-normal distribution, respectively, with Bonferroni, Dunnett's, or Dunn's posttests where appropriate. Statistical analyses were performed in GraphPad Prism software. A P value of less than 0.05 was considered significant, denoted as *P $\leqslant$ 0.05, **P $\leqslant$ 0.01, and ***P $\leqslant$ 0.001 for all analyses.

Chapter 4

Summary and Future Outlook

Recent research has revealed many mechanisms by which gut microbiota influences host immunity to defend against invading pathogens, but how microbiota directly antagonizes pathogen virulence is less studied. In this thesis, I describe how fatty acids, derived from both gut microbiota and diet, contribute to attenuation of virulence of enteric pathogen *Salmonella*.

In Chapter 1, I review how dietary and microbiota metabolites affect different aspects of host-microbe interactions. These metabolites are classified into microbial-associated molecular patterns and microbiota-derived secondary metabolites. Small molecules reviewed in this chapter not only enhances host innate and adaptive immunity, but also directly inhibit virulence of invading pathogens, providing colonization resistance to the host. These findings highlight the importance of understanding the intricate interactions between host and microbiota, and should provide insights in developing microbiota-targeting therapeutics for host physiology, immunity, and pathogen resistance.

In Chapter 2, I describe a mechanism by which microbiota-derived short-chain fatty acids inhibit virulence of *Salmonella* Typhimurium. We use a chemical reporter strategy to identify molecular targets of short-chain fatty acids in *Salmonella.* I demonstrate that alkynyl-functionalized short-chain fatty acids can be metabolized and covalently attached to proteins in *Salmonella*. Proteomic analysis reveal that HilA, a key virulence transcription regulator, is short-chain fatty acylated. I employ Amber Suppression Technology and CRISPR-Cas9 genome editing to faithfully mimic butyrylation on endogenous HilA. Biochemical and functional characterization show that acylation of HilA has site-specific effect, and K90 butyrylation affect HilA DNA-binding activity and *Salmonella* invasion in mice. Overall, our results discover a mechanism by which gut microbiota provides resistance against *Salmonella* through short-chain fatty acids.

In Chapter 3, I find that dietary long-chain fatty acids potently inhibit *Salmonella* virulence. Chemical proteomics with alkynyl-functionalized long-chain fatty acids reveal proteins that are long-chain fatty acylated in *Salmonella*, including HilA. Modification by long-chain fatty acids on HilA is post-translationally *N*-linked. Moreover, with photo-crosslinking unnatural amino acid, we discover that HilA forms homo-oligomers in *Salmonella*. Our data suggest that dietary long-chain fatty acids may interfere pathogenesis of *Salmonella* through post-translational modification, and further structural characterization of HilA may reveal novel target for treatment of *Salmonella* infection.

According to sequence homology, HilA belongs to OmpR/PhoB Response Regulator (RR) subfamily, which represents about one-third of all RRs. Response Regulators (RRs) are members within bacterial two-component systems that enable fast response to environmental signals and is critical for regulation of bacterial physiology. In most cases, RRs are activated by phosphate transfer from a cognate sensory histidine kinase to its aspartate residue in the receiver domain. Activated RRs then exert its function through its effector domain. OmpR/PhoB family RRs are often activated through phosphorylation-dependent mechanism. However, a lot of atypical OmpR/PhoB family RRs are activated through phosphorylation-independent mechanism (*219*, *220*). For example, HP1043, an atypical OmpR/PhoB family RR in *Helicobacter pylori*, binds to inverted repeat DNA sequence motifs with unknown regulatory activity (*220*–*222*). The aspartate that should have been phosphorylated in HP1043 is replaced by a lysine residue (*221*). Indeed, Asp67 of HilA, the conserved Asp residue that is supposed to be phosphorylated predicted by homology, could not be detected with phosphorylation (*185*). Our data suggests that Asp67 is unlikely to be in the receiver domain, as both Lys57 and Lys90, two residues flanking Asp67, resides in DNA-binding domain and directly affect DNA-binding activity of HilA. So far, no cognate sensory kinase for HilA has been reported. Our studies have demonstrated that HilA itself may be a sensor of environmental fatty acids, including short-chain fatty acids from the gut microbiota

and long-chain fatty acids from the diet. HilA may have evolved to use lysine residues instead of aspartate to sense environmental metabolites and regulate *Salmonella* virulence. Our research indicates that post-translational fatty acylation may represent one the sensing mechanisms by atypical OmpR/PhoB family RRs to environmental signals.

Even among atypical OmpR/PhoB family RR members, HilA is peculiar. RRs usually have an N-terminal receiver domain and a C-terminal effector domain (in most cases DNA-binding domain), yet HilA adopts a swapped domain architecture, with an N-terminal DNA-binding domain and a C-terminal TPR domain with unknown function. OmpR/PhoB family RRs often form homodimers and recognize tandem or inverted repeating DNA elements for transcription regulation (*223*, *224*). Yet the DNA-binding motif of HilA remains debated. Lostroh and Lee reported that HilA binds to a direct repeat sequence motif termed 'HilA Box' (*181*), but De Keersmaecker *et al.* identified other loci that were bound by HilA do not contain 'HilA Box' (*182*). They proposed a new 'HilA Box' motif (*182*), yet on closer analysis, that is essentially -35 element in prokaryotic promoters. The selectivity of HilA DNA-binding domain remains a mystery. Our HilA ChIP-Seq data may provide hints on the DNA motifs, and further in-depth data analysis is required.

While HilA may act as a sensor for fatty acids in *Salmonella*, *Salmonella* as well as other intestinal bacteria uses other proteins and mechanisms to sense fatty acids too. In the presence of high concentrations of fatty acids, bacterial cytoplasm would accumulate high-energy fatty acid intermediates, including fatty acyl-CoA and fatty acyl-phosphate (*194*, *195*, *225*). Chemical acylation of proteins by these intermediates as well as enzymatic acylation influences bacterial metabolism and physiology. For instance, dynamic acetylation level on metabolic enzymes in *Salmonella* regulates the direction of glycolysis versus gluconeogenesis, and the branching between citrate cycle and glyoxylate bypass (*173*). RcsB, a global regulatory RR that controls cells division, as well as capsule and flagellum biosynthesis in many bacteria, could be acetylated

through both enzymatic acetylation (*226*) and chemical acetylation (*227*), which inhibits its DNA binding activity. Similarly, PhoP, another highly conserved RR in bacteria that regulate many aspects of bacterial physiology including virulence, is dynamically acetylated at K201 that regulates its DNA binding ability (*228*). Fatty acids could be sensed through non-covalent binding mechanism as well. Virulence regulator HilD in *Salmonella* could bind to long-chain fatty acids and attenuate its DNA binding activity (*204*). *Vibrio cholerae* master virulence regulator ToxT could bind to *cis*-palmitoleate, which would affect ToxT DNA binding and reduce expression of toxin-coregulated pilus and cholera toxin in *V. cholerae* (*229*, *230*). All the examples above highlight the diverse mechanisms that intestinal bacteria use to sense and respond to the same sets of environmental signals, fatty acids.

Regulatory functions of acylation are also conserved in eukaryotes. Of note, acetylation has been well described as a post-translational modification that dynamically regulate cell physiology. The best example may be the 'histone code', which started with the discovery of acetylation on histone tails (*231*) and identification of Gcn5 family histone acetyltransferase (*232*). Dynamic acetylation on histone proteins regulates gene transcription, and is involved in inflammatory diseases (*233*), cancer (*234*), and other disorders (*235–237*). Interestingly, acetylation of histones appears to be responsive to acetyl-CoA level and regulates growth of budding yeast (*238*). Another example is the acetylation of p53, which regulates its DNA binding activity and transcription of p-53-responsive genes (*239*). With the advance of detection methods, many novel types of acylation have been reported, including propionylation (*240*), butyrylation (*240*), 2-hydroxyisobutyrylation (*241*), succinylation (*242*), malonylation (*242*), glutarylation (*243*), crotonylation (*244*), and β-hydroxybutyrylation (*245*), which collectively may regulate signal-dependent gene activation and metabolic stress (*246*, *247*). Some of these modifications are conserved in bacteria (*248*, *249*), and future studies on the functional consequences of these acylations in bacteria will help

elucidate mechanisms metabolic sensing by microbiota and guide metabolic approaches to enhance gut microbiota homeostasis and host health.

Appendix

Appendix Table 1. Bacterial strains used in this thesis. Strains are from Hang lab or generated in the thesis unless otherwise noted.

No.	Organism	Genotype	Plasmid	Antibiotic Resistance	Source	Note
2	*Salmonella typhimurium* 14028	WT	None	None	Craig Altier Lab	
3	*Salmonella typhimurium* 14028	*hilA::kan*	None	Kan	ORF deletion library	Kan:Kanamycin
4	*Salmonella typhimurium* 14028	*hilA::kan*	pBAD-HilA-HA-His	Amp, Kan		Amp: Ampicillin
5	*Salmonella typhimurium* 14028	*hilA::kan*	pBAD-HilAK90A-HA-His	Amp, Kan		
6	*Salmonella typhimurium* 14028	*hilA::kan*	pBAD-HilAK231A-HA-His	Amp, Kan		
7	*Salmonella typhimurium* 14028	*hilA::kan*	pBAD-HilAK324A-HA-His	Amp, Kan		
8	*Salmonella typhimurium* 14028	*hilA::kan*	pBAD-HilAK456A-HA-His	Amp, Kan		
9	*Salmonella typhimurium* 14028	*hilA::kan*	pBAD-HilAK533A-HA-His	Amp, Kan		
10	*Salmonella typhimurium* 14028	*hilA::kan*	pBAD-HilAK57A-HA-His	Amp, Kan		
11	*Salmonella typhimurium* 14028	*hilA::kan*	pBAD-HilAK90X-HA-His , pSupAR-MbPylRS-YAS-salmonella	Amp, Kan, Chlor		X: amber codon, tag
12	*Salmonella typhimurium* 14028	*hilA::kan*	pBAD-HilAK231X-HA-His , pSupAR-MbPylRS-YAS-salmonella	Amp, Kan, Chlor		Chlor: Chloramphenicol

13	*Salmonella typhimurium* 14028	*hilA::kan*	pBAD-HilAK324X-HA-His , pSupAR-MbPylRS-YAS-salmonella	Amp, Kan, Chlor		
14	*Salmonella typhimurium* 14028	*hilA::kan*	pBAD-HilAK456X-HA-His , pSupAR-MbPylRS-YAS-salmonella	Amp, Kan, Chlor		
15	*Salmonella typhimurium* 14028	*hilA::kan*	pBAD-HilAK533X-HA-His , pSupAR-MbPylRS-YAS-salmonella	Amp, Kan, Chlor		
16	*Salmonella typhimurium* 14028	*hilA::kan*	pBAD-HilAK57X-HA-His , pSupAR-MbPylRS-YAS-salmonella	Amp, Kan, Chlor		
33	*Escherichia coli*	DH5alpha	pSupAR-MbPylRS-YAS-salmonella	Chlor	plasmid from Peng R. Chen Lab	
34	*Escherichia coli*	DH5alpha	pBAD-HilA-HA-His	Amp		
44	*Salmonella typhimurium* 14028	*hilA*-HA	None	None		
56	*Salmonella typhimurium* 14028	WT	pCas9 , pKD46 $Temp^S$	Chlor, Amp		pKD46 Temperature sensitive (30C)
61	*Salmonella typhimurium* 14028	*hilA*-HA	pCas9 , pKD46 $Temp^S$	Chlor, Amp		
64	*Salmonella typhimurium* 14028	*hilA*K57A, g183t-HA	None	None		
65	*Salmonella typhimurium* 14028	*hilA*K57X, g183t-HA	None	None		
66	*Salmonella typhimurium* 14028	*hilA*K90A, c243g-HA	None	None		
67	*Salmonella typhimurium* 14028	*hilA*K90X, c243g-HA	None	None		

68	*Salmonella typhimurium* 14028	*hilA*K231A ,c672g-HA	None	None		
69	*Salmonella typhimurium* 14028	*hilA*K231X ,c672g-HA	None	None		
74	*Salmonella typhimurium* 14028	*hilA::kan*	pBAD-HilA-HA-His_Pat-Flag	Amp, Kan		
75	*Salmonella typhimurium* 14028	*hilA::kan*	pBAD-HilA-HA-His_CobB-Flag	Amp, Kan		
81	*Salmonella typhimurium* 14028	*hilA*K57X, g183t-HA	pSupAR-MbPylRS-YAS-salmonella	Chlor		
82	*Salmonella typhimurium* 14028	*hilA*K90X, c243g-HA	pSupAR-MbPylRS-YAS-salmonella	Chlor		
83	*Salmonella typhimurium* 14028	*hilA*K231X ,c672g-HA	pSupAR-MbPylRS-YAS-salmonella	Chlor		
84	*Salmonella typhimurium* 14028	*hilA*-HA	pSupAR-MbPylRS-YAS-salmonella	Chlor		
91	*Salmonella typhimurium* 14028	*hilA*K324X ,g987a-HA	pSupAR-MbPylRS-YAS-salmonella	Chlor		
92	*Salmonella typhimurium* 14028	*hilA*K456X -HA	pSupAR-MbPylRS-YAS-salmonella	Chlor		
93	*Escherichia coli*	DH5alpha	pCas9	Chlor	plasmid from Luciano Marraffini Lab	
94	*Escherichia coli*	DH5alpha	pCas9 (pWJ297)	Chlor	plasmid from Luciano Marraffini Lab	
99	*Salmonella typhimurium* 14028	*hilA*K527X -HA	pSupAR-MbPylRS-YAS-salmonella	Chlor		
100	*Salmonella typhimurium* 14028	*hilA*K533X ,c1575a-HA	pSupAR-MbPylRS-YAS-salmonella	Chlor		
109	*Salmonella typhimurium* 14028	*hilA*-HA	pKD46 $Temp^S$	Amp		

115	*Escherichia coli*	BL21(DE3)	pBAD-HilA-HA-His	Amp		
190	*Salmonella typhimurium* 14028	*hilA::kan*	pBAD-HilAK90X-HA-His , pSupAR-MmPylRS-WT-salmonella	Amp, Kan, Chlor		
191	*Salmonella typhimurium* 14028	*hilA::kan*	pBAD-HilAK231X-HA-His , pSupAR-MmPylRS-WT-salmonella	Amp, Kan, Chlor		
192	*Salmonella typhimurium* 14028	*hilA::kan*	pBAD-HilAK324X-HA-His , pSupAR-MmPylRS-WT-salmonella	Amp, Kan, Chlor		
193	*Salmonella typhimurium* 14028	*hilA::kan*	pBAD-HilAK456X-HA-His , pSupAR-MmPylRS-WT-salmonella	Amp, Kan, Chlor		
194	*Salmonella typhimurium* 14028	*hilA::kan*	pBAD-HilAK533X-HA-His , pSupAR-MmPylRS-WT-salmonella	Amp, Kan, Chlor		
196	*Salmonella typhimurium* 14028	WT	pSupAR-MbPylRS-YAS-salmonella	Chlor		
197	*Salmonella typhimurium* 14028	*hilA::kan*	pSupAR-MbPylRS-YAS-salmonella	Kan, Chlor		

Appendix Table 2. Primers used for qRT-PCR and ChIP-qPCR in this thesis.

No.	Primer Name	Sequence (5' to 3')
73	invF qPCR F	TCCTGAGTTTCGCGCTATTT
74	invF qPCR R	GTAACAGCGCCAGTACCTTAT
75	prgH qPCR F	ACAGCAGGCGTTACCTTATTC
76	prgH qPCR R	AATTGACGGGCTCTGAGTATTT
157	spaO qPCR F	CCGACCAATGCTGAACTTAAC
158	spaO qPCR R	TTCATGGATCTCAACGCCTAAG
159	orgB qPCR F	ATCAGACAATGGCCTGGAAG
160	orgB qPCR R	AAATCCCTTAGCCACTCATCC
199	PprgH qPCR F	AGAACGACAGACATCGCTAAC
200	PprgH qPCR R	CCGTTCAGTGAGCTGTTAAGTA
201	PinvF qPCR F	GGGCGGCATCAGTTTCATAA
202	PinvF qPCR R	GCATAGTTGTCAGCACCAGTTA

Appendix 2.1. Table of identified protein targets of Alk-3 in Stm through LFQ proteomic analysis.

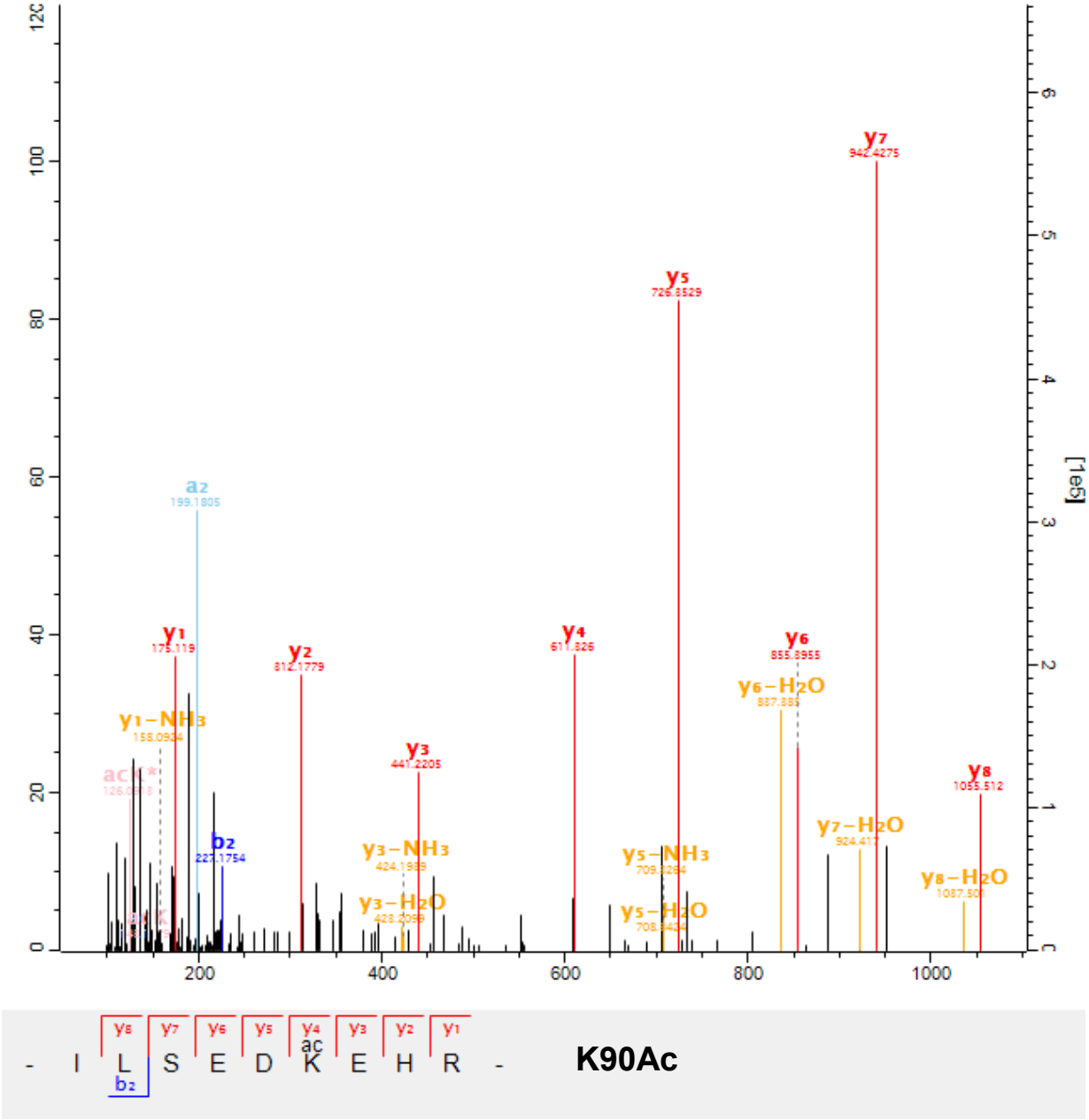
Protein IDs	Category	Description
hilA	SPI-1	Invasion protein regulator
sipD	SPI-1	Translocation machinery component
sopB	SPI-1	Secreted effector protein
sipA	SPI-1	Secreted effector protein
sipC	SPI-1	Cell invasion protein SipC
sicA	SPI-1	Secretion chaperone
fabD	Metabolism	Malonyl CoA-acyl carrier protein transacylase
gltA	Metabolism	Citrate synthase
citD	Metabolism	Citrate lyase acyl carrier protein
eno	Metabolism	Enolase
panD	Metabolism	Aspartate 1-decarboxylase
gcvH	Metabolism	Glycine cleavage system H protein
rpiA	Metabolism	Ribose-5-phosphate isomerase A
pgk	Metabolism	Phosphoglycerate kinase
sucC	Metabolism	Succinate--CoA ligase [ADP-forming] subunit beta

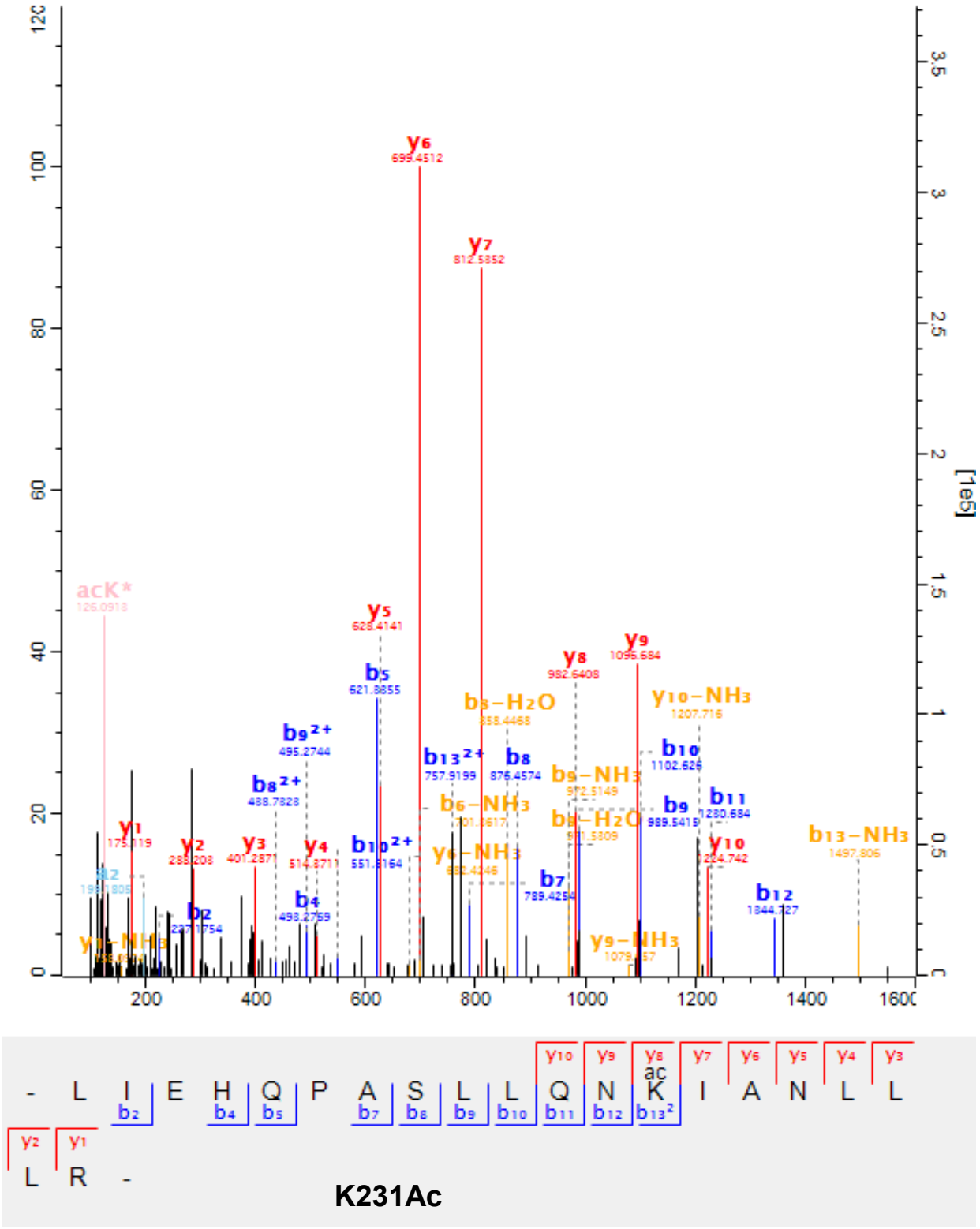
acnB	Metabolism	Aconitate hydratase B
asnB	Metabolism	Asparagine synthetase B
aspA	Metabolism	Aspartate ammonia-lyase
yqhD	Metabolism	Putative alcohol dehydrogenase
ribE	Metabolism	Riboflavin synthase subunit alpha
maeB	Metabolism	Malic enzyme
pykF	Metabolism	Pyruvate kinase
dctA	Metabolism	C4-dicarboxylate transport protein
tktA	Metabolism	Transketolase
pykA	Metabolism	Pyruvate kinase
dkgA	Metabolism	2,5-diketo-D-gluconate reductase A
pepD	Metabolism	Aminoacyl-histidine dipeptidase
nuoB	Metabolism	NADH-quinone oxidoreductase subunit B
glyA	Metabolism	Serine hydroxymethyltransferase
trxA	Metabolism	Thioredoxin
icdA	Metabolism	Isocitrate dehydrogenase [NADP]
guaC	Metabolism	GMP reductase
talB	Metabolism	Transaldolase

phoN	Metabolism	Acid phosphatase
gapA	Metabolism	Glyceraldehyde-3-phosphate dehydrogenase
cysK	Metabolism	Cysteine synthase
prsA	Metabolism	Ribose-phosphate pyrophosphokinase
tdcE	Metabolism	Pyruvate formate-lyase 4/2-ketobutyrate formate-lyase
adhE	Metabolism	Aldehyde-alcohol dehydrogenase
tyrS	Metabolism	Tyrosine--tRNA ligase
dnaK	Others	Chaperone protein DnaK
ydhD	Others	Glutaredoxin
grxC	Others	Glutaredoxin
yghA	Others	Oxidoreductase
nifU	Others	Iron-sulfur cluster assembly scaffold protein IscU
groEL	Others	60 kDa chaperonin
yfiD	Others	Autonomous glycyl radical cofactor
hns	Others	DNA-binding protein H-NS
katE	Others	Catalase
tpx	Others	Thiol peroxidase
erpA	Others	Iron-sulfur cluster insertion protein ErpA

greA	Others	Transcription elongation factor GreA
STM14_5121	Others	Putative inner membrane protein
yfiE	Others	Putative LysR family transcriptional regulator
pmbA	Others	Peptidase PmbA
yhgI	Others	Fe/S biogenesis protein NfuA

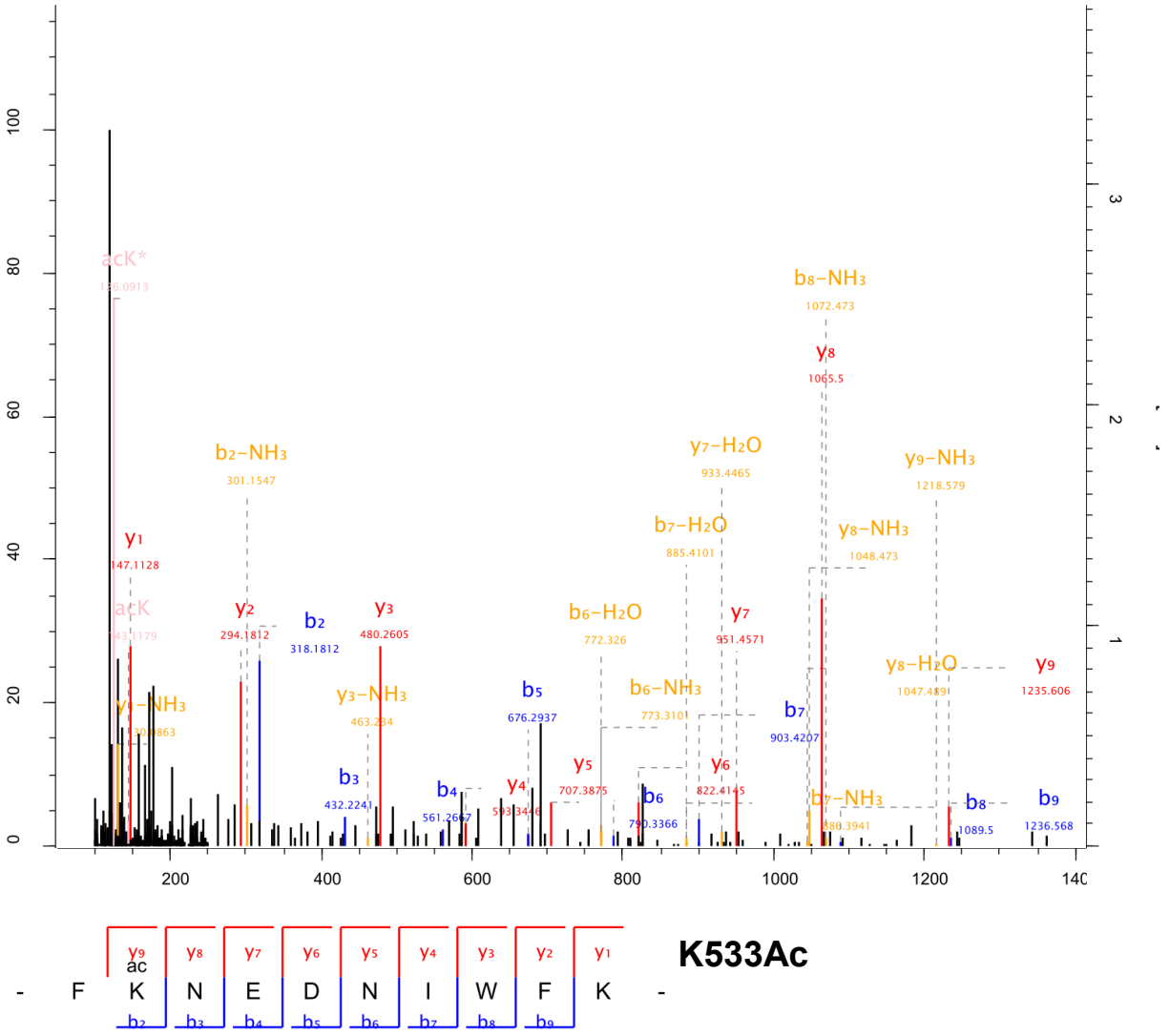
Appendix 2.2. MS/MS spectrum of detected acylated peptides of HilA.





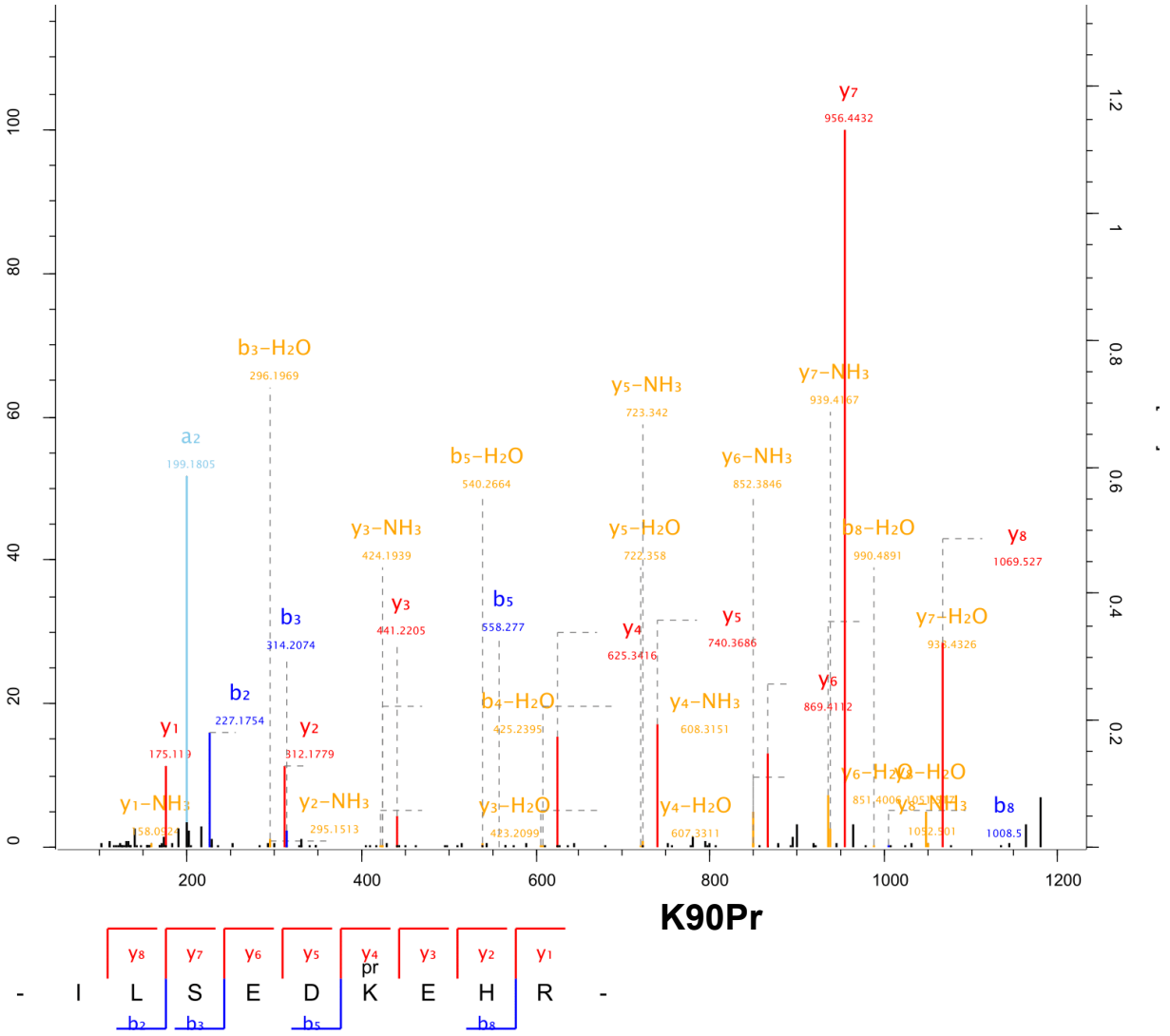
- L I E H Q P A S L L Q N K I A N L L
L R -

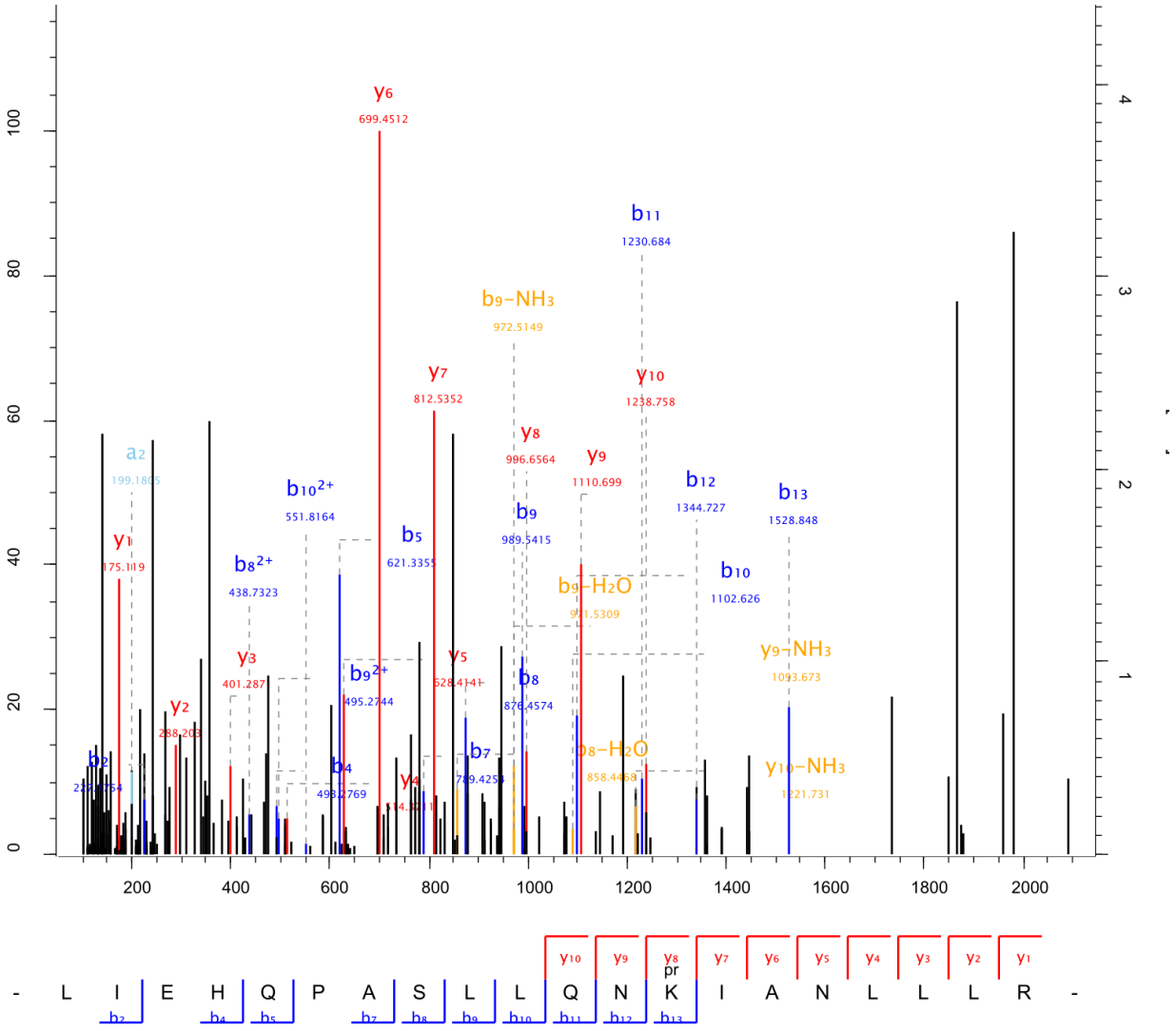
K231Ac

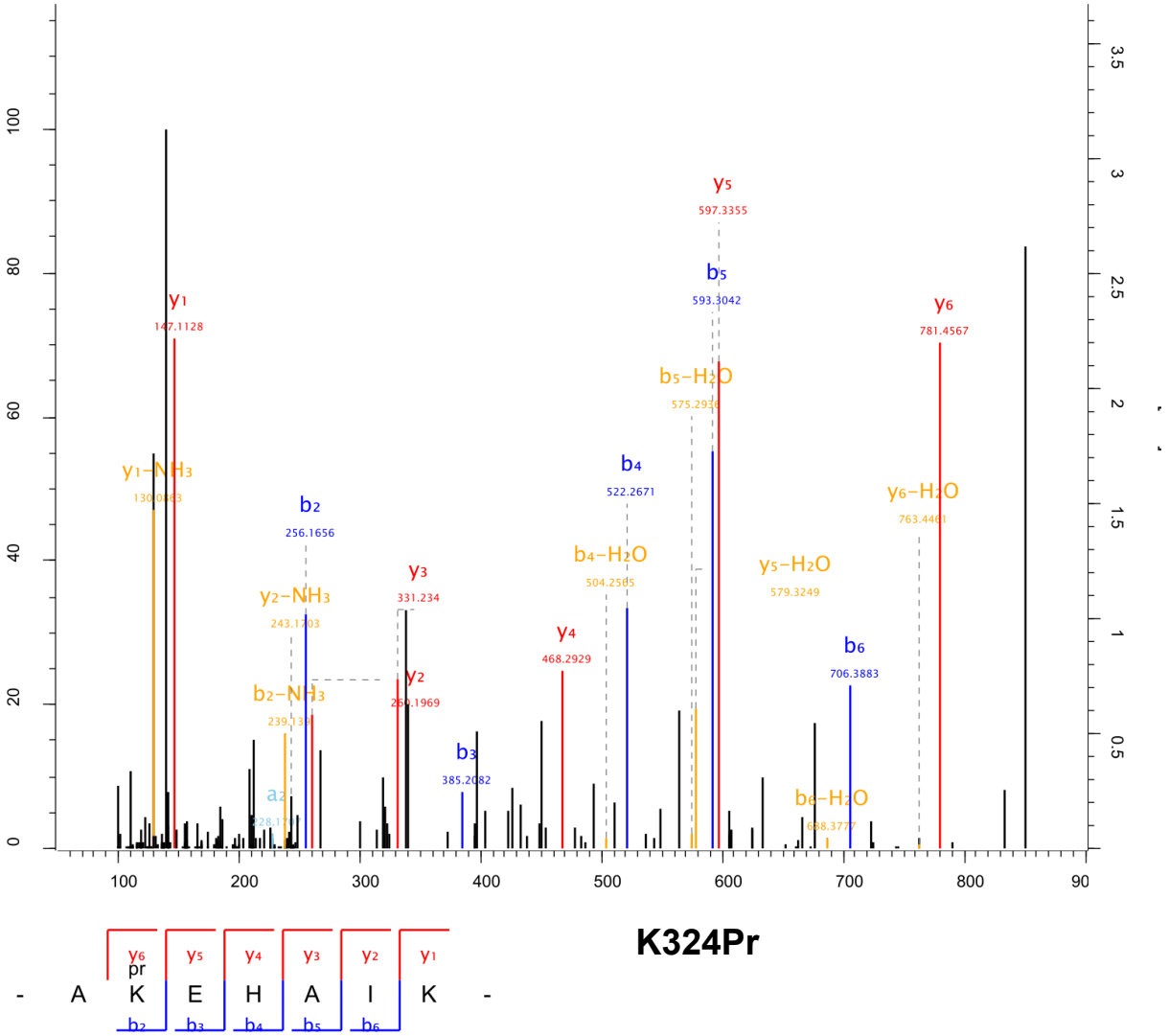

acK*
b8−NH3
1072.473
y8
1065.5
b2−NH3
301.1547
y7−H2O
933.4465
y9−NH3
1218.579
y1
147.1128
b7−H2O
885.4101
y8−NH3
1048.473
acK
y2
294.1812
b2
318.1812
y3
480.2605
b6−H2O
772.326
y7
951.4571
y8−H2O
1047.489
y9
1235.606
y3−NH3
463.234
b5
676.2937
b6−NH3
773.3101
b7
903.4207
b3
432.2241
b4
561.2667
y4
593.3446
y5
707.3875
y6
822.4145
b6
790.3366
b7−NH3
888.3941
b8
1089.5
b9
1236.568
200
400
600
800
1000
1200
0
20
40
60
80
100
1
2
3

y9 y8 y7 y6 y5 y4 y3 y2 y1
ac
- F K N E D N I W F K -
b2 b3 b4 b5 b6 b7 b8 b9

K533Ac

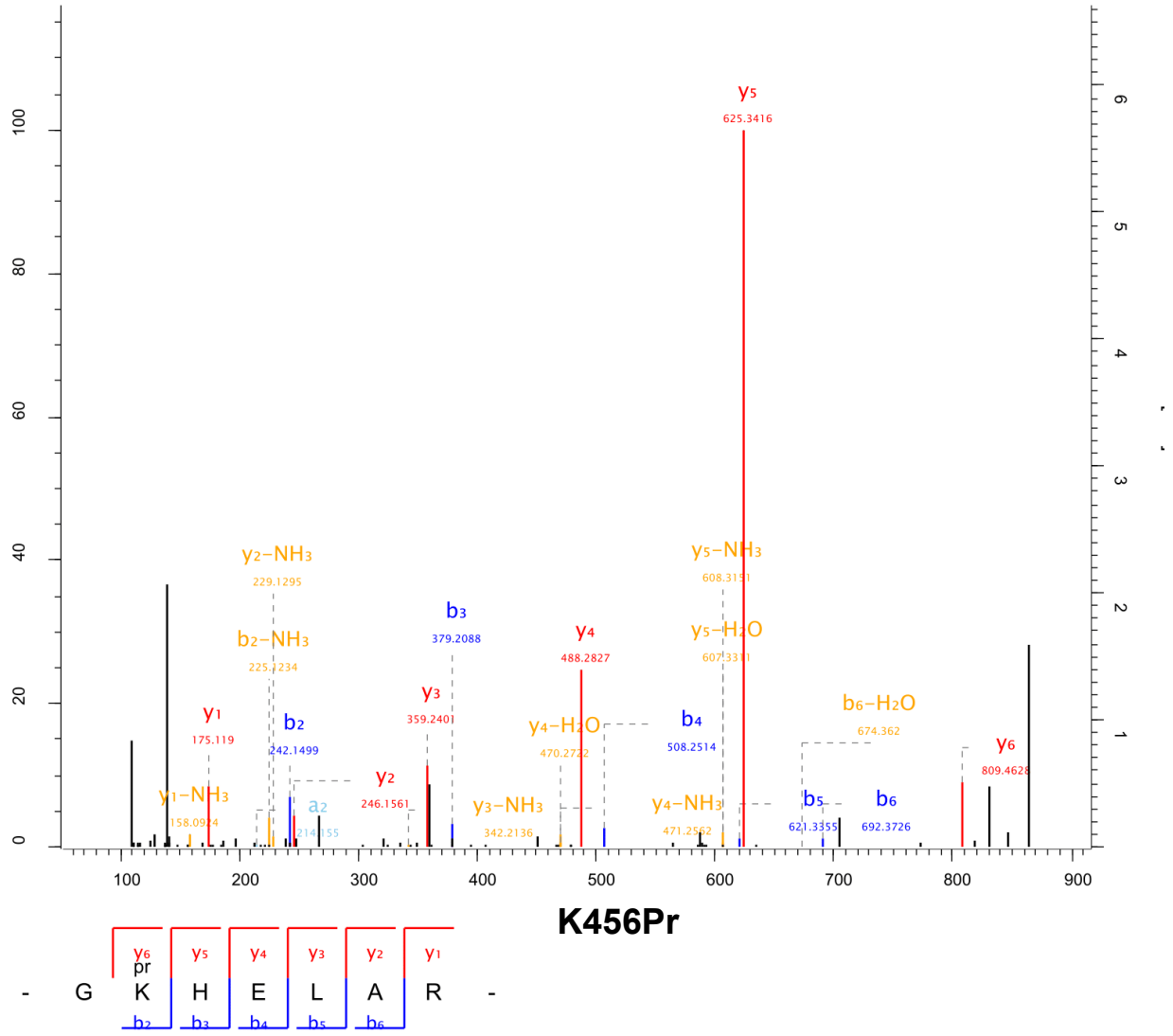

y7
956.4432
b3−H2O
296.1969
y5−NH3
723.342
y7−NH3
939.4167
a2
199.1805
b5−H2O
540.2664
y6−NH3
852.3846
y3−NH3
424.1939
y5−H2O
722.358
b8−H2O
990.4891
y8
1069.527
y3
441.2205
b5
558.277
b3
314.2074
y4
625.3416
y5
740.3686
y7−H2O
936.4326
b2
227.1754
y6
869.4112
y1
175.119
y2
312.1779
b4−H2O
425.2395
y4−NH3
608.3151
y6−H2O
y8−H2O
y1−NH3
y2−NH3
295.1513
y3−H2O
423.2099
y4−H2O
607.3311
y8−NH3
1052.501
b8
1008.5
K90Pr
y8 y7 y6 y5 y4 pr y3 y2 y1
- I L S E D K E H R -
b2 b3 b5 b8

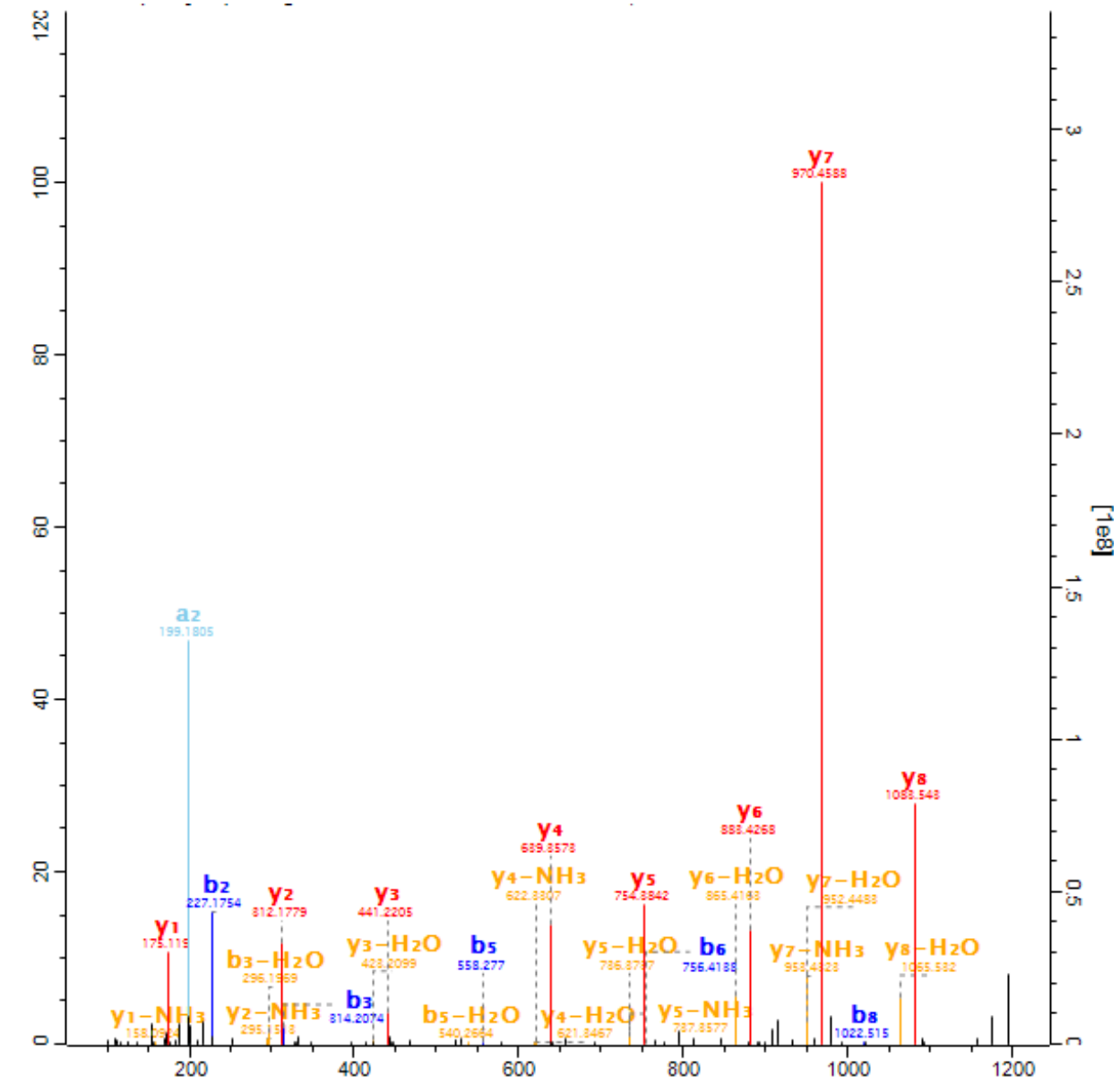

y6
699.4512
b11
1230.684
b9−NH3
972.5149
y7
812.5352
y10
1238.758
y8
996.6564
y9
1110.699
b12
1344.727
b13
1528.848
a2
199.1805
b10 2+
551.8164
b9
989.5415
y1
175.1119
b5
621.3355
b8 2+
438.7323
b9−H2O
971.5309
b10
1102.626
y3
401.287
y5
528.4141
b9 2+
495.2744
b8
876.4574
y9−NH3
1093.673
y2
288.2033
b7
789.4256
b4
498.2769
y4
b8−H2O
858.4468
y10−NH3
1221.731
200
400
600
800
1000
1200
1400
1600
1800
2000
0
20
40
60
80
100
1
2
3
4
- L I E H Q P A S L L Q N K I A N L L L R -
pr
K231Pr


y5
597.3355
b5
593.3042
y1
147.1128
y6
781.4567
b5−H2O
575.2936
y1−NH3
130.0863
b4
522.2671
b2
256.1656
y6−H2O
763.4461
b4−H2O
504.2565
y5−H2O
579.3249
y3
331.234
y2−NH3
243.1703
y4
468.2929
b6
706.3883
y2
269.1969
b2−NH3
239.1347
b3
385.2082
a2
228.1342
b6−H2O
688.3777
100
200
300
400
500
600
700
800
90
0
20
40
60
80
100
0.5
1
1.5
2
2.5
3
3.5

y6
pr
y5
y4
y3
y2
y1
-
A
K
E
H
A
I
K
-
b2
b3
b4
b5
b6

K324Pr


y5
625.3416
y2−NH3
229.1295
b2−NH3
225.1234
b3
379.2088
y4
488.2827
y5−NH3
608.3151
y5−H2O
607.3311
y1
175.119
b2
242.1499
y3
359.2401
y4−H2O
470.2722
b4
508.2514
b6−H2O
674.362
y6
809.4628
y1−NH3
158.0924
a2
214.1155
y2
246.1561
y3−NH3
342.2136
y4−NH3
471.2562
b5
621.3355
b6
692.3726
100
200
300
400
500
600
700
800
900
0
20
40
60
80
100
1
2
3
4
5
6
K456Pr
- G K H E L A R -
pr
y6 y5 y4 y3 y2 y1
b2 b3 b4 b5 b6

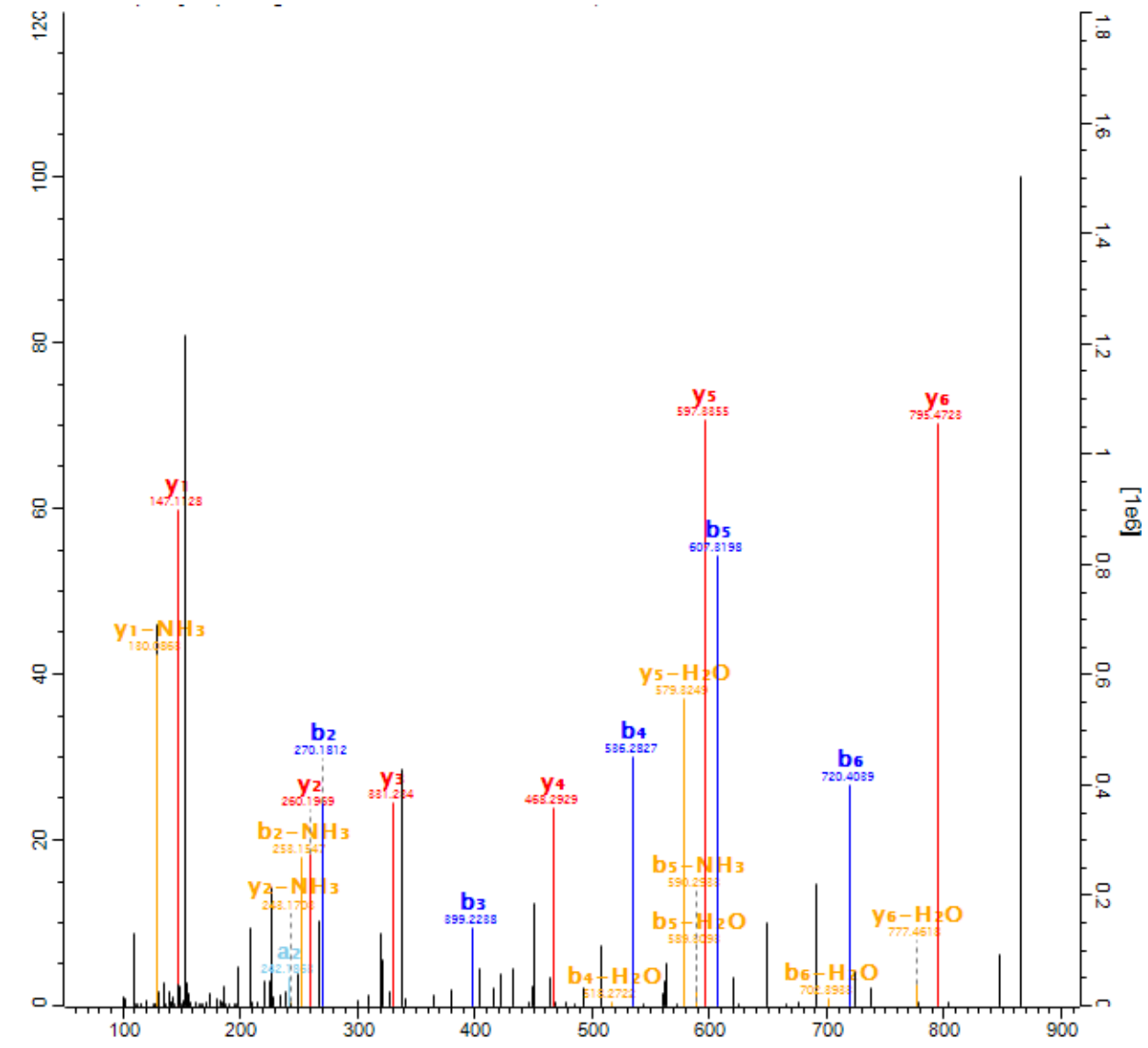


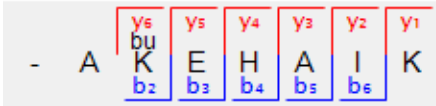
- I L S E D K E H R -

y8 y7 y6 y5 y4 y3 y2 y1

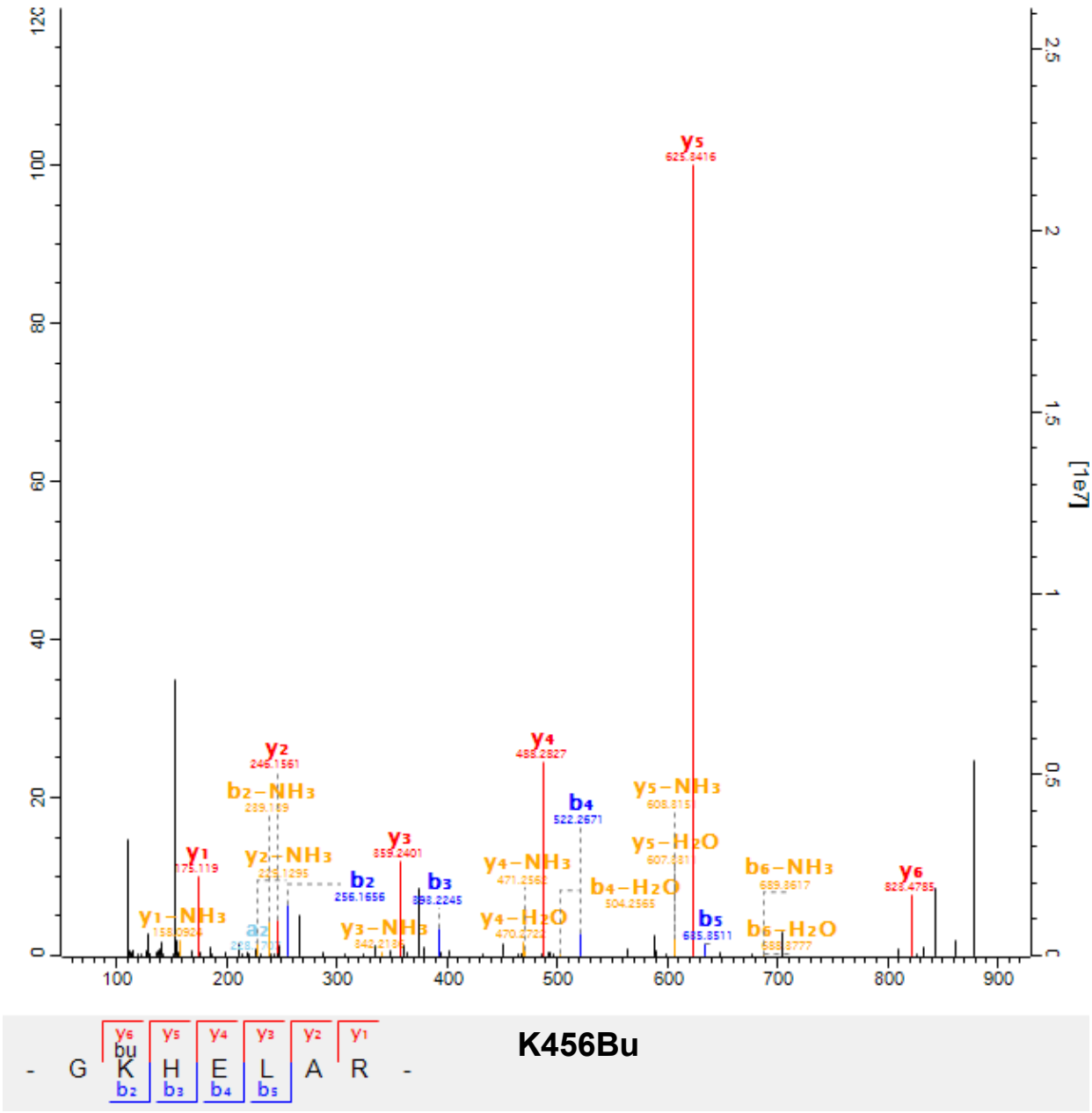
b2 b3 b5 bu b6 b8

K90Bu

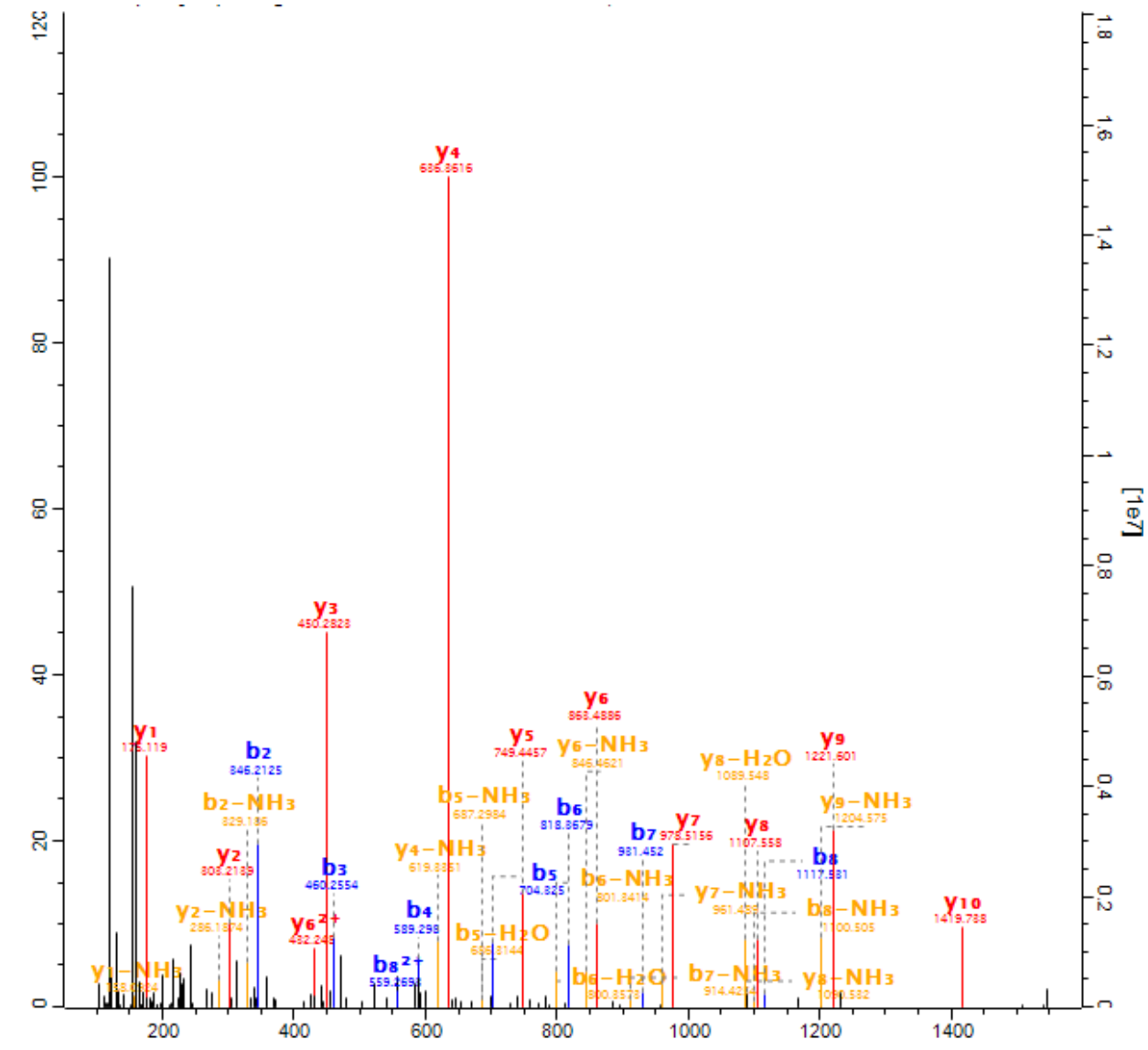

y5
597.8855
y6
795.4723
y1
147.1128
b5
607.8198
y1−NH3
180.0865
y5−H2O
579.8249
b2
270.1812
b4
586.2827
b6
720.4089
y2
260.1969
y3
331.204
y4
468.2929
b2−NH3
253.1547
b5−NH3
590.2933
y2−NH3
243.1703
b3
399.2288
b5−H2O
589.3093
a2
242.1856
y6−H2O
777.4618
b4−H2O
518.2722
b6−H2O
702.3983
[1e6]

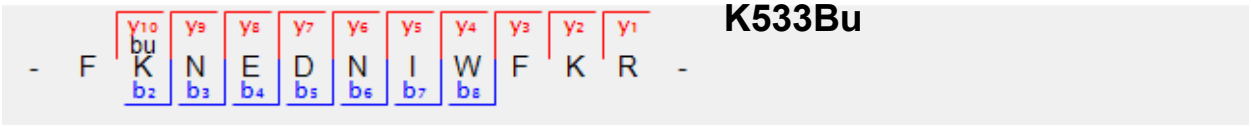

y6
bu
y5
y4
y3
y2
y1
-
A
K
E
H
A
I
K
-
b2
b3
b4
b5
b6

K324Bu


y5
625.8416
y4
488.2827
y2
246.1561
b2−NH3
289.169
b4
522.2671
y5−NH3
608.315
y5−H2O
607.331
y1
175.119
y2−NH3
229.1295
y3
359.2401
y4−NH3
471.2562
b6−NH3
689.3617
y6
829.4785
b2
256.1656
b3
358.2245
b4−H2O
504.2565
y1−NH3
158.0924
y3−NH3
342.2135
y4−H2O
470.2722
b5
685.3511
b6−H2O
671.3777
[1e7]
- G K H E L A R -
y6 y5 y4 y3 y2 y1
bu
b2 b3 b4 b5

K456Bu


y4
y3
y6
y1
y5
y6−NH3
y9
b2
y8−H2O
b2−NH3
b5−NH3
b6
y7
y8
y9−NH3
b7
y2
y4−NH3
b3
b5
b8
y7−NH3
y2−NH3
b4
b6−NH3
b8−NH3
y10
y6 2+
b5−H2O
y1−NH3
b8 2+
b6−H2O
b7−NH3
y8−NH3
[1e7]
0
20
40
60
80
100
120
200
400
600
800
1000
1200
1400


- F K N E D N I W F K R -
bu
y10 y9 y8 y7 y6 y5 y4 y3 y2 y1
b2 b3 b4 b5 b6 b7 b8

K533Bu

Appendix 3.1. List of identified protein targets of Alk-16 in Stm through LFQ proteomic analysis. SPI-1 proteins are labeled in red.

-log(*P*-value)	Difference	Gene ID	Description
5.295718	6.580152	sipA	Secreted effector protein
4.596443	6.067259	sopB	Inositol phosphate phosphatase SopB
4.13649	5.121576	sipC	Cell invasion protein SipC
6.095165	4.877037	sipD	Cell invasion protein SipD
2.53398	3.625011	sicA	Chaperone protein SicA
1.748172	3.163045	sopE2	Guanine nucleotide exchange factor sopE2
3.285982	2.883593	sptP	Secreted effector protein SptP
3.469847	2.314409	hilA	Transcriptional regulator HilA
1.132367	1.464825	sopA	E3 ubiquitin-protein ligase SopA
1.763296	7.616678	lpp1	Major outer membrane lipoprotein 1
2.750863	6.559837	acpP	Acyl carrier protein
3.096859	6.383107	pgk	Phosphoglycerate kinase
4.407329	5.763006	cadA	Lysine decarboxylase
5.138377	5.438326	ackA	Acetate kinase
3.090379	5.404028	pepD	Aminoacyl-histidine dipeptidase
3.898353	5.334023	osmE	DNA-binding transcriptional activator OsmE
4.931134	5.27616	yqhD	Putative alcohol dehydrogenase
4.530829	5.196304	arcA	Arginine deiminase
4.57988	5.141149	hns	DNA-binding protein H-NS
4.250792	4.936145	acrA	Acridine efflux pump

4.366618	4.865341	tpiA	Triosephosphate isomerase
5.431065	4.8517	sspA	Stringent starvation protein A
5.223858	4.850132	alaS	Alanine--tRNA ligase
4.132562	4.770939	manZ	Mannose-specific PTS system protein IID
5.108081	4.728649	yaeH	UPF0325 protein YaeH
3.584452	4.693852	icdA	Isocitrate dehydrogenase [NADP]
4.4016	4.628246	groL	60 kDa chaperonin
3.727492	4.582752	STM474_1440	Glutaredoxin
2.524992	4.523128	gltA	Citrate synthase
4.308787	4.522205	STMMW_07561	Phosphoglucomutase
3.714703	4.451301	STMMW_26761	ClpB protein (Heat shock protein f84.1)
3.93975	4.447678	ybis	Putative L,D-transpeptidase YbiS
2.529688	4.382948	rpiA	Ribose-5-phosphate isomerase A
5.23199	4.363451	sucC	Succinyl-CoA ligase [ADP-forming] subunit beta
1.892263	4.302311	dnaK	Chaperone protein DnaK
4.200961	4.257231	galU	UTP--glucose-1-phosphate uridylyltransferase subunit GalU
5.819958	4.253464	kdsA	2-dehydro-3-deoxyphosphooctonate aldolase
4.314986	4.210491	fabI	Enoyl-[acyl-carrier-protein] reductase [NADH] FabI

5.031606	4.200387	ptsI	Phosphoenolpyruvate-protein phosphotransferase
3.185759	4.171696	eno	Enolase
3.194797	4.155042	pps	Phosphoenolpyruvate synthase
3.803548	4.128049	STM474_1254	Putative lipoprotein
4.844141	4.1238	nuoL	NADH dehydrogenase subunit L
3.529328	4.120824	deoB	Phosphopentomutase
1.605582	4.050425	fbaB	Fructose-bisphosphate aldolase
3.909793	3.985844	acnB	Aconitate hydratase 2
4.419057	3.923382	adk	Adenylate kinase
4.563996	3.911688	minD	Site-determining protein
2.279087	3.896372	frdA	Fumarate reductase, flavoprotein subunit
1.51704	3.856174	slyB	Outer membrane lipoprotein SlyB
2.793875	3.838635	sucD	Succinyl-CoA ligase [ADP-forming] subunit alpha
3.937432	3.814405	glpK	Glycerol kinase
5.352378	3.801908	sdhB	Succinate dehydrogenase iron-sulfur subunit
4.735745	3.737444	nuoF	NADH-quinone oxidoreductase subunit F
2.224313	3.735641	frdB	Fumarate reductase iron-sulfur subunit
3.373764	3.731722	ucpA	Oxidoreductase UcpA
5.189995	3.686877	pta	Phosphate acetyltransferase
4.466336	3.670825	STMMW_42751	Aspartate ammonia-lyase

1.878152	3.666236	clpX	ATP-dependent Clp protease ATP-binding subunit ClpX
4.74674	3.658666	crp	cAMP-activated global transcriptional regulator CRP
5.391827	3.653837	purA	Adenylosuccinate synthetase
3.12594	3.630978	glpT	sn-glycerol-3-phosphate transporter
1.801252	3.627244	dcuA	Anaerobic C4-dicarboxylate transporter
3.527084	3.621531	STM4 74_18 15	Glutamate dehydrogenase
1.540743	3.597588	cyoB	Cytochrome o ubiquinol oxidase subunit I
2.761849	3.541516	uspF	Universal stress protein F
3.553989	3.507147	hldD	ADP-L-glycero-D-manno-heptose-6-epimerase
4.003492	3.49936	engD	Ribosome-binding ATPase YchF
5.152183	3.481234	gnd	6-phosphogluconate dehydrogenase, decarboxylating
1.469702	3.451141	deoD	Purine nucleoside phosphorylase DeoD-type
5.51632	3.429234	pheT	Phenylalanine--tRNA ligase beta subunit
3.090018	3.420237	tufA	Elongation factor Tu
2.059706	3.416838	fabB	3-oxoacyl-(Acyl carrier protein) synthase I
5.177576	3.342725	hldE	Bifunctional protein HldE
1.920311	3.330688	grxA	Glutaredoxin-1
4.319432	3.325662	glmS	Glutamine--fructose-6-phosphate aminotransferase [isomerizing]
5.492707	3.313935	eutB	Ethanolamine ammonia-lyase heavy chain
4.553025	3.278964	asnS	Asparagine--tRNA ligase

1.8994	3.271589	STMMW_30291	Fructose 1,6-bisphosphate aldolase
3.768561	3.233645	argS	Arginine--tRNA ligase
1.699932	3.221724	grxC	Glutaredoxin 3
4.27221	3.209486	talB	Transaldolase B
4.73276	3.207051	malK	Maltose/maltodextrin import ATP-binding protein MalK
4.079357	3.204319	thrS	Threonine--tRNA ligase
2.039398	3.199423	ybjP	Putative lipoprotein
3.932196	3.196885	yliJ	Glutathione s-transferase family protein
3.673113	3.174411	dapA	4-hydroxy-tetrahydrodipicolinate synthase
1.864859	3.173382	atpF	ATP synthase subunit b
1.535292	3.172801	yghA	Oxidoreductase
3.696875	3.170086	fbp	Fructose-1,6-bisphosphatase class 1
4.018392	3.146297	nifU	Scaffold protein
3.670025	3.113139	leuS	Leucine--tRNA ligase
3.72587	3.105	gst	Glutathionine S-transferase
2.026667	3.103836	pykF	Pyruvate kinase
3.001963	3.051551	torC	Trimethylamine N-oxide reductase cytochrome c-like subunit
3.44676	3.030305	sucA	Alpha-ketoglutarate decarboxylase
4.445307	3.025424	STMMW_33261	Uncharacterized protein
5.158218	2.994371	yebC	Probable transcriptional regulatory protein YebC

1.541114	2.987154	rfbH	Lipopolysaccharide biosynthesis protein RfbH
5.005963	2.973447	rpsA	30S ribosomal protein S1
4.120244	2.964425	uspA	Universal stress protein A
5.918274	2.956854	pfkB	Phosphofructokinase
1.509733	2.92706	ygaM	YgaM protein
3.131615	2.923406	fljB	Phase 2 flagellin
0.805356	2.919812	yajQ	UPF0234 protein YajQ
1.736869	2.900527	yfcZ	Uncharacterized protein
1.524307	2.89203	arcB	Ornithine carbamoyltransferase, catabolic
2.419709	2.885463	gudD	D-glucarate dehydratase
1.974192	2.870841	ndk	Nucleoside diphosphate kinase
2.177368	2.868642	secB	Protein-export protein SecB
3.435057	2.864764	rof	ROF protein
1.783416	2.84557	tkt	Transketolase
3.230551	2.839833	ileS	Isoleucine--tRNA ligase
1.787603	2.824103	STMMW_07911	Succinate dehydrogenase flavoprotein subunit
1.962253	2.818172	garR	Tartronate semialdehyde reductase
2.222904	2.817016	prs	Ribose-phosphate pyrophosphokinase
3.144408	2.814523	tyrS	Tyrosine--tRNA ligase
2.982671	2.80989	pykA	Pyruvate kinase II
3.905117	2.784853	mtlA	PTS system mannitol-specific EIICBA component

1.608646	2.781614	STMMW_12061	3-oxoacyl-[acyl-carrier-protein] synthase 2
1.87022	2.756329	STM474_4663	Carbamate kinase
2.606342	2.742773	citT	Citrate/succinate transport antiport protein
1.661152	2.730909	yieF	Putative oxidoreductase
3.582162	2.720367	aceE	Pyruvate dehydrogenase E1 component
4.148295	2.717626	bcp	Thioredoxin-dependent thiol peroxidase
2.109469	2.714926	cydB	Cytochrome d ubiquinol oxidase subunit II
3.847165	2.643891	STMMW_11331	Proline dehydrogenase (Proline oxidase)
3.704195	2.642329	STM474_3876	Putative secreted protein
3.262089	2.639904	valS	Valine--tRNA ligase
3.052889	2.612987	hemL	Glutamate-1-semialdehyde 2,1-aminomutase
3.763151	2.606649	tktA	Transketolase
2.879471	2.599024	nuoC	NADH-quinone oxidoreductase subunit C/D
2.844099	2.589397	STM474_2742	ATPase domain protein
1.797721	2.560793	traT	Conjugative transfer surface exclusion protein

2.844748	2.55012	glnA	Glutamine synthetase
4.034687	2.544812	pgi	Glucose-6-phosphate isomerase
1.445654	2.542115	STMMW_18181	L-serine deaminase 1
2.249804	2.521631	serS	Serine--tRNA ligase
3.39602	2.517659	gcvP	Glycine dehydrogenase (decarboxylating)
2.901281	2.436315	aspS	Aspartate--tRNA ligase
1.611601	2.427409	ftn	Ferritin
2.488933	2.419221	kbl	2-amino-3-ketobutyrate coenzyme A ligase
2.331812	2.381566	ygiB	UPF0441 protein YgiB
2.895535	2.37863	recA	Protein RecA
1.327359	2.336936	pnp	Polyribonucleotide nucleotidyltransferase
2.887168	2.330995	prlC	Oligopeptidase A
2.342801	2.317223	pflB	Formate acetyltransferase 1
1.866763	2.307178	cheA	Chemotaxis protein CheA
1.859378	2.286694	malE	Maltose-binding periplasmic protein
1.523602	2.285541	tdcG	L-serine deaminase
1.692734	2.284634	glnB	Nitrogen regulatory protein P-II 1
1.111207	2.279823	atpD	ATP synthase subunit beta
2.402536	2.272444	tsf	Elongation factor Ts
0.785455	2.259772	ompA	Outer membrane protein A
2.023052	2.253075	aspC	Aspartate aminotransferase
2.839981	2.23838	mreB	Rod shape-determining protein MreB

2.553685	2.23786	STMMW_07361	Putative monooxygenase
0.862548	2.23199	spaK	Surface presentation of antigens protein SpaK
0.959748	2.229185	STMMW_24911	NADP-dependent malate dehydrogenase (Decarboxylating)
3.206656	2.226079	rpoD	RNA polymerase sigma factor RpoD
3.614921	2.193863	glyA	Serine hydroxymethyltransferase
3.353257	2.188644	metK	S-adenosylmethionine synthase
3.954331	2.163292	fusA	Elongation factor G
2.321782	2.16064	aldB	Aldehyde dehydrogenase B
1.839531	2.155278	pepP	Proline aminopeptidase II
3.26356	2.147752	pyrG	CTP synthase
1.866638	2.140305	iadA	Isoaspartyl dipeptidase
1.142182	2.135898	lysS	Lysine--tRNA ligase
1.95664	2.134546	rfbG	CDP-glucose 4,6-dehydratase
3.693331	2.134274	hybC	Hydrogenase-2, large subunit
2.000398	2.132262	tig	Trigger factor
2.436818	2.129992	lpdA	Dihydrolipoyl dehydrogenase
1.571246	2.112889	STMMW_24331	Nucleoside permease NupC
1.040718	2.103135	nlpB	Outer membrane protein assembly factor BamC
1.725687	2.099536	guaA	GMP synthase [glutamine-hydrolyzing]

1.642941	2.099028	acnA	Aconitate hydratase
2.815118	2.085762	tsaB	tRNA threonylcarbamoyladenosine biosynthesis protein TsaB
4.52154	2.080546	STMMW_32171	Methyl-accepting chemotaxis protein II
0.962423	2.078973	dcuB	Anaerobic C4-dicarboxylate transporter
3.317019	2.071259	pepN	Aminopeptidase N
2.979041	2.060308	pckA	Phosphoenolpyruvate carboxykinase [ATP]
2.78949	2.05892	sodB	Superoxide dismutase
2.522927	2.04125	nuoG	NADH-quinone oxidoreductase subunit G
2.471814	2.008285	glpD	Glycerol-3-phosphate dehydrogenase
1.620935	2.001814	STM474_3949	Putative glycosyl hydrolase
1.149125	1.991829	pntA	NAD(P) transhydrogenase subunit alpha
1.421445	1.984356	speF	Ornithine decarboxylase
1.487247	1.972696	upp	Uracil phosphoribosyltransferase
1.044659	1.971578	rplK	50S ribosomal protein L11
0.883421	1.966462	nuoB	NADH-quinone oxidoreductase subunit B
2.223961	1.945998	ptsG	PTS system glucose-specific EIICB component
1.602224	1.915981	STMMW_29341	L-serine dehydratase 2 (L-serine deaminase 2)
1.293747	1.886269	malF	Maltose transport system permease protein MalF

1.035021	1.885867	rplE	50S ribosomal protein L5
2.973231	1.876769	typA	GTP-binding protein
1.119039	1.842197	nrdA	Ribonucleoside-diphosphate reductase 1 subunit alpha
1.790733	1.827587	aceF	Dihydrolipoyllysine-residue acetyltransferase
1.754495	1.82602	gyrB	DNA gyrase subunit B
2.68662	1.825929	udp	Uridine phosphorylase
0.778031	1.821534	rplL	50S ribosomal protein L7/L12
2.298098	1.81606	ppc	Phosphoenolpyruvate carboxylase
0.797527	1.800735	tdcE	Pyruvate formate-lyase 4/2-ketobutyrate formate-lyase
0.935107	1.775652	cydA	Cytochrome d terminal oxidase polypeptide subunit I
1.984189	1.760096	asd	Aspartate-semialdehyde dehydrogenase
1.816843	1.747447	fumA	Fumarate hydratase class I, anaerobic
1.756417	1.738461	fdoG	Hypothetical 1 formate dehydrogenase-O, major subunit
4.030025	1.718745	mdh	Malate dehydrogenase
3.111759	1.718309	metG	Methionine--tRNA ligase
1.039016	1.704385	rpoA	DNA-directed RNA polymerase subunit alpha
1.588926	1.650065	maeA	NAD-dependent malic enzyme
1.457687	1.630216	eutM	Ethanolamine utilization protein EutM
2.138438	1.627711	fumC	Fumarate hydratase class II
1.988542	1.624924	pepQ	Xaa-Pro dipeptidase
1.368701	1.589605	glpQ	Glycerophosphodiester phosphodiesterase
2.110415	1.589405	glyS	Glycine--tRNA ligase beta subunit
1.89163	1.559617	zwf	Glucose-6-phosphate 1-dehydrogenase

0.863203	1.543272	rpmB	50S ribosomal protein L28
1.336827	1.536856	atpA	ATP synthase subunit alpha
3.071922	1.524047	STMMW_15371	Uptake hydrogenase-1 large subunit
1.894378	1.495006	rihA	Pyrimidine-specific ribonucleoside hydrolase RihA
2.392272	1.484539	ldhA	D-lactate dehydrogenase
1.004855	1.458355	dapD	2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-succinyltransferase
1.279965	1.455299	ridA	Enamine/imine deaminase
3.241329	1.43513	STM474_4453	Putative methyl-accepting chemotaxis protein
0.957426	1.428329	rplJ	50S ribosomal protein L10
3.127597	1.427683	mglB	D-galactose-binding periplasmic protein
0.730977	1.403804	dctA	Aerobic C4-dicarboxylate transport protein
1.028964	1.401089	cysK	Cysteine synthase A
1.681067	1.392614	nusA	Transcription termination/antitermination protein NusA
2.181387	1.366551	STMMW_28001	Arabinose 5-phosphate isomerase
1.718342	1.315477	yeaG	Putative serine protein kinase
1.675064	1.26836	yifE	UPF0438 protein YifE
1.916112	1.257701	arnC	Undecaprenyl-phosphate 4-deoxy-4-formamido-L-arabinose transferase

1.425214	1.153601	ppa	Inorganic pyrophosphatase
3.203631	1.15095	talA	Transaldolase A
2.113997	1.121765	asnB	Asparagine synthetase

References

1. G. P. Donaldson, S. M. Lee, S. K. Mazmanian, Gut biogeography of the bacterial microbiota. *Nat. Rev. Microbiol.* **14** (2015), pp. 20–32.
2. M. G. Rooks, W. S. Garrett, Gut microbiota, metabolites and host immunity. *Nat. Rev. Immunol.* **16** (2016), pp. 341–352.
3. L. A. J. O'Neill, D. Golenbock, A. G. Bowie, The history of Toll-like receptors — redefining innate immunity. *Nat. Rev. Immunol.* **13**, 453–460 (2013).
4. W. Vollmer, D. Blanot, M. A. De Pedro, Peptidoglycan structure and architecture. *FEMS Microbiol. Rev.* **32** (2008), pp. 149–167.
5. W. Weidel, H. Pelzer, Bagshaped macromolecules - A new outlook on bacterial cell walls. *Adv. Enzym. Relat. Areas Mol. Biol.* **26**, 193–232 (1964).
6. C. Otten, M. Brilli, W. Vollmer, P. H. Viollier, J. Salje, Peptidoglycan in obligate intracellular bacteria. *Mol. Microbiol.* **107** (2018), doi:10.1111/mmi.13880.
7. W. Vollmer, Structural variation in the glycan strands of bacterial peptidoglycan. *FEMS Microbiol. Rev.* **32** (2008), pp. 287–306.
8. W. Vollmer, J. V. Höltje, The architecture of the murein (peptidoglycan) in gram-negative bacteria: Vertical scaffold or horizontal layer(s)? *J. Bacteriol.* **186** (2004), pp. 5978–5987.
9. W. Vollmer, S. J. Seligman, Architecture of peptidoglycan: more data and more models. *Trends Microbiol.* **18**, 59–66 (2010).
10. L. Franchi, N. Warner, K. Viani, G. Nuñez, Function of Nod-like receptors in microbial recognition and host defense. *Immunol. Rev.* **227**, 106–128 (2009).
11. S. E. Girardin *et al.*, Nod1 detects a unique muropeptide from gram-negative bacterial peptidoglycan. *Science (80-.).* **300**, 1584–1587 (2003).
12. M. Chamaillard *et al.*, An essential role for NOD1 in host recognition of bacterial peptidoglycan containing diaminopimelic acid. *Nat. Immunol.* **4**, 702–707 (2003).

13. S. E. Girardin *et al.*, Peptidoglycan Molecular Requirements Allowing Detection by Nod1 and Nod2. *J. Biol. Chem.* **278**, 41702–41708 (2003).

14. C. McDonald, N. Inohara, G. Nuñez, Peptidoglycan signaling in innate immunity and inflammatory disease. *J. Biol. Chem.* **280** (2005), pp. 20177–20180.

15. N. Inohara *et al.*, Nod1, an Apaf-1-like activator of caspase-9 and nuclear factor-κB. *J. Biol. Chem.* **274**, 14560–14567 (1999).

16. Y. Ogura *et al.*, Nod2, a Nod1/Apaf-1 Family Member That Is Restricted to Monocytes and Activates NF-??B. *J. Biol. Chem.* **276**, 4812–4818 (2001).

17. M. S. Hayden, S. Ghosh, Signaling to NF-kappaB. *Genes Dev.* **18**, 2195–2224 (2004).

18. S. E. Girardin *et al.*, CARD4/Nod1 mediates NF-kappaB and JNK activation by invasive Shigella flexneri. *EMBO Rep.* **2**, 736–742 (2001).

19. T. Hisamatsu *et al.*, CARD15/NOD2 functions as an antibacterial factor in human intestinal epithelial cells. *Gastroenterology*. **124**, 993–1000 (2003).

20. G. Ferwerda *et al.*, NOD2 and toll-like receptors are nonredundant recognition systems of Mycobacterium tuberculosis. *PLoS Pathog.* **1**, 0279–0285 (2005).

21. K. S. Kobayashi *et al.*, Nod2-dependent regulation of innate and adaptive immunity in the intestinal tract. *Science (80-.).* **307**, 731–734 (2005).

22. K. J. Rangan *et al.*, A secreted bacterial peptidoglycan hydrolase enhances tolerance to enteric pathogens. **353** (2016).

23. V. A. Pedicord *et al.*, *Sci. Immunol.*, in press, doi:10.1126/sciimmunol.aai7732.

24. G. Zhang, T. C. Meredith, D. Kahne, On the essentiality of lipopolysaccharide to Gram-negative bacteria. *Curr. Opin. Microbiol.* **16** (2013), pp. 779–785.

25. X. Wang, P. J. Quinn, Lipopolysaccharide: Biosynthetic pathway and structure modification. *Prog. Lipid Res.* **49** (2010), pp. 97–107.

26. R. G. Gaudet *et al.*, Cytosolic detection of the bacterial metabolite HBP activates TIFA-dependent innate immunity. *Science (80-.).* **348**, 1251–1255 (2015).

27. R. J. Malott *et al.*, Neisseria gonorrhoeae-derived heptose elicits an innate immune response and drives HIV-1 expression. *Proc. Natl. Acad. Sci.* **110**, 10234–10239 (2013).

28. S. R. Galvin, M. S. Cohen, The role of sexually transmitted diseases in HIV transmission. *Nat. Rev. Microbiol.* **2** (2004), pp. 33–42.

29. R. S. Gupta, E. Lorenzini, Phylogeny and molecular signatures (conserved proteins and indels) that are specific for the Bacteroidetes and Chlorobi species. *BMC Evol. Biol.* **7** (2007), doi:10.1186/1471-2148-7-71.

30. J. K. Fredrickson *et al.*, Aromatic-degrading Sphingomonas isolates from the deep subsurface. *Appl. Environ. Microbiol.* **61**, 1917–1922 (1995).

31. I. Olsen, E. Jantzen, Sphingolipids in bacteria and fungi. *Anaerobe*. **7**, 103–112 (2001).

32. B. A. Yard *et al.*, The Structure of Serine Palmitoyltransferase; Gateway to Sphingolipid Biosynthesis. *J. Mol. Biol.* **370**, 870–886 (2007).

33. W. Stoffel, K. Dittmar, R. Wilmes, Sphingolipid Metabolism in Bacteroideaceae. *Hoppe. Seylers. Z. Physiol. Chem.* **356**, 715–726 (1975).

34. L. C. Wieland Brown *et al.*, Production of α-Galactosylceramide by a Prominent Member of the Human Gut Microbiota. *PLoS Biol.* **11** (2013), doi:10.1371/journal.pbio.1001610.

35. D. An *et al.*, Sphingolipids from a symbiotic microbe regulate homeostasis of host intestinal natural killer T cells. *Cell*. **156**, 123–133 (2014).

36. V. A. Blaho, T. Hla, An update on the biology of sphingosine 1-phosphate receptors. *J. Lipid Res.* **55**, 1596–1608 (2014).

37. O. H. Nielsen, Y. Li, B. Johansson-Lindbom, M. Coskun, Sphingosine-1-Phosphate Signaling in Inflammatory Bowel Disease. *Trends Mol. Med.* **23** (2017), pp. 362–374.

38. L. J. Cohen *et al.*, Commensal bacteria make GPCR ligands that mimic human signalling molecules. *Nature*. **549**, 48–53 (2017).

39. R. P. Ryan, Y. Fouhy, J. F. Lucey, J. M. Dow, Cyclic di-GMP signaling in bacteria: Recent advances and new puzzles. *J. Bacteriol.* **188** (2006), pp. 8327–8334.

40. D. Kalia *et al.*, Nucleotide, c-di-GMP, c-di-AMP, cGMP, cAMP, (p)ppGpp signaling in bacteria and implications in pathogenesis. *Chem. Soc. Rev.* **42**, 305–341 (2013).

41. D. L. Burdette *et al.*, STING is a direct innate immune sensor of cyclic di-GMP. *Nature*. **478**, 515–518 (2011).

42. Q. Yin *et al.*, Cyclic di-GMP Sensing via the Innate Immune Signaling Protein STING. *Mol. Cell*. **46**, 735–745 (2012).

43. C. Shu, G. Yi, T. Watts, C. Cheng Kao, P. Li, Structure of STING bound to cyclic di-GMP reveals the mechanism of cyclic dinucleotide recognition by the immune system. *Nat. Struct. Mol. Biol.* **19**, 722–724 (2012).

44. L. Sun, J. Wu, F. Du, X. Chen, Z. J. Chen, Cyclic GMP-AMP synthase is a cytosolic DNA sensor that activates the type I interferon pathway. *Science (80-.).* **339**, 786–791 (2013).

45. J. Wu *et al.*, Cyclic GMP-AMP is an endogenous second messenger in innate immune signaling by cytosolic DNA. *Science (80-.).* **339**, 826–830 (2013).

46. P. Xia *et al.*, The ER membrane adaptor ERAdP senses the bacterial second messenger c-di-AMP and initiates anti-bacterial immunity article. *Nat. Immunol.* **19**, 141–150 (2018).

47. E. D. Sonnenburg, J. L. Sonnenburg, Starving our microbial self: The deleterious consequences of a diet deficient in microbiota-accessible carbohydrates. *Cell Metab.* **20** (2014), pp. 779–786.

48. J. H. Cummings, E. W. Pomare, W. J. Branch, C. P. Naylor, G. T. Macfarlane, Short chain fatty acids in human large intestine, portal, hepatic and venous blood. *Gut*. **28**, 1221–1227 (1987).

49. M. A. R. Vinolo *et al.*, Suppressive effect of short-chain fatty acids on production of proinflammatory mediators by neutrophils. *J. Nutr. Biochem.* **22**, 849–855 (2011).

50. P. M. Smith *et al.*, The Microbial Metabolites, Short-Chain Fatty Acids, Regulate Colonic Treg Cell Homeostasis. *Science (80-.).* **341**, 569–573 (2013).

51. Y. Furusawa *et al.*, Commensal microbe-derived butyrate induces the differentiation of

colonic regulatory T cells. *Nature*. **504**, 446–50 (2013).

52. N. Arpaia *et al.*, Metabolites produced by commensal bacteria promote peripheral regulatory T-cell generation. *Nature*. **504**, 451–5 (2013).

53. N. Singh *et al.*, Activation of Gpr109a, receptor for niacin and the commensal metabolite butyrate, suppresses colonic inflammation and carcinogenesis. *Immunity*. **40**, 128–139 (2014).

54. L. Macia *et al.*, Metabolite-sensing receptors GPR43 and GPR109A facilitate dietary fibre-induced gut homeostasis through regulation of the inflammasome. *Nat. Commun.* **6** (2015), doi:10.1038/ncomms7734.

55. E. Elinav *et al.*, NLRP6 inflammasome regulates colonic microbial ecology and risk for colitis. *Cell*. **145**, 745–757 (2011).

56. C. Voltolini *et al.*, A novel antiinflammatory role for the short-chain fatty acids in human labor. *Endocrinology*. **153**, 395–403 (2012).

57. A. T. Vieira *et al.*, A role for gut microbiota and the metabolite-sensing receptor GPR43 in a murine model of gout. *Arthritis Rheumatol.* **67**, 1646–1656 (2015).

58. A. Trompette *et al.*, Gut microbiota metabolism of dietary fiber influences allergic airway disease and hematopoiesis. *Nat. Med.* **20**, 159–166 (2014).

59. D. Erny *et al.*, Host microbiota constantly control maturation and function of microglia in the CNS. *Nat. Neurosci.* **18**, 965–977 (2015).

60. T. R. Sampson *et al.*, Gut Microbiota Regulate Motor Deficits and Neuroinflammation in a Model of Parkinson's Disease. *Cell*. **167**, 1469–1480.e12 (2016).

61. M. Usami *et al.*, Butyrate and trichostatin A attenuate nuclear factor κB activation and tumor necrosis factor α secretion and increase prostaglandin E2 secretion in human peripheral blood mononuclear cells. *Nutr. Res.* **28**, 321–328 (2008).

62. S. F. W. Kendrick *et al.*, Acetate, the key modulator of inflammatory responses in acute alcoholic hepatitis. *Hepatology*. **51**, 1988–1997 (2010).

63. P. V. Chang, L. Hao, S. Offermanns, R. Medzhitov, The microbial metabolite butyrate regulates intestinal macrophage function via histone deacetylase inhibition. *Proc. Natl. Acad. Sci.* **111**, 2247–2252 (2014).

64. N. Singh *et al.*, Blockade of dendritic cell development by bacterial fermentation products butyrate and propionate through a transporter (Slc5a8)-dependent inhibition of histone deacetylases. *J. Biol. Chem.* **285**, 27601–27608 (2010).

65. C. Chen *et al.*, Deacetylase inhibition promotes the generation and function of regulatory T cells. **13**, 1299–1307 (2007).

66. R. Schilderink, C. Verseijden, W. J. de Jonge, Dietary inhibitors of histone deacetylases in intestinal immunity and homeostasis. *Front. Immunol.* **4** (2013), doi:10.3389/fimmu.2013.00226.

67. F. Rivera-Chávez *et al.*, Depletion of Butyrate-Producing Clostridia from the Gut Microbiota Drives an Aerobic Luminal Expansion of Salmonella. *Cell Host Microbe*. **19**, 443–454 (2016).

68. M. X. Byndloss *et al.*, Microbiota-activated PPAR-γ signaling inhibits dysbiotic Enterobacteriaceae expansion. *Science (80-.).* **357**, 570–575 (2017).

69. S. D. Lawhon, R. Maurer, M. Suyemoto, C. Altier, Intestinal short-chain fatty acids alter Salmonella typhimurium invasion gene expression and virulence through BarA/SirA. *Mol. Microbiol.* **46**, 1451–1464 (2002).

70. I. Gantois *et al.*, Butyrate specifically down-regulates Salmonella pathogenicity island 1 gene expression. *Appl. Environ. Microbiol.* **72**, 946–949 (2006).

71. C.-C. C. Hung *et al.*, The intestinal fatty acid propionate inhibits Salmonella invasion through the post-translational control of HilD. *Mol. Microbiol.* **87**, 1045–1060 (2013).

72. D. N. Bronner *et al.*, Genetic Ablation of Butyrate Utilization Attenuates Gastrointestinal Salmonella Disease. *Cell Host Microbe*. **23**, 266–273.e4 (2018).

73. E. H. Harrison, MECHANISMS OF DIGESTION AND ABSORPTION OF DIETARY

VITAMIN A. *Annu. Rev. Nutr.* **25**, 87–103 (2005).

74. A. Larange, H. Cheroutre, Retinoic Acid and Retinoic Acid Receptors as Pleiotropic Modulators of the Immune System. *Annu. Rev. Immunol.* **34**, 369–394 (2016).

75. J. E. Balmer, R. Blomhoff, Gene expression regulation by retinoic acid. *J. Lipid Res.* **43**, 1773–1808 (2002).

76. R. Kurokawa *et al.*, Regulation of retinoid signalling by receptor polarity and allosteric control of ligand binding. *Nature*. **371**, 528–531 (1994).

77. R. Kurokawa *et al.*, Polarity-specific activities of retinoic acid receptors determined by a co-repressor. *Nature*. **377** (1995), pp. 451–454.

78. S. Westin *et al.*, Interactions controlling the assembly of nuclear-receptor heterodimers and co-activators. *Nature*. **395**, 199–202 (1998).

79. M. Raverdeau, K. H. G. Mills, Modulation of T Cell and Innate Immune Responses by Retinoic Acid. *J. Immunol.* **192**, 2953–2958 (2014).

80. E. J. Villablanca, Retinoic acid-producing dcs and gut-tropic FOXP3+ regulatory T cells in the induction of oral tolerance. *Oncoimmunology*. **2** (2013), , doi:10.4161/onci.22987.

81. O. Pabst, A. M. Mowat, Oral tolerance to food protein. *Mucosal Immunol.* **5** (2012), pp. 232–239.

82. R. Zeng *et al.*, Retinoic acid regulates the development of a gut-homing precursor for intestinal dendritic cells. *Mucosal Immunol.* **6**, 847–856 (2013).

83. R. Zeng, M. Bscheider, K. Lahl, M. Lee, E. C. Butcher, Generation and transcriptional programming of intestinal dendritic cells: Essential role of retinoic acid. *Mucosal Immunol.* **9**, 183–193 (2016).

84. J. A. Hill *et al.*, Retinoic Acid Enhances Foxp3 Induction Indirectly by Relieving Inhibition from CD4+CD44hi Cells. *Immunity*. **29**, 758–770 (2008).

85. S. Xiao *et al.*, Retinoic Acid Increases Foxp3+ Regulatory T Cells and Inhibits Development of Th17 Cells by Enhancing TGF-β-Driven Smad3 Signaling and Inhibiting IL-6 and IL-23

Receptor Expression. *J. Immunol.* **181**, 2277–2284 (2008).

86. G. Tejón *et al.*, Vitamin A Impairs the Reprogramming of Tregs into IL-17-Producing Cells during Intestinal Inflammation. *Biomed Res. Int.* **2015** (2015), doi:10.1155/2015/137893.

87. D. Mucida *et al.*, Reciprocal TH17 and regulatory T cell differentiation mediated by retinoic acid. *Science (80-.).* **317**, 256–260 (2007).

88. B. Cassani *et al.*, Gut-tropic T cells that express integrin α4β7 and CCR9 Are required for induction of oral immune tolerance in mice. *Gastroenterology*. **141**, 2109–2118 (2011).

89. U. Hadis *et al.*, Intestinal Tolerance Requires Gut Homing and Expansion of FoxP3+ Regulatory T Cells in the Lamina Propria. *Immunity*. **34**, 237–246 (2011).

90. M. H. Kim, E. J. Taparowsky, C. H. Kim Correspondence, C. H. Kim, Retinoic Acid Differentially Regulates the Migration of Innate Lymphoid Cell Subsets to the Gut. *Immunity*. **43**, 107–119 (2015).

91. E. J. Villablanca *et al.*, MyD88 and retinoic acid signaling pathways interact to modulate gastrointestinal activities of dendritic cells. *Gastroenterology*. **141**, 176–185 (2011).

92. K. G. McDonald *et al.*, Epithelial expression of the cytosolic retinoid chaperone cellular retinol binding protein II is essential for in vivo imprinting of local gut dendritic cells by lumenal retinoids. *Am. J. Pathol.* **180**, 984–997 (2012).

93. S. P. Spencer *et al.*, Adaptation of innate lymphoid cells to a micronutrient deficiency promotes type 2 barrier immunity. *Science (80-.).* **343**, 432–437 (2014).

94. D. Dodd *et al.*, A gut bacterial pathway metabolizes aromatic amino acids into nine circulating metabolites. *Nature*. **551**, 648–652 (2017).

95. E. A. Kiss *et al.*, Natural aryl hydrocarbon receptor ligands control organogenesis of intestinal lymphoid follicles. *Science (80-.).* **334**, 1561–1565 (2011).

96. Y. Li *et al.*, Exogenous stimuli maintain intraepithelial lymphocytes via aryl hydrocarbon receptor activation. *Cell*. **147**, 629–640 (2011).

97. I. Monteleone *et al.*, Aryl Hydrocarbon Receptor-Induced Signals Up-regulate IL-22

Production and Inhibit Inflammation in the Gastrointestinal Tract. *Gastroenterology*. **141**, 237–248.e1 (2011).

98. N. Tijet *et al.*, Aryl hydrocarbon receptor regulates distinct dioxin-dependent and dioxin-independent gene batteries. *Mol Pharmacol*. **69**, 140–153 (2006).

99. S. A. Kliewer, B. Goodwin, T. M. Willson, The nuclear pregnane X receptor: A key regulator of xenobiotic metabolism. *Endocr. Rev.* **23**, 687–702 (2002).

100. L. Cervantes-Barragan *et al.*, *Science (80-.).*, in press, doi:10.1126/science.aah5825.

101. T. Sujino *et al.*, Tissue adaptation of regulatory and intraepithelial CD4+ T cells controls gut inflammation. *Science (80-.).* **352**, 1581–1586 (2016).

102. T. Zelante *et al.*, Tryptophan catabolites from microbiota engage aryl hydrocarbon receptor and balance mucosal reactivity via interleukin-22. *Immunity*. **39**, 372–385 (2013).

103. M. Wlodarska *et al.*, Indoleacrylic Acid Produced by Commensal Peptostreptococcus Species Suppresses Inflammation. *Cell Host Microbe*. **22**, 25–37.e6 (2017).

104. M. Venkatesh *et al.*, Symbiotic bacterial metabolites regulate gastrointestinal barrier function via the xenobiotic sensor PXR and toll-like receptor 4. *Immunity*. **41**, 296–310 (2014).

105. M. Dusseaux *et al.*, Human MAIT cells are xenobiotic resistant, tissue-targeted, CD161hi IL-17 secreting T cells. *Blood*. **117**, 1250–1260 (2011).

106. R. J. Napier, E. J. Adams, M. C. Gold, D. M. Lewinsohn, The role of mucosal associated invariant T cells in antimicrobial immunity. *Front. Immunol.* **6** (2015), doi:10.3389/fimmu.2015.00344.

107. J. E. Ussher, P. Klenerman, C. B. Willberg, Mucosal-associated invariant T-cells: New players in anti-bacterial immunity. *Front. Immunol.* **5** (2014), , doi:10.3389/fimmu.2014.00450.

108. T. S. C. Hinks, Mucosal-associated invariant T cells in autoimmunity, immune-mediated diseases and airways disease. *Immunology*. **148**, 1–12 (2016).

109. E. Treiner *et al.*, Selection of evolutionarily conserved mucosal-associated invariant T cells by MR1. *Nature*. **422**, 164–169 (2003).

110. K. Hashimoto, M. Hirai, Y. Kurosawa, A gene outside the human MHC related to classical HLA class I genes. *Science*. **269**, 693–695 (1995).

111. L. Kjer-Nielsen *et al.*, MR1 presents microbial vitamin B metabolites to MAIT cells. *Nature*. **491**, 717–723 (2012).

112. A. J. Corbett *et al.*, T-cell activation by transitory neo-antigens derived from distinct microbial pathways. *Nature*. **509**, 361–5 (2014).

113. S. Mondot, P. Boudinot, O. Lantz, MAIT, MR1, microbes and riboflavin: a paradigm for the co-evolution of invariant TCRs and restricting MHCI-like molecules? *Immunogenetics*. **68**, 537–548 (2016).

114. J. Y. L. Chiang, Bile acids: regulation of synthesis. *J. Lipid Res.* **50**, 1955–1966 (2009).

115. C. Thomas *et al.*, TGR5-Mediated Bile Acid Sensing Controls Glucose Homeostasis. *Cell Metab.* **10**, 167–177 (2009).

116. F. Kuipers, V. W. Bloks, A. K. Groen, Beyond intestinal soap - Bile acids in metabolic control. *Nat. Rev. Endocrinol.* **10** (2014), pp. 488–498.

117. M. Makishima *et al.*, Identification of a nuclear receptor for bile acids. *Science*. **284**, 1362–1365 (1999).

118. D. J. Parks *et al.*, Bile acids: Natural ligands for an orphan nuclear receptor. *Science (80-.).* **284**, 1365–1368 (1999).

119. H. Wang, J. Chen, K. Hollister, L. C. Sowers, B. M. Forman, Endogenous bile acids are ligands for the nuclear receptor FXR/BAR. *Mol. Cell*. **3**, 543–553 (1999).

120. Y. Kawamata *et al.*, A G protein-coupled receptor responsive to bile acids. *J. Biol. Chem.* **278**, 9435–9440 (2003).

121. T. Maruyama *et al.*, Identification of membrane-type receptor for bile acids (M-BAR). *Biochem. Biophys. Res. Commun.* **298**, 714–719 (2002).

122. T. Inagaki *et al.*, Regulation of antibacterial defense in the small intestine by the nuclear bile acid receptor. *Proc. Natl. Acad. Sci.* **103**, 3920–3925 (2006).

123. Y.-D. Wang, W.-D. Chen, D. Yu, B. M. Forman, W. Huang, The G-protein coupled bile acid receptor Gpbar1 (TGR5) negatively regulates hepatic inflammatory response through antagonizing Nuclear Factor κB. *Hepatology*. **54**, 1421–1432 (2011).

124. T. W. H. Pols *et al.*, TGR5 activation inhibits atherosclerosis by reducing macrophage inflammation and lipid loading. *Cell Metab.* **14**, 747–757 (2011).

125. V. Keitel *et al.*, The G-protein coupled bile salt receptor TGR5 is expressed in liver sinusoidal endothelial cells. *Hepatology*. **45**, 695–704 (2007).

126. V. Keitel, M. Donner, S. Winandy, R. Kubitz, D. Häussinger, Expression and function of the bile acid receptor TGR5 in Kupffer cells. *Biochem. Biophys. Res. Commun.* **372**, 78–84 (2008).

127. R. Pellicciari, G. Costantino, S. Fiorucci, Farnesoid X receptor: From structure to potential clinical applications. *J. Med. Chem.* **48** (2005), pp. 5383–5403.

128. S. Fiorucci, G. Rizzo, A. Donini, E. Distrutti, L. Santucci, Targeting farnesoid X receptor for liver and metabolic disorders. *Trends Mol. Med.* **13** (2007), pp. 298–309.

129. K. Otte *et al.*, Identification of farnesoid X receptor beta as a novel mammalian nuclear receptor sensing lanosterol. *Mol. Cell. Biol.* **23**, 864–72 (2003).

130. H. Sato *et al.*, Novel potent and selective bile acid derivatives as TGR5 agonists: Biological screening, structure-activity relationships, and molecular modeling studies. *J. Med. Chem.* **51**, 1831–1841 (2008).

131. R. Pellicciari *et al.*, Nongenomic actions of bile acids. Synthesis and preliminary characterization of 23- and 6,23-alkyl-substituted bile acid derivatives as selective modulators for the G-protein coupled receptor TGR5. *J. Med. Chem.* **50**, 4265–4268 (2007).

132. A. Nguyen, B. Bouscarel, Bile acids and signal transduction: Role in glucose homeostasis.

Cell. Signal. **20** (2008), pp. 2180–2197.

133. T. Claudel, G. Zollner, M. Wagner, M. Trauner, Role of nuclear receptors for bile acid metabolism, bile secretion, cholestasis, and gallstone disease. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1812** (2011), pp. 867–878.

134. G. Porez, J. Prawitt, B. Gross, B. Staels, Bile acid receptors as targets for the treatment of dyslipidemia and cardiovascular disease. *J. Lipid Res.* **53**, 1723–1737 (2012).

135. C. Degirolamo, S. Modica, G. Palasciano, A. Moschetta, Bile acids and colon cancer: Solving the puzzle with nuclear receptors. *Trends Mol. Med.* **17** (2011), pp. 564–572.

136. V. Urdaneta, J. Casadesús, Interactions between Bacteria and Bile Salts in the Gastrointestinal and Hepatobiliary Tracts. *Front. Med.* **4** (2017), doi:10.3389/fmed.2017.00163.

137. K. J. Rangan, H. C. Hang, Biochemical Mechanisms of Pathogen Restriction by Intestinal Bacteria. *Trends Biochem. Sci.* **42** (2017), pp. 887–898.

138. K. Gotoh *et al.*, Bile acid-induced virulence gene expression of Vibrio parahaemolyticus reveals a novel therapeutic potential for bile acid sequestrants. *PLoS One*. **5** (2010), doi:10.1371/journal.pone.0013365.

139. P. Li *et al.*, Bile salt receptor complex activates a pathogenic type III secretion system. *Elife*. **5** (2016), doi:10.7554/eLife.15718.

140. R. Thanissery, J. A. Winston, C. M. Theriot, Inhibition of spore germination, growth, and toxin activity of clinically relevant C. difficile strains by gut microbiota derived secondary bile acids. *Anaerobe*. **45**, 86–100 (2017).

141. C. G. Buffie *et al.*, Precision microbiome reconstitution restores bile acid mediated resistance to Clostridium difficile. *Nature*. **517**, 205–208 (2015).

142. C. R. Eade *et al.*, Bile acids function synergistically to repress invasion gene expression in salmonella by destabilizing the invasion regulator hilD. *Infect. Immun.* **84**, 2198–2208 (2016).

143. M. Levy *et al.*, Microbiota-Modulated Metabolites Shape the Intestinal Microenvironment by Regulating NLRP6 Inflammasome Signaling. *Cell*. **163**, 1428–1443 (2015).

144. L. J. Cohen *et al.*, Functional metagenomic discovery of bacterial effectors in the human microbiome and isolation of commendamide, a GPCR G2A/132 agonist. *Proc. Natl. Acad. Sci.* **112**, E4825–E4834 (2015).

145. C. J. Guo *et al.*, Discovery of Reactive Microbiota-Derived Metabolites that Inhibit Host Proteases. *Cell*. **168**, 517–526.e18 (2017).

146. M. Zimmermann, M. A. Fischbach, A family of pyrazinone natural products from a conserved nonribosomal peptide synthetase in staphylococcus aureus. *Chem. Biol.* **17**, 925–930 (2010).

147. A. M. O'Hara, F. Shanahan, The gut flora as a forgotten organ. *EMBO Rep.* **7** (2006), pp. 688–693.

148. P. D. Cani, Gut microbiota-at the intersection of everything? *Nat. Rev. Gastroenterol. Hepatol.* **14** (2017), pp. 321–322.

149. J. Tap *et al.*, Towards the human intestinal microbiota phylogenetic core. *Environ. Microbiol.* **11**, 2574–2584 (2009).

150. M. X. Byndloss, A. J. Bäumler, The germ-organ theory of non-communicable diseases. *Nat. Rev. Microbiol.* **16**, 103–110 (2018).

151. R. Sender, S. Fuchs, R. Milo, Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLoS Biol.* **14** (2016), doi:10.1371/journal.pbio.1002533.

152. P. J. Turnbaugh *et al.*, An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature*. **444**, 1027–31 (2006).

153. B. Stecher, W.-D. Hardt, The role of microbiota in infectious disease. *Trends Microbiol.* **16**, 107–14 (2008).

154. D. A. Hill, D. Artis, Intestinal bacteria and the regulation of immune cell homeostasis. *Annu. Rev. Immunol.* **28**, 623–67 (2010).

155. S. Nell, S. Suerbaum, C. Josenhans, The impact of the microbiota on the pathogenesis of IBD: lessons from mouse infection models. *Nat. Rev. Microbiol.* **8**, 564–77 (2010).

156. A. Shreiner, G. B. Huffnagle, M. C. Noverr, The “Microflora Hypothesis” of allergic disease. *Adv. Exp. Med. Biol.* **635**, 113–34 (2008).

157. L. Wen *et al.*, Innate immunity and intestinal microbiota in the development of Type 1 diabetes. *Nature*. **455**, 1109–13 (2008).

158. M. Austin, M. Mellow, W. M. Tierney, Fecal Microbiota Transplantation in the Treatment of Clostridium difficile Infections. *Am. J. Med.* (2014), doi:10.1016/j.amjmed.2014.02.017.

159. A. Haraga, M. B. Ohlson, S. I. Miller, Salmonellae interplay with host cells. *Nat. Rev. Microbiol.* **6**, 53–66 (2008).

160. J. E. Galán, Salmonella interactions with host cells: type III secretion at work. *Annu. Rev. Cell Dev. Biol.* **17**, 53–86 (2001).

161. T. P. Moest, S. Méresse, Salmonella T3SSs: Successful mission of the secret(ion) agents. *Curr. Opin. Microbiol.* **16** (2013), pp. 38–44.

162. J. R. Ellermeier, J. M. Slauch, Adaptation to the host environment: regulation of the SPI1 type III secretion system in Salmonella enterica serovar Typhimurium. *Curr. Opin. Microbiol.* **10**, 24–9 (2007).

163. J. S. Yount *et al.*, Palmitoylome profiling reveals S-palmitoylation-dependent antiviral activity of IFITM3. *Nat. Chem. Biol.* **6**, 610–614 (2010).

164. K. J. Rangan, Y.-Y. Yang, G. Charron, H. C. Hang, Rapid visualization and large-scale profiling of bacterial lipoproteins with chemical reporters. *J. Am. Chem. Soc.* **132**, 10628–10629 (2010).

165. Y.-Y. Yang, J. M. Ascano, H. C. Hang, Bioorthogonal chemical reporters for monitoring protein acetylation. *J. Am. Chem. Soc.* **132**, 3640–3641 (2010).

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

167. S. Uzzau, N. Figueroa-Bossi, S. Rubino, L. Bossi, Epitope tagging of chromosomal genes in Salmonella. *Proc. Natl. Acad. Sci.* **98**, 15264–15269 (2001).

168. W. Jiang, D. Bikard, D. Cox, F. Zhang, L. a Marraffini, RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* **31**, 233–9 (2013).

169. H. H. Wang *et al.*, Genome-scale promoter engineering by coselection MAGE. *Nat. Methods*. **advance on**, 591–3 (2012).

170. P. Mali *et al.*, RNA-guided human genome engineering via Cas9. *Science (80-.).* **339**, 823–826 (2013).

171. M. Jinek *et al.*, RNA-programmed genome editing in human cells. *Elife*. **2013** (2013), doi:10.7554/eLife.00471.

172. E. Verdin, M. Ott, Acetylphosphate: A novel link between Lysine Acetylation and intermediary metabolism in bacteria. *Mol. Cell*. **51** (2013), pp. 132–134.

173. Q. Wang *et al.*, Acetylation of metabolic enzymes coordinates carbon source utilization and metabolic flux. *Science*. **327**, 1004–7 (2010).

174. T. Nakagawa *et al.*, Sir2-dependent activation of acetyl-CoA synthetase by deacetylation of active lysine. *Science*. **298**, 2390–2 (2002).

175. F. Liu *et al.*, Acetylome analysis reveals diverse functions of lysine acetylation in Mycobacterium tuberculosis. *Mol. Cell. Proteomics*. **13**, 3352–66 (2014).

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

177. S. Tyanova *et al.*, The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods*. **13** (2016), pp. 731–740.

178. V. Bajaj, R. L. Lucas, C. Hwang, C. A. Lee, Co-ordinate regulation of Salmonella typhimurium invasion genes by environmental and regulatory factors is mediated by control of hilA expression. *Mol. Microbiol.* **22**, 703–714 (1996).

179. V. Bajaj, C. Hwang, C. A. Lee, hilA is a novel ompR/toxR family member that activates the expression of Salmonella typhimurium invasion genes. *Mol. Microbiol.* **18**, 715–727 (1995).

180. C. P. Lostroh, V. Bajaj, C. A. Lee, The cis requirements for transcriptional activation by HilA, a virulence determinant encoded on SPI-1. *Mol. Microbiol.* **37**, 300–315 (2000).

181. C. P. Lostroh, C. A. Lee, The HilA box and sequences outside it determine the magnitude of HilA-dependent activation of P(prgH) from Salmonella pathogenicity island 1. *J. Bacteriol.* **183**, 4876–85 (2001).

182. S. C. J. De Keersmaecker *et al.*, Microarray analysis and motif detection reveal new targets of the Salmonella enterica serovar Typhimurium HilA regulatory protein, including hilA itself. *J. Bacteriol.* (2005), doi:10.1128/JB.187.13.4381-4391.2005.

183. K. Eichelberg, J. E. Galan, Differential Regulation of Salmonella typhimurium Type III Secreted Proteins by Pathogenicity Island 1 (SPI-1)-Encoded Transcriptional Activators InvF and HilA. *Infect. Immun.* **67**, 4099–4105 (1999).

184. J. D. Boddicker, B. M. Knosp, B. D. Jones, Transcription of the Salmonella invasion gene activator, hilA, requires HilD activation in the absence of negative regulators. *J. Bacteriol.* **185**, 525–533 (2003).

185. A. Sturm, Bistable regulation of ttss-1 genes in Salmonella Typhimurium (2011), doi:10.3929/ethz-a-006715280.

186. D. E. Kim, D. Chivian, D. Baker, Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res.* **32** (2004), doi:10.1093/nar/gkh468.

187. H. Neumann, S. Y. Peak-Chew, J. W. Chin, Genetically encoding N(epsilon)-acetyllysine in recombinant proteins. *Nat. Chem. Biol.* **4**, 232–4 (2008).

188. M. J. Gattner, M. Vrabel, T. Carell, Synthesis of ε-N-propionyl-, ε-N-butyryl-, and ε-N-crotonyl-lysine containing histone H3 using the pyrrolysine system. *Chem. Commun. (Camb).* **49**, 379–81 (2013).

189. J. Li *et al.*, Ligand-free palladium-mediated site-specific protein labeling inside gram-

negative bacterial pathogens. *J. Am. Chem. Soc.* **135**, 7330–7338 (2013).

190. S. H. Sternberg, S. Redding, M. Jinek, E. C. Greene, J. A. Doudna, DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature*. **507**, 62–67 (2014).

191. D. L. Court, J. a Sawitzke, L. C. Thomason, Genetic engineering using homologous recombination. *Annu. Rev. Genet.* **36**, 361–88 (2002).

192. I. M. V Thijs *et al.*, Delineation of the Salmonella enterica serovar Typhimurium HilA regulon through genome-wide location and transcript analysis. *J. Bacteriol.* **189**, 4587–4596 (2007).

193. J. Garrity, J. G. Gardner, W. Hawse, C. Wolberger, J. C. Escalante-Semerena, N-Lysine Propionylation Controls the Activity of Propionyl-CoA Synthetase * □ S. *Publ. JBC Pap. Press* (2007), doi:10.1074/jbc.M704409200.

194. B. Weinert *et al.*, Acetyl-Phosphate is a critical determinant of Lysine Acetylation in E.coli. *Mol. Cell*. **51**, 265–272 (2013).

195. M. L. Kuhn *et al.*, Structural, kinetic and proteomic characterization of acetyl phosphate-dependent bacterial protein acetylation. *PLoS One*. **9**, e94816 (2014).

196. S. Tu *et al.*, YcgC represents a new protein deacetylase family in prokaryotes. *Elife*. **4**, 1–18 (2015).

197. R. A. Daly, C. P. Lostroh, Genetic analysis of the Salmonella transcription factor HilA. *Can. J. Microbiol.* **54**, 854–60 (2008).

198. A. R. Horswill, J. C. Escalante-Semerena, Salmonella typhimurium LT2 catabolizes propionate via the 2-methylcitric acid cycle. *J. Bacteriol.* **181**, 5615–5623 (1999).

199. B. Stecher *et al.*, Salmonella enterica Serovar Typhimurium Exploits Inflammation to Compete with the Intestinal Microbiota. *PLoS Biol.* **5**, e244 (2007).

200. J. Cox *et al.*, Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. Proteomics*. **13**, 2513–26 (2014).

201. F. Van Immerseel *et al.*, The use of organic acids to combat Salmonella in poultry: a mechanistic explanation of the efficacy. *Avian Pathol.* **35**, 182–8 (2006).

202. F. Van Immerseel *et al.*, Medium-chain fatty acids decrease colonization and invasion through hilA suppression shortly after infection of chickens with Salmonella enterica serovar enteritidis. *Appl. Environ. Microbiol.* **70**, 3582–3587 (2004).

203. G. G. Viarengo *et al.*, Unsaturated long-chain free fatty acids are input signals of the Salmonella enterica PhoP/PhoQ regulatory system. *J. Biol. Chem.* **288**, 0–26 (2013).

204. Y. A. Golubeva, J. R. Ellermeier, J. E. Cott Chubiz, J. M. Slauch, Intestinal Long-Chain Fatty Acids Act as a Direct Signal To Modulate Expression of the Salmonella Pathogenicity Island 1 Type III Secretion System. *MBio*. **7**, 1–9 (2015).

205. X. Gao, E. Pujos-Guillot, J. L. Sébédio, in *Analytical Chemistry* (2010), vol. 82, pp. 6447–6456.

206. S. H. Iram, J. E. Cronan, The β-oxidation systems of Escherichia coli and Salmonella enterica are not functionally equivalent. *J. Bacteriol.* **188**, 599–608 (2006).

207. J. S. Yount, G. Charron, H. C. Hang, Bioorthogonal proteomics of 15-hexadecynyloxyacetic acid chemical reporter reveals preferential targeting of fatty acid modified proteins and biosynthetic enzymes. *Bioorg. Med. Chem.* **20**, 650–4 (2012).

208. A. Percher *et al.*, Mass-tag labeling reveals site-specific and endogenous levels of protein S-fatty acylation. *Proc. Natl. Acad. Sci.* **113**, 4302–4307 (2016).

209. D. Senyilmaz *et al.*, Regulation of mitochondrial morphology and function by stearoylation of TFR1. *Nature* (2015), doi:10.1038/nature14601.

210. H. Jiang *et al.*, SIRT6 regulates TNF-α secretion through hydrolysis of long-chain fatty acyl lysine. *Nature*. **496** (2013), pp. 110–3.

211. M. Zhang *et al.*, A genetically incorporated crosslinker reveals chaperone cooperation in acid resistance. *Nat. Chem. Biol.* **7**, 671–7 (2011).

212. S. Lin *et al.*, Site-specific incorporation of photo-cross-linker and bioorthogonal amino acids

into enteric bacterial pathogens. *J. Am. Chem. Soc.* (2011), doi:10.1021/ja209008w.

213. T. Peng, H. C. Hang, Bifunctional fatty acid chemical reporter for analyzing S-palmitoylated membrane protein-protein interactions in mammalian cells. *J. Am. Chem. Soc.* **137**, 556–9 (2015).

214. M. M. Zhang, P.-Y. J. Wu, F. D. Kelly, P. Nurse, H. C. Hang, Quantitative control of protein s-palmitoylation regulates meiotic entry in fission yeast. *PLoS Biol.* **11** (2013), p. e1001597.

215. P. Stanley, L. C. Packman, V. Koronakis, C. Hughes, Fatty acylation of two internal lysine residues required for the toxic activity of Escherichia coli hemolysin. *Science (80-.).* **266**, 1992–1996 (1994).

216. G. Charron *et al.*, Robust fluorescent detection of protein fatty-acylation with chemical reporters. *J. Am. Chem. Soc.* **131**, 4967–4975 (2009).

217. M. Zhang *et al.*, A genetically incorporated crosslinker reveals chaperone cooperation in acid resistance. *Nat. Chem. Biol.* **7**, 671–677 (2011).

218. A. Shevchenko, H. Tomas, J. Havliš, J. V. Olsen, M. Mann, In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protoc.* **1**, 2856–2860 (2007).

219. D. Ruiz *et al.*, Phosphorylation-independent activation of the atypical response regulator NblR. *Microbiology*. **154**, 3002–3015 (2008).

220. E. Hong *et al.*, Structure of an atypical orphan response regulator protein supports a new phosphorylation-independent regulatory mechanism. *J. Biol. Chem.* **282**, 20667–20675 (2007).

221. I. Delany, G. Spohn, R. Rappuoli, V. Scarlato, Growth phase-dependent regulation of target gene promoters for binding of the essential orphan response regulator HP1043 of Helicobacter pylori. *J. Bacteriol.* **184**, 4800–4810 (2002).

222. M. Nguyen, J. Yoon, M. Cho, S. Lee, Prokaryotic 2-component systems and the OmpR/PhoB superfamily. *Can J Microbiol*. **61**, 799–810 (2015).

223. S. L. Harlocker, L. Bergstrom, M. Inouye, Tandem binding of six OmpR proteins to the ompF upstream regulatory sequence of Escherichia coli. *J. Biol. Chem.* **270**, 26849–26856 (1995).

224. R. Gao, A. M. Stock, Molecular strategies for phosphorylation-mediated regulation of response regulator activity. *Curr. Opin. Microbiol.* **13** (2010), pp. 160–167.

225. A. Abouelfetouh *et al.*, The E. coli sirtuin CobB shows no preference for enzymatic and nonenzymatic lysine acetylation substrate sites, doi:10.1002/mbo3.223.

226. S. Thao, C.-S. Chen, H. Zhu, J. C. Escalante-Semerena, Nε-lysine acetylation of a bacterial transcription factor inhibits Its DNA-binding activity. *PLoS One*. **5**, e15123 (2010).

227. L. I. Hu *et al.*, Acetylation of the response regulator RcsB controls transcription from a small RNA promoter. *J. Bacteriol.* **195**, 4174–4186 (2013).

228. J. Ren *et al.*, Acetylation of Lysine 201 Inhibits the DNA-Binding Ability of PhoP to Regulate Salmonella Virulence, 1–29 (2016).

229. M. J. Lowden *et al.*, Structure of Vibrio cholerae ToxT reveals a mechanism for fatty acid regulation of virulence genes. *Proc. Natl. Acad. Sci.* **107**, 2860–2865 (2010).

230. S. C. Plecha, J. H. Withey, The mechanism for inhibition of *Vibrio cholerae* ToxT activity by the unsaturated fatty acid components of bile. *J. Bacteriol.* (2015), doi:10.1128/JB.02409-14.

231. V. Allfrey, R. Faulkner, A. Mirsky, Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. *Proc. Natl. Acad. Sci. U. S. A.* **315**, 786–794 (1964).

232. J. E. Brownell *et al.*, Tetrahymena histone acetyltransferase A: A homolog to yeast Gcn5p linking histone acetylation to gene activation. *Cell*. **84**, 843–851 (1996).

233. P. J. Barnes, I. M. Adcock, K. Ito, Histone acetylation and deacetylation: Importance in inflammatory lung diseases. *Eur. Respir. J.* **25** (2005), pp. 552–563.

234. M. A. Glozak, E. Seto, Histone deacetylases and cancer. *Oncogene*. **26**, 5420–5432 (2007).

235. C. L. Zhang *et al.*, Class II histone deacetylases act as signal-responsive repressors of cardiac hypertrophy. *Cell*. **110**, 479–488 (2002).

236. B. Tang, B. Dean, E. A. Thomas, Disease- and age-related changes in histone acetylation at gene promoters in psychiatric disorders. *Transl. Psychiatry*. **1** (2011), doi:10.1038/tp.2011.61.

237. J. Lee, Y. J. Hwang, K. Y. Kim, N. W. Kowall, H. Ryu, Epigenetic Mechanisms of Neurodegeneration in Huntington's Disease. *Neurotherapeutics*. **10** (2013), pp. 664–676.

238. L. Cai, B. M. Sutter, B. Li, B. P. Tu, Acetyl-CoA Induces Cell Growth and Proliferation by Promoting the Acetylation of Histones at Growth Genes. *Mol. Cell*. **42**, 426–437 (2011).

239. W. Gu, R. G. Roeder, Activation of p53 sequence-specific DNA binding by acetylation of the p53 C-terminal domain. *Cell*. **90**, 595–606 (1997).

240. Y. Chen *et al.*, Lysine Propionylation and Butyrylation Are Novel Post-translational Modifications in Histones. *Mol. Cell. Proteomics*. **6**, 812–819 (2007).

241. L. Dai *et al.*, Lysine 2-hydroxyisobutyrylation is a widely distributed active histone mark. *Nat. Chem. Biol.* (2014), doi:10.1038/nchembio.1497.

242. Z. Xie *et al.*, Lysine Succinylation and Lysine Malonylation in Histones. *Mol. Cell. Proteomics*. **11**, 100–107 (2012).

243. M. Tan *et al.*, Lysine glutarylation is a protein posttranslational modification regulated by SIRT5. *Cell Metab.* **19**, 605–617 (2014).

244. M. Tan *et al.*, Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell*. **146**, 1016–1028 (2011).

245. Z. Xie *et al.*, Metabolic Regulation of Gene Expression by Histone Lysine β-Hydroxybutyrylation. *Mol. Cell*. **62**, 194–206 (2016).

246. A. F. Kebede *et al.*, Histone propionylation is a mark of active chromatin. *Nat. Struct. Mol. Biol.* **24**, 1048–1056 (2017).

247. B. R. Sabari, D. Zhang, C. D. Allis, Y. Zhao, Metabolic regulation of gene expression

through histone acylations. *Nat. Rev. Mol. Cell Biol.* **18**, 90–101 (2016).

248. L. Xie *et al.*, Proteome-wide Lysine Glutarylation Profiling of the Mycobacterium tuberculosis H37Rv. *J. Proteome Res.* **15**, 1379–1385 (2016).

249. L. Qian *et al.*, Global profiling of protein lysine malonylation in Escherichia coli reveals its role in energy metabolism. *J. Proteome Res.* **15**, 2060–2071 (2016).