Title
Probing post-translational modifications in environmentally impactful bacteria using mass spectrometry

Author
Muroski, John

Publication Date
2021

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Probing post-translational modifications in environmentally impactful bacteria using mass spectrometry

A dissertation submitted in partial satisfaction of the requirements for
the degree Doctor of Philosophy
in Biochemistry, Molecular, and Structural Biology

by

John Muroski

2021

ABSTRACT OF THE DISSERTATION

Probing post-translational modifications in environmentally impactful bacteria using mass spectrometry

by

John Muroski

Doctor of Philosophy in Biochemistry, Molecular, and Structural Biology

University of California, Los Angeles, 2021

Professor Joseph Ambrose Loo, Chair

Microbes are found in all corners of the globe, and demonstrate extreme diversity in the roles they play within their niches. Syntrophic bacteria are a group of anaerobic microbes that break down fatty and aromatic acids, a process that forces them to survive at the edges of thermodynamic feasibility. Both the energetically limited state they exist, in combination with the large abundance of reactive acyl-CoA species present as metabolic intermediates, suggest a possibility of widespread lysine acylation, a common class of post-translational modification. We hypothesized that reactive acyl-CoA species would modify lysine residue with the corresponding acyl-lysine modification. Mass spectrometry was used to identify a wide array of lysine acylations in two species of bacteria, *Syntrophus aciditrophicus* and *Syntrophomonas wolfei.* As overly broad analysis could result in misidentifications, we identified diagnostic ion markers and developed methods utilizing these markers for detecting acyl-lysine modifications that could confidently identify even previously unknown modifications. Shotgun proteomics

identified seven acylations in *S. aciditrophicus*, including those previously unidentified in the bacterial domain. Modifications identified were correlated with acyl-CoA intermediates in the cells. Acylations were enriched in pathways associated with aromatic acid degradation pathways. Mass spectrometry data acquired for the *S. wolfei* system revealed the same trends and suggested that lysine acylations changed quantitatively under different growth conditions. Together, the data support the "carbon stress" model of lysine acylation, whereby a buildup of reactive acyl-CoA species will non-enzymatically modify lysine residue in response to cellular conditions. The data also lays out a roadmap for future studies to determine the role these acylations play in regulating syntrophic cell metabolism.

The dissertation of John Muroski is approved.

James U. Bowie

Steven G. Clarke

Robert P. Gunsalus

Joseph Ambrose Loo, Committee Chair

University of California, Los Angeles
2021

Table of Contents

Acknowledgements

I would like to acknowledge my advisor, Professor Joseph Loo, for his patience and help with this dissertation and assistance during graduate school. It is thanks to him that I had many great opportunities to present my work and interact with many scientists to grow my scientific basis. A very special thanks as well to the microbial proteomics members of the lab, Dr. Rachel Ogorzalek Loo and Dr. Hong Nguyen for helping provide both a foundation for my understanding of mass spectrometry and proteomics, as well as helping design projects and generate data. In particular I would also like to thank Janine Fu for her tireless efforts in making much of the work presented here possible. Her support both inside and outside of lab has been invaluable. I thank collaborators Professor Robert Clubb, Professor Robert Gunsalus, Professor Michael McInerney, and Professor Karla Satchell for helping provide opportunities to study a wide range of microbial systems. I would also like to acknowledge other members of the Loo lab both former and current for thoughtful discussions over the course of my PhD. I would like to thank the Philip Whitcome and Audree Fowler Fellowships for financial support. I'd also like to thank my committee, Professor Steven Clarke, Professor James Bowie, Professor Robert Gunsalus and my Committee Chair Professor Joseph Loo.

Chapter Two of the dissertation is a published as a version of the published manuscript: *Muroski, J. M., *Fu, J. Y., Nguyen, H. H., Ogorzalek Loo, R. R., & Loo, J. A. (2021). Leveraging Immonium Ions for Targeting Acyl-Lysine Modifications in Proteomic Datasets. Asterisks denote the authors contributed equally.

Chapter Three of the dissertation is a version of the manuscript in preparation: Muroski, J. M., Fu, J. Y., Nguyen, H. H., Wofford N.Q., Mouttaki, H., James, K.L., McInerney, M.J., Gunsalus,

R.P., Loo, J. A., & Ogorzalek Loo, R. R. The acyl-proteome of *Syntrophus aciditrophicus* reveals metabolic relationship with benzoate degradation.

Chapter Four is pretty cool too.

Vita

EDUCATION

Master of Science in Biochemistry, Molecular, and Structural Biology, 2017
University of California, Los Angeles, Los Angeles, California
Graduate Research Advisor: Joseph A. Loo

Bachelor of Science in Biochemistry, 2015
University of Rochester, Rochester, New York
Cum Laude
Undergraduate Research Advisor: Gloria M. Culver

HONORS AND AWARDS

UCLA Chemistry and Biochemistry Excellence in Research Fellowship (2020)
UCLA Audree V. Fowler Fellowship (2019-2020)
UCLA Whitcome Fellowship (2018-2019)
UCLA Dorothy Radcliffe Dee Fellowship (2016)
University of Rochester Dean's List, 6x (2012 – 2015)
University of Rochester Transfer Dean's Scholarship (2012 – 2014)

RESEARCH INTERESTS

The primary goals of my thesis work has been to understand post-translational modifications (PTMs) and the effects of these modifications on bacterial proteomes. Of particular interest have been the impact of acyl-modifications; as the growing literature on the topic is expanding rapidly. The regulation of these ubiquitous PTMs are also of interest with respect to the *sirtuin* family of enzymes. Particularly, these modifications are being studied in environmentally impactful bacteria, where they have a significant presence and variety in metabolic pathways.

PEER-REVIEWED PUBLICATIONS

5. **Muroski, J.M.**, Fu, J.Y., Nguyen, H.H., Loo, R.R.O., and Loo, J., 2020. "Leveraging immonium ions for targeting acyl-lysine modifications in proteomic datasets". *Proteomics, 21*(3-4), 2000111.

4. Sue, C.S., McConnell, S.A., Ellis-Guardiola, K., **Muroski, J.M.**, McAllister, R., Yu, J., Alverez, A., Chang, C., Loo, R.R.O., Loo, J., Ton-That, H. and Clubb, R.T., 2020. "Kinetics and Optimization of the Lysine-Isopeptide Bond Forming Sortase Enzyme from *Corynebacterium diphtheriae*". *Bioconjugate Chemistry, 31* (6), 1624-1634

3. Herrera, A., **Muroski, J.**, Sengupta, R., Nguyen, H.H., Agarwal, S., Loo, R.R.O., Mattoo, S., Loo, J.A., and Satchell, K.J.F., 2020. "N-terminal autoprocessing and acetylation of MARTX Makes-Caterpillars Floppy-like effector is stimulated by ADP-Ribosylation Factor 1 in advance of Golgi fragmentation". *Cellular Microbiology*, *22* (2), e13133.

2. McConnell, S.A., Amer, B.R., **Muroski, J.**, Fu, J., Chang, C., Loo, R.R.O., Loo, J., Osipiuk, J., Ton-That, H. and Clubb, R.T., 2018. "Protein Labeling via a Specific Lysine-Isopeptide Bond using the Pilin Polymerizing Sortase from *Corynebacterium diphtheriae*". *Journal of the American Chemical Society, 140* (27), 8420-8423.

1. Chang, C., Amer, B. R., Osipiuk, J., McConnell, S. A., Huang, I. H., Hsieh, V., Fu, J., Nguyen, H. H., **Muroski, J.**, Flores, E., Ogorzalek Loo, R. R., Loo, J. A., Putkey, J. A., Joachimiak, A., Das, A., Clubb, R. T., & Ton-That, H. (2018). "In vitro reconstitution of sortase-catalyzed pilus polymerization reveals structural elements involved in pilin cross-linking." *Proceedings of the National Academy of Sciences of the United States of America, 115* (24), E5477–E5486.

PRESENTATIONS

Oral Presentations
Muroski, J., Fu, J., Nguyen, HHN., McInerny, MJ., Loo, RRO., Loo, JA.
"Identifying the Protein Acylome in Syntophus Aciditrophicus and its Relationship to Metabolic Intermediates" December 2020, SoCal Mass Spec 2020

Muroski, J., Nguyen, HHN., Fu, J., McInerny, MJ., Loo, RRO., Loo, JA.
"Characterization of acyl modifications in syntrophic bacteria proteomes" September 2019, UCLA Molecular Biology Institute Retreat

Muroski, J., Fu, J., Loo, RRO., Loo, JA.
"Characterization of a Novel Sirtuin Found in Syntrophic Bacteria" January 2019

Muroski, J., Nguyen, HHN… Loo, JA.
"Utilizing Immonium Ions to Characterize Lysine Acylation in Syntrophic Bacteria" April 2018

Poster Presentations
Muroski, J., Fu, J., Natale, E., Takayesu, A., Mahoney, B., Martinez, OE., Morselli, M., Pellegrini, M., Loo, RRO., Chou, K., Clubb, R., Loo, JA., Gunsalus, R., Yeates, T.
"Metabolism in Microbial Communities and the Associated Biochemistry of Polymer Deconstruction under Study at the UCLA-DOE Institute for Genomics and Proteomics" February 2021, Department of Energy 2021 Genomic Sciences Program Annual Principal Investigator (PI) Meeting

Muroski, J., Nguyen, HHN., McInerny, MJ., Loo, RRO., Loo, JA.
"The metabolic state of a syntrophic bacterium affects the proteomic acylation profile" April 2020, Human Proteomics Organization

Muroski, J., Fu, J., Loo, RRO., Loo, JA.
"Characterizing a novel bacterial sirtuin protein in an intact protein system" April 2019, American Society of Biochemistry and Molecular Biology

Muroski, J., Sedighian, F., Gunsalus, RP., Loo, JA., Loo, RRO.
"Characterization of the Methanosaeta concilii Sheath, a Remarkably Resilient Protein Assembly" February 2019, Department of Energy 2019 Genomic Sciences Program Annual Principal Investigator (PI) Meeting

Muroski, J., Nguyen, HHN., McInerny, MJ., Loo, RRO., Loo, JA.
"Electrospray ionization mass spectrometry provides kinetic insight into a novel bacterial sirtuin protein in an intact protein system" June 2018, American Society of Mass Spectrometry Conference

Muroski, J., Nguyen, HHN., Loo, RRO., Loo, JA.
"Succinyl-CoA Potentially Regulates the Oligomeric Forms of Cyanobacterial Citrate Synthase through Succinylation" June 2017, American Society of Mass Spectrometry Conference

TEACHING EXPERIENCE

Teaching Assistant, UCLA
Chem184: Chemical Instrumentation, Winter 2018
Chem153L: Biochemical Methods I, Winter 2017
Chem153B: DNA, RNA, and Protein Synthesis, Fall 2016
Chem153A: Biochemistry, Intro to Structure, Enzymes, and Metabolism, Winter/Spring 2016

OUTREACH ACTIVITIES

Graduate Student Mentor
University Research Center (URC)-Sciences, University of California, Los Angeles, 2017-Present
- URC helps undergraduate students develop skills necessary for engaging in and conducting research in an academic environment. Primary duties for this position involved giving workshops, feedback sessions, and presentations to students interested in scientific research. We also worked closely with UCLA to focus our attention on underrepresented groups in STEM.

Chair of Recruitment
Graduate Biochemistry Student Association, University of California, Los Angeles, 2018-2019
- Principle duties included designing activities and coordinating with department heads to recruit graduate students to UCLA's BMSB program. Hosting social events and tours of the campus and local community were critical to the role.

SPUR Mentor
Summer Programs for Undergraduate Research (SPUR) UCLA, 2017
HBCU Mentor
Summer Programs for Undergraduate Research (SPUR) UCLA, 2017

Chapter 1

Overview: Topics of Key Importance

1.1 Mass spectrometry-based proteomics

The field of biology is incoherent without an understanding of proteins. The final product of molecular biologist's "central dogma", these macromolecules are often the genotype's effectors that produce a given phenotype. While critically important, analyzing proteins at the systems level has proven to be much more difficult than profiling other nucleic acid based biological macromolecules like DNA and RNA. Whereas DNA and RNA are each only composed of four different types of monomeric units, proteins in most species are composed of 20, greatly increasing the complexity of the analysis. Each type of nucleic acid base differs only by the presence of few hydrogen donors or acceptors that allow for distinction. By comparison, the 20 amino acids differ rather significantly by chemical property; some are acidic, some basic, some are polar, others hydrophobic, and many differ in reactivity under different conditions. This variety once again adds to the challenges of developing a universal technique that can determine the identity and sequence of proteins at the systems level. Compounding the challenge of complexity is that unlike nucleic acids, the production of protein is not "semi-conserved". That is to say, one protein molecule cannot be used as a template to produce others. As such, protein molecules must be *directly* measured in some way to understand sequence, identity, and quantitative information creating a complete divergence from the tools and techniques commonly used in other areas of the study of biological macromolecules.

The primary tool used to study proteomics and often proteins more generally is mass spectrometry (MS). MS is a tool that chemists can use to measure the mass-to-charge ratio (*m/z*) of analytes, and from this are able to determine the mass of the molecule. This feature in and of itself has proven a powerful tool, as since the discovery of electrospray ionization (ESI) and matrix assisted laser desorption ionization (MALDI) it has been possible to detect entire proteins

using these ionization methods and take measurements that can identify chemical properties of the molecules (1, 2). However, given the complexity of biological systems, like whole-cell lysates, this feature is not be enough. Tandem mass spectrometry (MS/MS) is required to truly understand peptide composition. Critically, emerging at the time, was an understanding of collision-induced dissociation (CID) of peptides (3), which allowed for the possibility to study complex proteomic milieus through direct measurements of the peptides. Briefly, upon induction of CID using inert gas, the peptide bonds break, typically at very specific junctures along the protein backbone (3). When the peptides are fragmented a single time, this results in a pattern of ions that can be identified as differing by a single amino acid residue within each "mass shift" in the spectra. While other mechanisms of fragmentation have been developed in the decades since (4–6), the general concept of using mass shifts to identify the sequence and identity of proteins and peptides has remained. When coupled with a separation technique, most commonly liquid chromatography, entire proteomes can be analyzed, demonstrating great power with respect to understanding the mechanisms of underlying processes and disease states (7).

While development on the instrumentation side was critical to allow for the analysis of protein and peptides, the next hurdle that needed to be overcome to run a complete proteomics experiment was on the analysis side, considering a single LC-MS/MS can easily generate on the order of 10,000 tandem MS spectra. Even a brilliant mass spectrometrist who can sequence a fragment spectrum at a pace of 2 minutes per spectra would require near 14 sleepless nights straight to analyze a single one hour run. On this front, the first software to help this analysis was a tool known as SEQUEST (8). To tackle this issue, the algorithm developed would use a known genomic database and utilize theoretical sequences of peptides generated by a protease (8). Upon generation of the theoretical peptides, computed spectra would then be generated from these

peptides using known patterns in fragmentation previously identified and the overlap of theoretical and experimentally measured fragments were compared to one another (8). Since this progenitor, many alternative algorithms have been developed using the same approach of database searching to address limitations of data size (9–11). In more recent years, the challenges of peptide spectral interpretation have moved away from identifying standard peptides resulting directly from standard mRNA translation, and more towards identifying non-canonical peptides, or those peptides that have a modification or deviation from the genomic template. Open search strategies have led the way in this arena (12, 13). Algorithms like those utilized in MSFragger take a very different approach to identifying sequence and modification from spectra. Rather than using simple peak matching strategies with sequence databases, MSFragger decomposes fragment spectra to their individual fragments and groups spectra by similarities across spectra (14). From this categorization, similar peptides with an unspecified mass shift can be identified, thus allowing *a posteriori* assignment of modifications based on mass shift. From this type of analysis, we can begin to uncover what the Nesvizhskii group and others have coined as the "dark matter" of the proteome, that is proteinaceous components that are present in systems but remain unidentified due to a lack of knowledge of them or limitations in searching capabilities (14, 15). Only through careful deliberation, consideration of incorrect assignments, and chemical and biological understanding, can this "dark matter" be identified and elucidated, as overly simplistic assignment strategies may provide easy yet incorrect answers. To do less would be anathema to the scientific method and the field of proteomics.

While the field continues to expand rapidly in a variety of different areas, and methods further expand to help address specific questions, these tools have combined to form the basis for mass spectrometry-based proteomic analysis.

1.2 Post-translational modifications: Acylation

Proteins play critical functional roles in the cell. Beyond the 20 canonical amino acids encoded by the genome, and two non-canonical residues present in some microbial systems, proteins can be further modified to fulfill their functional role or in response to changing environmental conditions. These post-translational modifications (PTMs) occur directly to proteins and are categorized into different groups based on their physicochemical properties, which may differ drastically. Phosphorylation and methylation for instance involve placement of a single small moiety on an amino acid side chain, which may alter protein function, and are thus typically enzymatically controlled. Similar to these are PTMs generated by spontaneous reactions, such as oxidation, amidation, or deamidation which may also slightly alter the composition of an amino acid side chain but do so often as a function of time in physiological conditions. Glycosylation is much more complex, as it involves the addition of sugars, very often long and complex chains of sugars, that are heterogenous in nature (16). Still other PTMs, like isoapartate formation, don't involve a change in the chemical composition of a protein, rather simply the isomerization of a single residue (17). PTMs can also involve the binding of multiple protein together via a cross-link, and may be multiple homologs bound together, or involve an entirely unique class of proteins that bind, à la ubiquitylation.

Different classes of PTMs often have generalizable functions for the role they play in a cell. Phosphorylation often plays a role in cell signaling, a process that is conserved from both bacterial to mammalian systems and can be involved in signals as simple as bacterial motility (18) to inter-tissue signaling at the level of an organism (19). Lipidation is often seen as a necessary PTM for the proper localization of protein that must bind to a membranous region of

the cell or membrane compartment. A variety of protein cross-links may occur through a variety of chemical means that are necessary for the structural stability of protein superstructures; from collagen or elastin in humans, to the pilin formation in microbes (20, 21). Ubiquitylation tends to signal to other protein that a protein is marked for degradation (22). These modifications all play integral roles within the cell, and therefore maintain various mechanisms of regulation.

Physicochemical differences also mean that different techniques are used to characterize classes of modifications and must be optimized. While mass spectrometry remains the gold standard of PTM analysis, the ways in which samples are prepared for MS analysis can be very different. The negatively charged phosphorylation moieties can bind to metal stationary phases, such as titanium oxide, to identify and enrich these PTMs (23). Glycans, despite their extreme heterogeneity, can be bound to lectins for enrichment thereby allowing for specific analysis (16). Ubiquitylation, while being defined as the addition of an entire protein, can be identified by a distinct di-glycine addition to the ε-amino group of lysine that is left after the digest of the protein complex by trypsin protease (24). Many of the modifications and classes of modifications that exist have specific robust methods of identification and characterization.

Acylation is a common class of PTM that has been neglected thus far, as it tends to buck many of the trends and generalizations that can be made about other classes of PTMs. Acyl modifications may exist in a variety of different subclasses. Chain length and composition may vary greatly, allowing for much longer and more complex modifications (beyond acetylation) of crotonylation, succinylation, and 3-hydroxy-3-methyl-glutarylation among many others (25, 26). Recently, the addition of an aromatic moiety has been identified in the form of a benzoyl modification on histone side chains (27). Aside from the propensity of acylating primary amines, like N-termini and lysine side chains, and to a lesser extent cysteine thioesterification, there is

little obvious overlap in these modifications structure that can be exploited. Individual acyl groups may have antibodies generated to enrich for them (28), but antibodies may be prone to sequence biases or cross-reactivity. In terms of cellular function, there is not a universal role that can be determined. Modification of histone tails by acyl groups has been demonstrated to modulate gene expression (29). However, binding to enzymatic proteins has also shown that these modifications have the ability to inhibit as well as activate enzyme activity (30–33). While most other complex PTMs mentioned have very tightly regulated and enzyme mediated paths to modification, acylations do have enzymes that modify proteins (34), but also have the ability to occur spontaneously (25). The presence of any number of reactive acyl-CoA species (RACS) may react with and subsequently covalently modify lysine side chains under physiological conditions (25, 35). In the bacterial domain this phenomenon has also been identified with the highly reactive acetylphosphate metabolite (36, 37). This lack of specificity to which protein can be modified, and great diversity of acyl groups already identified in biological systems, means there is great importance for the comprehensive surveying of acyl modifications.

1.3 Diagnostic ions

At the nexus of mass spectrometry and PTM analysis is the important concept of diagnostic ions. While mass shifts occurring in mass spectra are historically used to sequence peptides, this single identifier alone may have ambiguity, particularly as complexity increases (38, 39). Many factors may contribute to the ambiguity, but often incomplete or “chimeric” spectra, that is a spectra containing fragments from two coeluting peptides, can result in incorrect assignments, as can isobaric or isomeric analytes under investigation (39–41). The inclusion of PTMs, including those yet unidentified, increases the theoretical search space used to identify peptide identity,

exacerbating the issue. To overcome this limitation, diagnostic ions, that is fragment ions specific to a particular feature of a precursor ion that can be used to "diagnose" its presence, can be incorporated into identification-based analysis when appropriate. While often overlooked by standard algorithms in the MS-proteomics field, they can provide substantial information to validate otherwise ambiguous spectra.

Early investigations into the fragmentation of peptides identified that standard CID fragmentation methods produced a significant portion of low-mass ions that were associated with amino acids in the peptide themselves (42). Many of the low mass ions generated by fragmentation methods utilizing collisional activation are part of the class of ions known as immonium ions (43), the abundance of which vary depending on the amino acid (42). Some residues may also have other low mass ions that present themselves in high abundance; for instance while the immonium ion of a lysine residue (*m/z* 101) is rarely in high abundance, it has derivatives that tend to be in high abundance at 84 and 129 (42, 44). Though this type of analysis on canonical amino acids may not seem as important as it does not convey sequence information as readily or clearly as sequencing larger ion fragments will, it quickly found a purpose in helping to distinguish leucine residue from its isomer, isoleucine, a feature first reported by Biemann (45), but not incorporated into common distinct workflows subsequently afterwards (46). While this feature of a peptide or protein could be applied with some confidence using a genomic database, PTMs are not hardwired into the genomic code, therefore making these types of low-mass ions critical for confident assignment.

Therefore, just as each amino acid has its own set of diagnostic ions, so too does every modified amino acid residue have ions that may allow for unique distinction (47, 48). However, given the challenges of identifying certain classes of PTMs, due to either the ambiguity of the

modification or the complexity of the class, three stand out as key examples of why diagnostic ions are a necessary feature of MS based PTM analysis; methylation, acylation, and glycosylation.

Methylation is a challenge to identify as a modification because it is simply the addition of one or more methyl groups to an amino acid side chain. This means that peptides may easily end up being isomeric with peptides that simply have different sequences, making the distinction nearly impossible with sub-optimal MS fragmentation. Compounding this issue, lower resolution instruments may have difficulty distinguishing between lysine trimethylation (+42.0470) and acetylation (+42.0106), which is a particular problem for epigenetics in which both modifications are potentially present on identical residues under different conditions (49). To overcome these issues, low-mass ions have become critical to validate the presence of methylation. With the addition of methyl group(s) on the lysine side chains, immonium ions have been used to help add confidence to the presence of a modification (50, 51). Monomethyl-lysine also have the diagnostic marker present resulting from the cyclization of the lysine immonium ion at *m/z* 98.0964, a feature that, critically, neither di- or trimethylation had (51). While trimethylation has an immonium at *m/z* 143.1543 that could be used for distinction, it also has the added benefit of a 59.0735 Da neutral loss (51) that can be used to assemble the final piece of the methylated lysine puzzle. Methylated arginine is a bit more complicated overall. While there are no diagnostic markers for monomethylated arginine, there are two for dimethylated arginine that are dependent on structural elements of the modification. The guanidinium group at the head of an arginine side chain means that there are two amine groups that can potentially be modified either once or twice. In a case where the arginine head is dimethylated symmetrically, that is one on each

amine, they will produce a diagnostic dimethyl-carbodiimidium ion at *m/z* 71.0604 (52). Asymmetric dimethylation will produce a dimethylammonium ion at *m/z* 46.0651 (52).

The study of lysine acylation has also found utility in the use of diagnostic ions. As mentioned, both trimethyl-lysine and acetyl-lysine are isobaric modifications and may be difficult or impossible to distinguish with low resolution mass spectrometers. To confidently identify acetyl-lysine ions, the cyclized immonium ion (*m/z* 126.0913) can be utilized as it is not only a more specific marker than its immonium ion at *m/z* 143.1179 (53); because trimethyl-lysine immonium ion cannot cyclize by the same mechanism, it does not result in an ion at the same nominal mass of a spectra (51), allowing for clear distinction. Additionally, the cyclized immonium ion acts as a potential marker for acyl modifications beyond the use of characterizing histone acetylation. Because acetylation is robust in cellular proteins, attempts have been made to use the ion quantitatively to understand stoichiometry of acetylation within compartments of the cell (54), highlighting ways in which this particular ion may extend beyond simple identification. However, while the acetyl-lysine immonium ion has become well characterized, the expansion of acyl-lysine modifications identified has forced consideration of this ion in modifications beyond acetylation. Recently, work was done to identify potential diagnostic ions in a number of different modifications through empirical analysis of peptide spectra (55). Using this analysis, ions specific to several types of acyl-lysine modifications were identified. All of the ions specific to these modifications corresponded to the cyclized immonium ion, as did acetylation (55). While this background covers what has been does previously, Chapter 2 will discuss new work that has been done elucidating the properties and uses of additional acyl immonium ions (56).

The identification of glycans is largely possible with the assistance of their oxonium ions that can be used in a diagnostic capacity. Glycosylated peptides are notoriously difficult to analyze, owing in part to their propensity for glycosidic bond cleavage upon CID fragmentation, which leaves the peptide intact (16). To aid in the identification of peptides that are modified with sugars, oxonium ions were identified as reporters that can distinguish between hexoses, N-acetylhexoseamines, and deoxyhexoses (57). Unlike most other classes of PTMs, glycans can form long oligomeric chains, the composition of which can vary in a heterogeneous population. The composition of complex oligosaccharide chains can also be distinguished by utilizing the profiles of diagnostic oxonium ions present (58, 59). Given the differences in profiles, as well as the complexity of modifications, there has been much consideration about how robustly these ions could be used to act as structural markers. Indeed, methods are being developed that incorporate these markers into analyses that identify glycopeptides both via optimization of instrument methods (60, 61) as well as targeted data interpretation (62, 63).

Ultimately, diagnostic ions play a critical role in mass spectrometry analysis. It can be used to distinguish isobaric or isomeric amino acids or PTMs. As the portfolio of modifications that are being reported increases, confidence in spectral identification becomes more challenging. In this way, diagnostic ions can act as confidence markers and can contribute greatly to PTM assignments in ways that may not be as necessary for peptides contain only more distinguishable amino acid components. While the analysis of most complex class of PTMs, such as glycans, utilizes diagnostic ions to a very large degree, incorporating them universally into analysis workflows may be an afterthought. For acylation in particular, the expanding number of modifications that exist, as well as the spontaneous nature in which acylation may occur, may make the incorporation of diagnostic ions a necessary or welcomed feature of their analysis.

1.4 Syntrophy in bacteria

Environmental bacteria play a critical role in bioremediation and the cycling of carbon around the globe (64–66). Fermenting microbes begin the process by breaking down larger macromolecules to their monomeric subunits, and ultimately to small molecules that are challenging to degrade. These molecules often include aromatic or fatty acids, as well as some amino acids (67). Syntrophs very often fill the role of degrading these small molecules to the more fundamental constituents of acetate, formate, hydrogen, and carbon dioxide (64). The products of syntrophs are then utilized by methanogens to generate methane that is released into the atmosphere.

Syntrophs, despite the critical role with which they play in this process, have not been as closely studied as the other groups of bacteria in this process. One of the primary reasons for this knowledge gap is the difficulty in which these microbes are to isolate in lab settings. Illustrative of this challenge is the discovery made of the once studied species of *Methanobacillus omelinskii*, an organism that was of interest due to its perceived ability to oxidize ethanol into methane (68, 69). In 1967, Bryant *et al.* demonstrated that this reaction was not carried out by a single species, but that the cultures of *M. omelinskii* were two separate but associated species that carried out two separate reactions but require the presence of the other organism to function (70). The reason for this is that one species, deemed the “S organism” oxidized ethanol in a reaction that on its own was endergonic, but when coupled with a methanogen would be driven forward (70). Thus, the concept of syntrophs, bacteria that associate by necessity to carry out energetically challenging reactions, began.

In the decades that followed, a much more widespread characterization of syntrophic bacteria began. Other bacteria were identified that had the ability to oxidize ethanol in the presence of methanogens (71–73). Substrate oxidation was soon expanded beyond ethanol. Aromatic and fatty acids were later found to be frequently degraded in anaerobic conditions by species of syntrophic bacteria (74–77). The wide variety of substrates that these bacteria can break down, as well as the diversity of species involved in these processes leaves much room for understanding the mechanisms behind these processes. Investigation of these organisms is largely focused on energy production, as many of the reactions that are completed by the organism would, under standard state conditions, would require an input of energy into the system (67), making it difficult for these organisms to survive and reproduce. One of the unique mechanisms by which organisms can do this has been found in the syntrophic species *Syntrophus aciditrophicus*, whereby acetyl-CoA can be used to generate adenosine triphosphate (ATP) (78). This unique feature highlights not only the energy limited state of the cell, but the abundance and importance of acyl-CoA intermediates in these species. The oxidation of aromatic and aliphatic acids in bacteria typically requires the formation of a thioester with CoA, which scaffolds intermediates throughout the degradation process (79, 80). This produces a wide variety of RACS, which have the potential to act as regulatory elements in the bacterial proteomes. Despite the known function of some modifications (32, 81, 82) and the potential function for other modifications, few systems have been studied for the presence or role of these modifications. Technological limitations in how these systems are analyzed have played a role in the inability of many groups to take a comprehensive look at bacterial acyl-proteomes.

While much of the work to follow will focus on digging deep into bacterial proteomes to fully understand acyl-modifications, this work will also highlight how systems can be studied at the community level. As demonstrated by the story of *M. omelinskii*, bacterial communities may have emergent properties that a single species in lab conditions cannot recapitulate. This anecdote, and the existence of syntrophic bacteria at large, highlight the necessity of understanding organisms, environments, and interactions at increasing scales to fully understand how complex biological processes occur. Increasingly the study of microbiomes at the community scale have been utilized to understand how environmental and medical processes occur (83–85). The expansion of tools and considerations in studying these systems may help to determine mechanisms underlying larger and more complex processes.

1.5 References

1. Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., and Whitehouse, C. M. (1989) Electrospray ionization for mass spectrometry of large biomolecules. *Science.* **246**, 64–71
2. Karas, M., and Hillenkamp, F. (1988) Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10 000 Daltons. *Anal. Chem.* **60**, 2299–2301
3. Biemann, K. (1988) Contributions of mass spectrometry to peptide and protein structure. *Biol. Mass Spectrom.* **16**, 99–111
4. Syka, J. E. P., Coon, J. J., Schroeder, M. J., Shabanowitz, J., and Hunt, D. F. (2004) Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 9528–9533
5. Zubarev, R. A., Kelleher, N. L., and McLafferty, F. W. (1998) Electron capture dissociation of multiply charged protein cations. A nonergodic process. *J. Am. Chem. Soc.* **120**, 3265–3266
6. Ly, T., and Julian, R. R. (2009) Ultraviolet photodissociation: developments towards applications for mass-spectrometry-based proteomics. *Angew. Chemie - Int. Ed.* **48**, 7130–7137
7. Aebersold, R., and Mann, M. (2003) Mass spectrometry-based proteomics. *Nature*. **422**, 198–207
8. Eng, J. K., McCormack, A. L., and Yates, J. R. (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5**, 976–989
9. Perkins, D. N., Pappin, D. J. C., Creasy, D. M., and Cottrell, J. S. (1999) Probability-based

protein identification by searching sequence databases using mass spectrometry data. in *Electrophoresis*, pp. 3551–3567, Wiley-VCH Verlag, **20**, 3551–3567

10. Tyanova, S., Temu, T., and Cox, J. (2016) The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.* **11**, 2301–2319
11. Schirmer, E. C., Yates-Iii, J. R., and Gerace, L. (2009) MudPIT: A Powerful Proteomics Tool for Discovery. *Discov. Med.* **2**, 38–39
12. Ye, D., Fu, Y., Sun, R. X., Wang, H. P., Yuan, Z. F., Chi, H., and He, S. M. (2010) Open MS/MS spectral library search to identify unanticipated post-translational modifications and increase spectral identification rate. *Bioinformatics*. 10.1093/bioinformatics/btq185
13. Bittremieux, W., Meysman, P., Noble, W. S., and Laukens, K. (2018) Fast Open Modification Spectral Library Searching through Approximate Nearest Neighbor Indexing. *J. Proteome Res.* **17**, 3463–3474
14. Kong, A. T., Leprevost, F. V., Avtonomov, D. M., Mellacheruvu, D., and Nesvizhskii, A. I. (2017) MSFragger: Ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nat. Methods*. **14**, 513–520
15. Skinner, O. S., and Kelleher, N. L. (2015) Illuminating the dark matter of shotgun proteomics. *Nat. Biotechnol.* **33**, 717–718
16. Nilsson, J., Halim, A., Grahn, A., and Larson, G. (2013) Targeting the glycoproteome. *Glycoconj. J.* **30**, 119–136
17. Aswad, D. W., Paranandi, M. V., and Schurter, B. T. (2000) Isoaspartate in peptides and proteins: Formation, significance, and analysis. *J. Pharm. Biomed. Anal.* **21**, 1129–1136
18. Baker, M. D., Wolanin, P. M., and Stock, J. B. (2006) Signal transduction in bacterial chemotaxis. *BioEssays*. **28**, 9–22

19. Humphrey, S. J., James, D. E., and Mann, M. (2015) Protein Phosphorylation: A Major Switch Mechanism for Metabolic Regulation. *Trends Endocrinol. Metab.* **26**, 676–687

20. Eyre, D. R., Paz, M. A., and Gallop, P. M. (1984) Cross-Linking in Collagen and Elastin. *Annu. Rev. Biochem.* **53**, 717–748

21. Hae, J. K., Coulibaly, F., Clow, F., Proft, T., and Baker, E. N. (2007) Stabilizing isopeptide bonds revealed in gram-positive bacterial pilus structure. *Science (80-.).* **318**, 1625–1628

22. Lecker, S. H., Goldberg, A. L., and Mitch, W. E. (2006) Protein degradation by the ubiquitin-proteasome pathway in normal and disease states. *J. Am. Soc. Nephrol.* **17**, 1807–1819

23. Pinkse, M. W. H., Uitto, P. M., Hilhorst, M. J., Ooms, B., and Heck, A. J. R. (2004) Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-NanoLC-ESI-MS/MS and titanium oxide precolumns. *Anal. Chem.* **76**, 3935–3943

24. Xu, G., Paige, J. S., and Jaffrey, S. R. (2010) Global analysis of lysine ubiquitination by ubiquitin remnant immunoaffinity profiling. *Nat. Biotechnol.* 10.1038/nbt.1654

25. Trub, A. G., and Hirschey, M. D. (2018) Reactive Acyl-CoA Species Modify Proteins and Induce Carbon Stress. *Trends Biochem. Sci.* **43**, 369–379

26. Wagner, G. R., and Hirschey, M. D. (2014) Nonenzymatic Protein Acylation as a Carbon Stress Regulated by Sirtuin Deacylases. *Mol. Cell*. **54**, 5–16

27. Huang, H., Zhang, D., Wang, Y., Perez-Neut, M., Han, Z., Zheng, Y. G., Hao, Q., and Zhao, Y. (2018) Lysine benzoylation is a histone mark regulated by SIRT2. *Nat. Commun.* **9**, 1–11

28. Guan, K. L., Yu, W., Lin, Y., Xiong, Y., and Zhao, S. (2010) Generation of acetyllysine antibodies and affinity enrichment of acetylated peptides. *Nat. Protoc.* **5**, 1583–1595

29. Turner, B. M. (2000) Histone acetylation and an epigenetic code. *BioEssays*. **22**, 836–845

30. Cain, J. A., Solis, N., and Cordwell, S. J. (2014) Beyond gene expression: The impact of protein post-translational modifications in bacteria. *J. Proteomics*. **97**, 265–286

31. Christensen, D. G., Xie, X., Basisty, N., Byrnes, J., McSweeney, S., Schilling, B., and Wolfe, A. J. (2019) Post-translational Protein Acetylation: An elegant mechanism for bacteria to dynamically regulate metabolic functions. *Front. Microbiol.* **10**, 1604

32. Garrity, J., Gardner, J. G., Hawse, W., Wolberger, C., and Escalante-Semerena, J. C. (2007) N-lysine propionylation controls the activity of propionyl-CoA synthetase. *J. Biol. Chem.* **282**, 30239–30245

33. Ott, M., Schnölzer, M., Garnica, J., Fischle, W., Emiliani, S., Rackwitz, H. R., and Verdin, E. (1999) Acetylation of the HIV-1 tat protein by p300 is important for its transcriptional activity. *Curr. Biol.* **9**, 1489–1493

34. McClure, J. J., Inks, E. S., Zhang, C., Peterson, Y. K., Li, J., Chundru, K., Lee, B., Buchanan, A., Miao, S., and Chou, C. J. (2017) Comparison of the Deacylase and Deacetylase Activity of Zinc-Dependent HDACs. *ACS Chem. Biol.* **12**, 1644–1655

35. Baeza, J., Smallegan, M. J., and Denu, J. M. (2015) Site-Specific Reactivity of Nonenzymatic Lysine Acetylation. *ACS Chem. Biol.* **10**, 122–128

36. VanDrisse, C. M., and Escalante-Semerena, J. C. (2019) Protein Acetylation in Bacteria. *Annu. Rev. Microbiol.* **73**, 111–132

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

38. Fu, Y. (2012) Bayesian false discovery rates for post-translational modification proteomics. *Stat. Interface*. **5**, 47–59

39. Kim, M. S., Zhong, J., and Pandey, A. (2016) Common errors in mass spectrometry-based analysis of post-translational modifications. *Proteomics*. **16**, 700–714

40. Lee, S., Tan, M., Dai, L., Kwon, O. K., Yang, J. S., Zhao, Y., and Chen, Y. (2013) MS/MS of synthetic peptide is not sufficient to confirm new types of protein modifications. *J. Proteome Res.* **12**, 1007–1013

41. El Kennani, S., Crespo, M., Govin, J., and Pflieger, D. (2018) Proteomic analysis of histone variants and their PTMs: Strategies and pitfalls. *Proteomes*. 10.3390/proteomes6030029

42. Falick, A. M., Hines, W. M., Medzihradszky, K. F., Baldwin, M. A., and Gibson, B. W. (1993) Low-mass ions produced from peptides by high-energy collision-induced dissociation in tandem mass spectrometry. *J. Am. Soc. Mass Spectrom.* **4**, 882–893

43. Renner, D., and Spiteller, G. (1986) Mechanism of fragmentation reactions of [MH]+ ions obtained from peptides by liquid secondary ion mass spectrometry. *Biol. Mass Spectrom.* **13**, 405–410

44. Zhang, P., Chan, W., Ang, I. L., Wei, R., Lam, M. M. T., Lei, K. M. K., and Poon, T. C. W. (2019) Revisiting Fragmentation Reactions of Protonated α-Amino Acids by High-Resolution Electrospray Ionization Tandem Mass Spectrometry with Collision-Induced Dissociation. *Sci. Rep.* **9**, 1–10

45. Biemann, K., Seibl, J., and Gapp, F. (1961) Mass Spectra of Organic Molecules. I. Ethyl Esters of Amino Acids. *J. Am. Chem. Soc.* **83**, 3795–3804

46. Armirotti, A., Millo, E., and Damonte, G. (2006) How to Discriminate Between Leucine

and Isoleucine by Low Energy ESI-TRAP MS n. 10.1016/j.jasms.2006.08.011

47. Hung, C. W., Schlosser, A., Wei, J., and Lehmann, W. D. (2007) Collision-induced reporter fragmentations for identification of covalently modified peptides. *Anal. Bioanal. Chem.* **389**, 1003–1016

48. Gehrig, P. M., Nowak, K., Panse, C., Leutert, M., Grossmann, J., Schlapbach, R., and Hottiger, M. O. (2020) Gas-Phase Fragmentation of ADP-Ribosylated Peptides: Arginine-Specific Side-Chain Losses and Their Implication in Database Searches. 10.1021/jasms.0c00040

49. Bowman, G. D., and Poirier, M. G. (2015) Post-translational modifications of histones that influence nucleosome dynamics. *Chem. Rev.* **115**, 2274–2295

50. Couttas, T. A., Raftery, M. J., Bernardini, G., and Wilkins, M. R. (2008) Immonium Ion Scanning for the Discovery of Post-Translational Modifications and Its Application to Histones. *J. Proteome Res.* **7**, 2632–2641

51. Zhang, K., Yau, P. M., Chandrasekhar, B., New, R., Kondrat, R., Imai, B. S., and Bradbury, M. E. (2004) Differentiation between peptides containing acetylated or tri-methylated lysines by mass spectrometry: An application for determining lysine 9 acetylation and methylation of histone H3. *Proteomics*. **4**, 1–10

52. Brame, C. J., Moran, M. F., and McBroom-Cerajewski, L. D. B. (2004) A mass spectrometry based method for distinguishing between symmetrically and asymmetrically dimethylated arginine residues. *Rapid Commun. Mass Spectrom.* **18**, 877–881

53. Trelle, M. B., and Jensen, O. N. (2008) Utility of immonium ions for assignment of ε-N-acetyllysine-containing peptides by tandem mass spectrometry. *Anal. Chem.* **80**, 3422–3430

54. Nakayasu, E. S., Wu, S., Sydor, M. A., Shukla, A. K., Weitz, K. K., Moore, R. J., Hixson, K. K., Kim, J.-S., Petyuk, V. A., Monroe, M. E., Pasa-Tolic, L., Qian, W.-J., Smith, R. D., Adkins, J. N., and Ansong, C. (2014) A Method to Determine Lysine Acetylation Stoichiometries. *Int. J. Proteomics*. **2014**, 1–8

55. Zolg, D. P., Wilhelm, M., Schmidt, T., Médard, G., Zerweck, J., Knaute, T., Wenschuh, H., Reimer, U., Schnatbaum, K., and Kuster, B. (2018) ProteomeTools: Systematic Characterization of 21 Post-translational Protein Modifications by Liquid Chromatography Tandem Mass Spectrometry (LC-MS/MS) Using Synthetic Peptides. *Mol. Cell. Proteomics*. **17**, 1850–1863

56. Muroski, J. M., Fu, J. Y., Nguyen, H. H., Ogorzalek Loo, R. R., and Loo, J. A. (2021) Leveraging Immonium Ions for Targeting Acyl-Lysine Modifications in Proteomic Datasets. *Proteomics*. **21**, 2000111

57. Carr, S. A., Huddleston, M. J., and Bean, M. E. (1993) *Selective identification and differentiation of N-and O-linked oligosaccharides in glycoproteins by liquid chromatography-mass spectrometry*, Cambridge University Press

58. Halim, A., Westerlind, U., Pett, C., Schorlemer, M., Rüetschi, U., Brinkmalm, G., Sihlbom, C., Lengqvist, J., Larson, G., and Nilsson, J. (2014) Assignment of saccharide identities through analysis of oxonium ion fragmentation profiles in LC-MS/MS of glycopeptides. *J. Proteome Res.* **13**, 6024–6032

59. Yu, J., Schorlemer, M., Gomez Toledo, A., Pett, C., Sihlbom, C., Larson, G., Westerlind, U., and Nilsson, J. (2016) Distinctive MS/MS Fragmentation Pathways of Glycopeptide-Generated Oxonium Ions Provide Evidence of the Glycan Structure. *Chem. - A Eur. J.* **22**, 1114–1124

60. Ritchie, M. A., Gill, A. C., Deery, M. J., and Lilley, K. (2002) Precursor ion scanning for detection and structural characterization of heterogeneous glycopeptide mixtures. *J. Am. Soc. Mass Spectrom.* **13**, 1065–1077

61. Madsen, J. A., Farutin, V., Lin, Y. Y., Smith, S., and Capila, I. (2018) Data-independent oxonium ion profiling of multi-glycosylated biotherapeutics. *MAbs*. **10**, 968–978

62. Stadlmann, J., Taubenschmid, J., Wenzel, D., Gattinger, A., Dürnberger, G., Dusberger, F., Elling, U., Mach, L., Mechtler, K., and Penninger, J. M. (2017) Comparative glycoproteomics of stem cells identifies new players in ricin toxicity. *Nature*. **549**, 538–542

63. Bollineni, R. C., Koehler, C. J., Gislefoss, R. E., Anonsen, J. H., and Thiede, B. (2018) Large-scale intact glycopeptide identification by Mascot database search. *Sci. Rep.* 10.1038/s41598-018-20331-2

64. McInerney, M. J., Sieber, J. R., and Gunsalus, R. P. (2009) Syntrophy in anaerobic global carbon cycles. *Curr. Opin. Biotechnol.* **20**, 623–632

65. Stams, A. J. M., Sousa, D. Z., Kleerebezem, R., and Plugge, C. M. (2012) Role of syntrophic microbial communities in high-rate methanogenic bioreactors. *Water Sci. Technol.* **66**, 352–362

66. Boll, M., Geiger, R., Junghare, M., and Schink, B. (2020) Microbial degradation of phthalates: biochemistry and environmental implications. *Environ. Microbiol. Rep.* **12**, 3–15

67. Schink, B. (1997) Energetics of syntrophic cooperation in methanogenic degradation. *Microbiol. Mol. Biol. Rev.* **61**, 262–80

68. Heukelekian, H., and Heinemann, B. (1939) Studies on the Methane-Producing Bacteria:

I. Development of a Method for Enumeration on JSTOR. *Sewage Work. J.* **11**, 426–435

69. Barker, H. A. (1941) Studies on the methane fermentation V. Biochemical activities of Methanobacterium omelianskii. *J. Biol. Chem.* **137**, 153–167

70. Bryant, M. P., Wolin, E. A., Wolin, M. J., and Wolfe, R. S. (1967) Methanobacillus omelianskii, a symbiotic association of two species of bacteria. *Arch. Mikrobiol.* **59**, 20–31

71. Ben-Bassat, A., Lamed, R., and Zeikus, J. G. (1981) Ethanol production by thermophilic bacteria: Metabolic control of end product formation in Thermoanaerobium brockii. *J. Bacteriol.* **146**, 192–199

72. Schink, B. (1984) Fermentation of 2,3-butanediol by Pelobacter carbinolicus sp. nov. and Pelobacter propionicus sp. nov., and evidence for propionate formation from C2 compounds. *Arch. Microbiol.* **137**, 33–41

73. Bryant, M. P., Campbell, L. L., Reddy, C. A., and Crabill, M. R. (1977) Growth of desulfovibrio in lactate or ethanol media low in sulfate in association with H2 utilizing methanogenic bacteria. *Appl. Environ. Microbiol.* **33**, 1162–1169

74. McInerney, M. J., Bryant, M. P., Hespell, R. B., and Costerton, J. W. (1981) Syntrophomonas wolfei gen. nov. sp. nov., an Anaerobic, Syntrophic, Fatty Acid-Oxidizing Bacterium. *Appl. Environ. Microbiol.* **41**, 1029–39

75. Boone, D. R., and Bryant, M. P. (1980) Propionate-degrading bacterium, Syntrophobacter wolinii sp. nov. gen. nov., from methanogenic ecosystems. *Appl. Environ. Microbiol.* **40**, 626–632

76. Mountfort, D. O., and Bryant, M. P. (1982) Isolation and characterization of an anaerobic syntrophic benzoate-degrading bacterium from sewage sludge. *Arch. Microbiol.* **133**, 249–

256

77. Jackson, B. E., Bhupathiraju, V. K., Tanner, R. S., Woese, C. R., and McInerney, M. J. (1999) Syntrophus aciditrophicus sp. nov., a new anaerobic bacterium that degrades fatty acids and benzoate in syntrophic association with hydrogen- using microorganisms. *Arch. Microbiol.* **171**, 107–114

78. James, K. L., Ríos-Hernández, L. A., Wofford, N. Q., Mouttaki, H., Sieber, J. R., Sheik, C. S., Nguyen, H. H., Yang, Y., Xie, Y., Erde, J., Rohlin, L., Karr, E. A., Loo, J. A., Loo, R. R. O., Hurst, G. B., Gunsalus, R. P., Szweda, L. I., and McInerney, M. J. (2016) Pyrophosphate-dependent ATP formation from acetyl coenzyme a in Syntrophus aciditrophicus, a new twist on ATP formation. *MBio*. 10.1128/mBio.01208-16

79. Heider, J., and Fuchs, G. (1997) Anaerobic metabolism of aromatic compounds. *Eur. J. Biochem.* **243**, 577–596

80. Müller, N., Worm, P., Schink, † Bernhard, Stams, A. J. M., and Plugge, C. M. (2010) Syntrophic butyrate and propionate oxidation processes: from genomes to reaction mechanismse mi4_147 489..499. 10.1111/j.1758-2229.2010.00147.x

81. Crosby, H. A., Heiniger, E. K., Harwood, C. S., and Escalante-Semerena, J. C. (2010) Reversible Nε-lysine acetylation regulates the activity of acyl-CoA synthetases involved in anaerobic benzoate catabolism in Rhodopseudomonas palustris. *Mol. Microbiol.* **76**, 874–888

82. Gardner, J. G., Grundy, F. J., Henkin, T. M., and Escalante-Semerena, J. C. (2006) Control of Acetyl-Coenzyme A Synthetase (AcsA) Activity by Acetylation/Deacetylation without NAD+ Involvement in Bacillus subtilis. *J. Bacteriol.* **188**, 5460–5468

83. Rocca, J. D., Simonin, M., Blaszczak, J. R., Ernakovich, J. G., Gibbons, S. M., Midani, F.

S., and Washburne, A. D. (2019) The Microbiome Stress Project: Toward a Global Meta-Analysis of Environmental Stressors and Their Effects on Microbial Communities. *Front. Microbiol.* **9**, 3272

84. Trinh, P., Zaneveld, J. R., Safranek, S., and Rabinowitz, P. M. (2018) One Health Relationships Between Human, Animal, and Environmental Microbiomes: A Mini-Review. *Front. Public Heal.* **6**, 235

85. Tasnim, N., Abulizi, N., Pither, J., Hart, M. M., and Gibson, D. L. (2017) Linking the gut microbial ecosystem with the environment: Does gut health depend on where we live? *Front. Microbiol.* 10.3389/fmicb.2017.01935

CHAPTER 2

Leveraging Immonium Ions for Targeting Acyl-Lysine Modifications in Proteomic Datasets

The following is a reprint of research article from

Proteomics

08 September 2020

DOI: 10.1002/pmic.202000111

RESEARCH ARTICLE

Proteomics
Proteomics and Systems Biology
www.proteomics-journal.com

Leveraging Immonium Ions for Targeting Acyl-Lysine Modifications in Proteomic Datasets

John M. Muroski, Janine Y. Fu, Hong Hanh Nguyen, Rachel R. Ogorzalek Loo, and Joseph A. Loo**

Acyl modifications vary greatly in terms of elemental composition and site of protein modification. Developing methods to identify acyl modifications more confidently can help to assess the scope of these modifications in large proteomic datasets. The utility of acyl-lysine immonium ions is analyzed for identifying the modifications in proteomic datasets. It is demonstrated that the cyclized immonium ion is a strong indicator of acyl-lysine presence when its rank or relative abundance compared to other ions within a spectrum is considered. Utilizing a stepped collision energy method in a shotgun experiment highlights the immonium ion. By implementing an analysis that accounted for features within each MS^2 spectrum, the method clearly identifies peptides with short chain acyl-lysine modifications from complex lysates. Immonium ions can also be used to validate novel acyl modifications; in this study, the first examples of 3-hydroxylpimelyl-lysine modifications are reported and they are validated using immonium ions. Overall these results solidify the use of the immonium ion as a marker for acyl-lysine modifications in complex proteomic datasets.

1. Introduction

As proteomic technology advances, the depth and breadth of proteomic data that can be analyzed has increased concurrently. This increased scope of proteomic information has resulted in a plethora of data on post-translational modifications (PTMs), which can affect protein function and modulate protein activity in a cell without the energetic burden of synthesizing new proteins.[1–5] These modifications fall into different classes based on their physicochemical properties, such as phosphorylation, glycosylation, oxidation, and many more. One class of modification that has been observed across biological systems is lysine acylation.[6–8] Acetylation has been long known as an epigenetic regulator, modulating protein expression through the modified lysine side chains on histone tails.[9–11] Later, it was discovered that acetyl and other acyl modifications not only impact histone function, but are also ubiquitous in mammalian metabolic pathways as well as in other eukaryotic and prokaryotic systems.[6,9] In eukaryotic organisms, these modifications have been shown to correlate with aging as well as metabolic state.[7,8] In prokaryotes, these modifications impact the activity of enzymes in metabolic pathways.[12,13] Acylation even plays a role in viruses, as viral proteins have been shown to require acetylation to function.[14]

Acyl modifications have proven to be challenging to identify on a proteome-wide scale for a number of reasons. These modifications tend to have low stoichiometries relative to unmodified proteoforms; for example, mitochondrial acylation stoichiometries are around 0.02%.[15] This low abundance makes it difficult to consistently identify modified peptides in complex samples using untargeted proteomic methods, and enrichment strategies during sample preparation are often necessary.[16] A second challenge is that acyl modifications come in many different varieties. Chain length, elemental composition, and degree of unsaturation differ between different types of acylations.[4] For example, an acetyl group and a propionyl group differ by one carbon. Although it is often unclear whether these modifications assume different functions, delineating their presence is important. However, creating such a list requires knowing a priori which acyl compositions might be present. Acylation can occur spontaneously; that is, reactive intermediates such as acetyl phosphate or an acyl-CoA can modify certain primary amines without enzymatic catalysis.[17] Such processes suggest that a given residue could be tagged by many acyl

J. M. Muroski, J. Y. Fu, Dr. H. H. Nguyen, Prof. J. A. Loo
Department of Chemistry and Biochemistry
University of California
Los Angeles, CA 90095, USA
E-mail: JLoo@chem.ucla.edu
Dr. R. R. Ogorzalek Loo, Prof. J. A. Loo
David Geffen School of Medicine, Department of Biological Chemistry
University of California
Los Angeles, CA 90095, USA
E-mail: RLoo@mednet.ucla.edu
Dr. R. R. Ogorzalek Loo, Prof. J. A. Loo
UCLA-DOE Institute
University of California
Los Angeles, CA 90095, USA
Dr. R. R. Ogorzalek Loo, Prof. J. A. Loo
UCLA Molecular Biology Institute
University of California
Los Angeles, CA 90095, USA
Dr. H. H. Nguyen
TRANSMED Co. Ltd.
Ho Chi Minh City Vietnam

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/pmic.202000111

DOI: 10.1002/pmic.202000111

compositions, even on a single peptide. Not including all possible acyl modifications during database searching in proteomic workflows may miss critical information about a given protein, particularly when quantification is sought. Broadening the number of modifications considered, however, can increase the false discovery rate (FDR) or, if FDR is treated properly, reduce the sensitivity for identifying peptides. All of these challenges add to the difficulty in characterizing the full range of acyl modifications.

Physicochemical differences in PTMs have meant that different experimental strategies may need to be considered if each modification is to be detected and localized optimally. For example, peptide phosphates are often labile to collision induced dissociation (CID) and the associated mass shift may not be apparent in MS^2 spectra or the PTM may migrate to another residue.[18] To overcome this limitation, phosphoproteomic studies have adopted strategies that utilize electron transfer dissociation (ETD) as an alternative MS^2 dissociation method.[18,19] Glycosylated peptides also present challenges for MS fragmentation given the fragility and complexity of their structures. One strategy for identifying these modifications uses collision energy stepping in conjunction with detecting the characteristic oxonium ions associated with specific glycans.[20] Oxonium ions[21–23] are low mass ions resulting from fragmentation of oligosaccharides and glycopeptides.

Similarly, immonium ions can serve as diagnostic markers for specific acyl-lysine PTMs.[24,25] Immonium ions are internal product ions resulting from two-bond cleavages that retain a single amino acid side chain. These ions, common in peptide tandem mass spectra, can often verify the presence of certain amino acids.[26,27] It has been shown that, in addition to the canonical amino acids, immonium ions can be generated for acyl-lysine residues. The immonium ion for acetyl-lysine is observed at m/z 143.1179; however, a related ion observed at m/z 126.0913 that originates from cyclization has shown special utility.[24,28] Other acyl-lysines present unique immonium ions.[25] These diagnostic indicators are typically used for post-identification validation, because their presence and intensity depend heavily on sequence context and instrument parameters. Here, we posit a means to overcome limitations in the use of immonium ions by optimizing collision energies and by defining comparative criteria. This strategy can be extended to identifying novel acyl PTMs in large proteomic data sets.

Significance Statement

Acyl-lysine modifications come in a variety of elemental compositions. There is increasing evidence that these modifications can have a functional effect on proteins and are present in species across all domains of life. Here, we describe a new method that can allow for more confident identification of acyl modifications in proteomes by utilizing their immonium ions. We also report the first observation of a 3-hydroxypimelyl lysine.

2. Experimental Section

2.1. Reagents and Materials

Synthetic peptides (lyophilized, >95% purity) were obtained from Genscript, Inc. and reconstituted in water. The sequences, which originate from *Syntrophus aciditrophicus* are: **K**STPEAMAK, F**K**DEIPVVIK, STDP**K**GPSVR, with the lysine residues indicated in bold containing an acetyl-, butyryl-, or crotonyl-modification on the ϵ-amine.

2.2. Preparation and Digestion of Acyl-Bovine Serum Albumin

Acetylated-bovine serum albumin (BSA, Promega Product #R3961) was diluted in 100 mM ammonium bicarbonate. Butyrylated-BSA was prepared in a process adapted from Baez, et al.[29] Butyric anhydride (~25 µmol) was added to 100 µL of a 1 mg mL^{-1} solution of BSA (Sigma Aldrich, Product #A8022) in 100 mM ammonium bicarbonate. The solution was incubated for 20 min at 4 °C, after which the solution pH was adjusted to ~8 using ammonium hydroxide. The incubation/pH adjustment process was repeated two more times. To reverse adventitious *O*-acylation, hydroxylamine hydrochloride was added to 50% w/v of the final concentration (e.g., 55 mg added to 110 µL) and the pH was readjusted with NH_4OH to ~8. The solution was incubated at room temperature overnight. Butyrylated-BSA was then buffer exchanged into 100 mM ammonium bicarbonate using 10 kD MWCO Amicon spin filters (Millipore).

Acylated-BSA preparations were heated to 95 °C for 10 min, disulfide-reduced with 20 mM dithiothreitol (DTT) for 1 h at 60 °C, and alkylated with 50 mM iodoacetamide for 45 min at room temperature in the dark. Excess iodoacetamide was quenched with DTT and the samples were digested overnight with endoproteinase GluC (1:100) at room temperature (New England Biolabs, Product #P8100S). Digested peptides were dried in a vacuum concentrator, acidified with 0.1% acetic acid, desalted with STAGE tips assembled from 3M Empore C18 Solid Phase Extraction Disks,[30] and dried again. Peptides were reconstituted in LC-MS injection buffer (3% acetonitrile, 0.1% formic acid) and quantified by Pierce Quantitative Fluorometric Peptide Assay (Thermo Scientific, Product #23290).

2.3. Peptide Preparation from *S. aciditrophicus* Cells

Cells were harvested from tricultures of *S. aciditrophicus*, *Methanosaeta concilli*, and *Methanospirillum hungatei* grown with benzoate as the carbon source. Peptides were prepared from cell pellets using enhanced filter-aided sample preparation (eFASP) as described by Erde et al.[31,32] Briefly, cells were lysed in 4.0% v/v ammonium lauryl sulfate, 0.1% w/v sodium deoxycholate, and 5 mM tris(2-carboxyethyl)phosphine in 100 mM ammonium bicarbonate. The lysate was exchanged into a buffer containing 8 M urea, 0.1% w/v sodium deoxycholate, and 0.1% w/v *n*-octyl glucoside using a 10 kDa Microcon ultrafiltration unit (Millipore). Within the ultrafiltration unit, proteins were alkylated in 17 mM iodoacetamide and digested overnight at 37 °C with trypsin in a buffer containing 0.1% w/v sodium deoxycholate, 0.1% w/v *n*-octyl glucoside, and 100 mM ammonium bicarbonate. Peptides were desalted with STAGE tips, as described earlier.

2.4. Mass Spectrometry (LC-MS/MS) Analysis

Processed peptides were measured by reversed phase liquid chromatography-tandem mass spectrometry (LC-MS/MS) on an EASY nLC1000 (Thermo Scientific) coupled to a quadrupole orbitrap mass spectrometer (Q-Exactive, Thermo Scientific). Acyl-BSA peptides (100 ng) were loaded onto an Acclaim PepMap100 C18 trap column (Thermo Scientific, Product #16-494-6, 75 μm × 2 cm, 100 Å) and separated on an Acclaim PepMap RSLC C18 analytical column (Thermo Scientific, Product #03-251-873, 75 μm × 25 cm, 100 Å). Buffer A (0.1% formic acid) and buffer B (0.1% formic acid in 100% acetonitrile) were employed in the 300 nL min^{-1} gradient: 3–35% B in 30 min, 35–50% B in 5 min, and 50–80% B in 2 min.

Synthetic, acylated peptide standards (1 fmol of each) were spiked into 100 ng of a HeLa tryptic digest standard (Thermo Scientific, #PI88329). For the HeLa digest, 100 ng was loaded onto the column, whereas 200 ng of the triculture digest (containing *S. aciditrophicus* and two other organisms) was loaded. HeLa and triculture analyses used the gradient 3–20% B in 62 min, 20–30% B in 31 min, 30–50% in 5 min, and 50–80% in 2 min.

The mass spectrometer was operated in a data-dependent acquisition mode (m/z 300–1800). MS scans were acquired at 70 000 resolution using an automatic gain control (AGC) target of 1E6 (maximum fill: 100 ms). Collision induced dissociation MS/MS spectra were acquired at 17 500 resolution, AGC (maximum fill: 80 ms) of 1E5, and a normalized collision energy of 27 (unless otherwise indicated) on the top 10 abundant precursor ions. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD018758.

2.5. Proteomic Data Analysis

2.5.1. Acyl-BSA Data

RAW files were converted into MGF format and peak lists were submitted to Mascot (version 2.5; Matrix Science) and searched against the BSA sequence supplemented with protein sequences of common contaminants. GluC was specified as the cleavage enzyme with up to six missed cleavages considered, and a precursor mass tolerance of 10 ppm and product mass error of 0.02 Da. Cysteine carbamidomethylation (+57.021464), methionine oxidation (+15.994915), and the respective acyl-lysine modification, acetyl (+42.010565) or butyryl (+70.041865) were set as variable modifications. Peptide spectral matches (PSMs) were filtered to 1% false discovery rate using the target-decoy strategy.

2.5.2. Acyl Modifications in HeLa and S. aciditrophicus

All HeLa cell data was analyzed using Mascot (version 2.5). Files were searched against the UniProt Human database (as of January 23, 2019) supplemented with common laboratory contaminants and the three spiked synthetic peptide sequences. The *S. aciditrophicus* triculture RAW files were processed through ProteomeDiscoverer (version 1.4), using Mascot for the database search. Files were searched against UniProt *S. aciditrophicus, M. concilli, and M. hungatei* sequence databases that were concatenated and supplemented with contaminant sequences (as of July 8, 2019). The search parameters for both datasets were: enzyme specificity, trypsin; maximum number of missed cleavages, 2; precursor mass tolerance, 10 ppm; product mass error allowed, 0.02 Da; variable modifications included cysteine carbamidomethylation, methionine oxidation, lysine acetylation, lysine butyrylation, and lysine crotonylation (+68.026215). PSMs were filtered to 1% false discovery rate using the target-decoy strategy.

2.5.3. Immonium Ion Analysis for Acylated BSA

From the Mascot search results (DAT files), a more stringent secondary filter was applied to the data to increase the confidence in PSMs identified. PSMs with an ion score >25 were considered for immonium ion analysis and an in-house Python script was utilized to extract the corresponding MS/MS spectra containing the immonium ion of interest for further characterization. For all datasets, the mass tolerance was set to 10 ppm for the immonium ion of interest. Similar analysis was performed on the Svinkina et al.[33] data set (MassIVE MSV000079068).

3. Results

3.1. The *m/z* 126 Immonium Ion is Present in Great Abundance in an Acetyl-Lysine Data Set

Previous reports have indicated that the m/z 126 ion (126.0913) (hereafter called the "126 ion") is diagnostic for acetyl-lysine.[24] To determine the prevalence of the 126 ion in PSMs, we investigated peptides from GluC-digested acetylated BSA, focused on comparing matches with and without acetyl-lysine, as indicated by the Mascot search. Of those identified as acetylated (335/571), 97.3% displayed the 126 ion in MS/MS spectra (**Table 1**). However, the 126 ion was also present in 65.7% of non-acetylated spectra. There are several possible ways that non-acetyl-lysine containing peptides can yield an ion with the exact mass of the 126 ion, including *"a*-ion" type products that subsequently lose NH_3 from sequences containing Gly-Ile, Gly-Leu, or Ala-Val and the reversed sequences. To verify that the observation of 126 ions from non-acetylated peptides also occurs with complex lysates, we performed the same analyses on a published dataset that utilized immunoprecipitation to enrich acetylated peptides from Jurkat E6-1 cells.[33] From this dataset, 96% of acetylated PSMs presented a 126 ion, but 73.9% of the non-acetylated PSMs also displayed a 126 product (Table 1). That the 126 ion appears in acetylated PSMs agrees with previous work detailing its sensitivity as an acetyl-lysine marker, but its presence in non-acetylated PSMs raises questions about its specificity.[24] This difference may reflect the higher sensitivity of current mass spectrometers as compared to those used previously.[24] It may also reflect the previous study's metrics for determining false positives, which relied on the presence of specific dipeptide cleavages that may have been absent in spectra with limited sequence information.[24]

Table 1. Distribution of PSMs containing m/z 126 and 129 immonium ions from acetylated BSA (27 NCE) and from Jurkat E6-1 cells (25 NCE) (Svinkina et al.)[33].

		PSMs with 126 ion		PSMs without 126 ion		PSMs with 129 ion		PSMs without 129 ion	
Acetyl-BSA	Non-acetyl	155	(65.7%)	81	(34.3%)	212	(89.8%)	24	(10.2%)
	Acetyl	326	(97.3%)	9	(2.7%)	304	(90.7%)	31	(9.3%)
Svinkina et al.	Non-acetyl	4458	(73.9%)	1577	(26.1%)	5806	(96.2%)	229	(3.8%)
	Acetyl	4415	(96%)	185	(4%)	4255	(92.5%)	345	(7.5%)

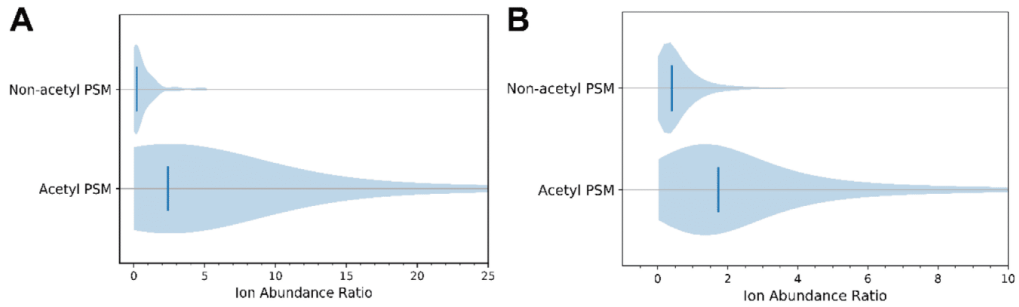


Figure 1. Ion abundance ratios [126]/[129] for acetylated and non-acetylated spectra. A) Violin plots of the [126]/[129] abundance ratio for all PSMs in acetylated-BSA. Vertical lines denote median ion abundance ratios 0.2 and 2.4 for non-acetyl and acetyl PSMs, respectively. B) Violin plots of the [126]/[129] ion abundance ratio for all PSMs in the Jurkat E6-1 cell dataset of Svinkina et al.[33] Vertical lines denote the median ion abundance ratios of 0.4 and 1.7 for non-acetyl and acetyl PSMs, respectively.

3.2. Using the 126:129 Ion Abundance Ratio as an Acetyl-Lysine Indicator

Given the ubiquity of the 126 ion in MS^2 spectra, a more specific diagnostic ion metric is needed to distinguish between acetylated and non-acetylated PSMs. Other low mass ions that may indicate the presence of an unmodified lysine include the m/z 101 (101.1079) immonium ion and a diagnostic ion at m/z 129 (129.1023).[26] The m/z 101 ion was rarely observed within lysine-containing PSMs and, when present, did not clearly differentiate acetylated from non-acetylated peptides (Table S1, Supporting Information). The m/z 129 diagnostic ion (hereafter called the "129 ion"), was present more often and proved to better indicate the presence of unmodified lysine (**Figure 1**). Similar to the low specificity of the 126 ion for acetyl-lysine containing peptides, the 129 ion was not very specific for unmodified lysine-containing peptides; 89.8% and 90.7% of non-acetyl and acetyl PSMs, respectively, contained the 129 ion (Table 1). The prevalence of the 129 ion in acetyl and non-acetyl PSMs was also noted by Svinkina, et al.[33]

Given the presence of 126 and 129 ions in both acetylated and non-acetylated PSMs, we considered whether their abundance ratio could yield an improved diagnostic for acetylation. The abundance of the 129 ion was compared to that of the 126 ion. Only 13.6% of non-acetylated PSMs had a 126 ion of greater abundance than the 129 ion, while 72.8% of acetylated PSMs contained 126 ions at abundances exceeding the 129 ions. Peak intensity ratios for acetylated and non-acetylated peptide spectra clearly differ, with acetyl PSMs having a [126]/[129] ratio greater than 1 (Figure 1A). In a large majority of cases where the [126]/[129] ratio is less than 1, the PSMs were identified as possessing an additional unmodified lysine residue (94%). Employing the [126]/[129] ratio to discriminate between non-acetylated and acetylated PSMs greatly increases the specificity as compared to using the 126 ion alone. Similar trends can be ascertained from the data published by Svinkina, et al.,[33] with only 17.8% of non-acetylated PSMs and 71.8% of acetylated PSMs displaying the 126 ion at greater abundance than 129 (Figure 1B). This data suggests that the trends identified here are generalizable to large datasets and can increase the reliability for assigning spectra with lysine acetylation.

For acetylated peptide PSMs, the 126 ion tended to be more abundant than other ions in the MS^2 spectrum. The ion abundances within each spectrum containing a 126 ion were ranked; ~80% of acetyl PSMs had m/z 126 as one of the 10 most abundant ions, as compared to 16% of non-acetyl PSMs (**Figure 2**A). The trend is also true in the re-mined Jurkat dataset (Figure 2B).[33]

Butyrylated BSA was also analyzed in order to determine if a similar trend applied to other short-chain acylations. Data was limited, likely due to the low efficiency of butyrylation in aqueous solution. Nevertheless, the trends mirrored those for acetyl BSA. The presence of the m/z 126 analogue for butyryl-lysine (m/z 154.1232) in a spectrum corresponded to a

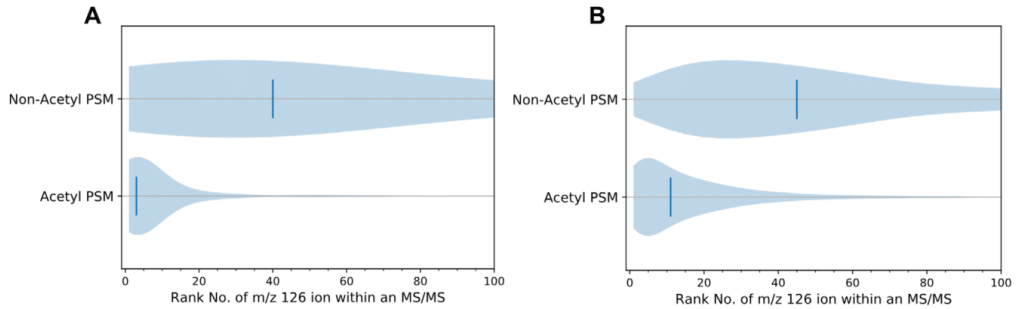


Figure 2. Abundance rank *m/z* 126 in MS/MS spectra of non-acetylated and acetylated peptides. A) Violin plots ranking the 126 ion abundance to all other ions in a spectrum for all acetyl-BSA PSMs. Vertical lines denote median rankings 40 and 3 for non-acetyl and acetyl peptides, respectively. B) Violin plots ranking the 126 ion abundance to all other ions present in a spectrum for all Jurkat E6-1 PSMs (dataset of Svinkina et al.[33]). Vertical lines denote median rankings 45 and 11 for non-acetylated and acetylated PSMs, respectively.

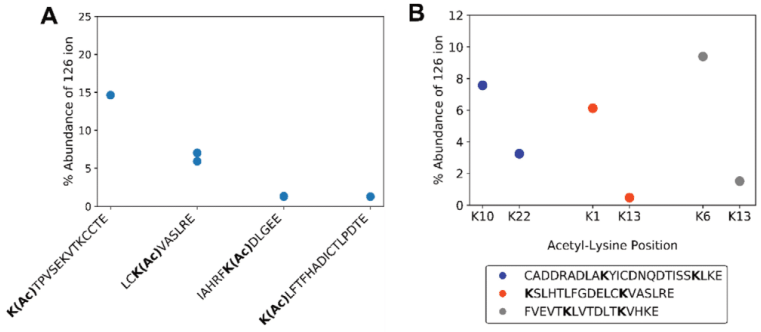


Figure 3. Dependence of 126 ion abundance on sequence context within fragmentation spectra. A) 126 ion abundance versus total product ion abundance for acetyl-BSA peptides. B) 126 ion abundance versus total product ion abundance for isobaric acetyl-BSA peptides. A single PSM of the same precursor charge state was selected for each acetyl-lysine position.

butyrylated PSM 87.1% of the time (Table S2 and Figure S2, Supporting Information), while 96.4% of butyrylated peptide spectra presented a 154 ion, indicating that it is both specific and sensitive.

3.3. Sequence Dependence of Immonium Ion Formation

Efficient formation of immonium ions can depend on a modified residue's location in a particular peptide sequence. Sequence-specific fragmentation is a well elucidated phenomenon, as certain local residues can affect CID fragmentation; for example, when a proline or glycine residue is near.[34–38] Some examples of this sequence dependence are shown in **Figure 3**. Generally an acetyl-lysine at the N-terminal position will yield the strongest signal for the 126 ion (Figure 3A).[24] This observation can be rationalized by recognizing that the first step in forming immonium ions requires an *N*-terminal amine. Hence, an internal lysine would require two fragmentation events, while an *N*-terminal lysine requires only one. Sequence composition is also important. For example, two different peptides with acetylation at the K1 position yield very different relative abundances for the 126 ions (Figure 3A). Peptides containing multiple lysines that are acetylated closer to the *N*-terminus tend to yield more abundant 126 ions (Figure 3B). Thus, there can be a localization bias when using the cyclized immonium ion to identify and/or validate acyl modifications.

Normalized collision energy (NCE) was increased in an attempt to maximize the 126 ion's signal. Global analysis with this optimization indicates, indeed, that the elevated NCE increases the relative abundance of the 126 ion from the acetylated BSA sample (**Figure 4**A). As per the violin plots, the 126 ion relative abundance increases with increasing NCE. Tandem mass spectra for individual peptides were examined to ensure that this trend applied to modifications at varied sequence locations. Figure 4B–D shows examples for different modified sequences. All of these peptides show that increased collisional energy increases the relative abundance of the immonium ion, congruent with the global analysis of Figure 4A.

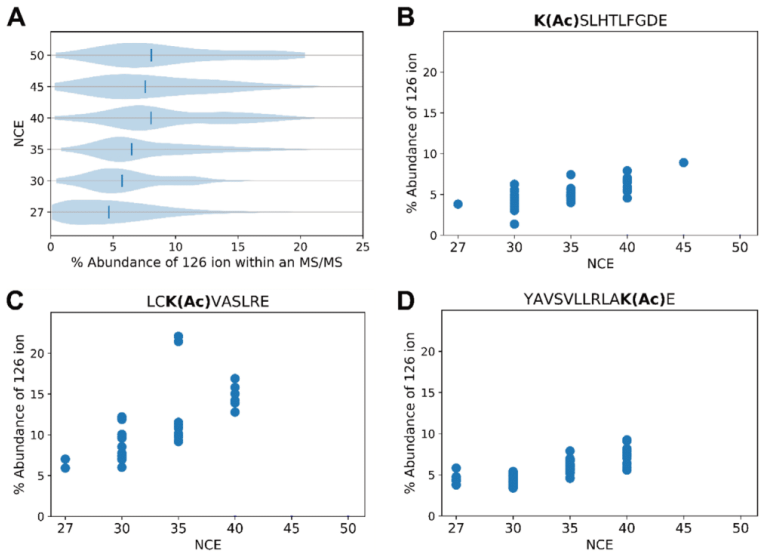


Figure 4. Abundance of the 126 ion versus total product ion abundance in MS^2 spectra of acetyl PSMs. A) Each violin plot shows data obtained at a given NCE that reveals for each acetyl-PSM the percentage of total ion signal due to m/z 126. Vertical lines indicate the median relative abundances of the 126 ion. B–D) Percentage of total signal due to the 126 ion versus NCE for acetylated peptides selected for their acetyl-lysine position and sequence context. Some peptides were not identified at higher NCEs.

3.4. Stepped Collisional Activation Mitigates a Loss of Sequence Identifications

Increasing collision energy increases the intensities of low mass peaks, such as immonium ions, but in doing so, information critical to identifying the peptide sequence may be lost. At an NCE of 27, typical for most proteomic CID fragmentation, over 200 unique peptides could be identified from acetylated-BSA LC-MS/MS runs. At 30 NCE, the number of identified peptides dropped moderately to just below 100. However, at NCE 35 and 40, the number falls dramatically to just over 20 (**Figure 5**). While higher collision energies highlight the 126 ion in MS^2 spectra, the benefit comes at the cost of poor sequence information, making it less likely that peptides can be matched to the correct sequence confidently.

To recover this information, a stepped collision energy (CE) approach was taken to enhance the intensity of the immonium ion peaks while retaining enough high mass information to identify the peptide sequence. In the stepped CE strategy, precursor ions are fragmented at low and high NCE, and the product ions from each CE are pooled for detection. Figure 5 shows how stepped CE recovers the sequence information lost by using excessive collision energy. Several stepped collision energy combinations (NCEs of 27/40, 27/45, and 27/50) were tested to optimally balance maximal signal for the 126 ion with retention of sequence-related information. A stepped NCE method of 27/40 was found to be optimum, recovering 80% of unique peptides relative to the single energy method at 27 NCE (Figure 5). There was still some loss of peptides identified due to the complexity of

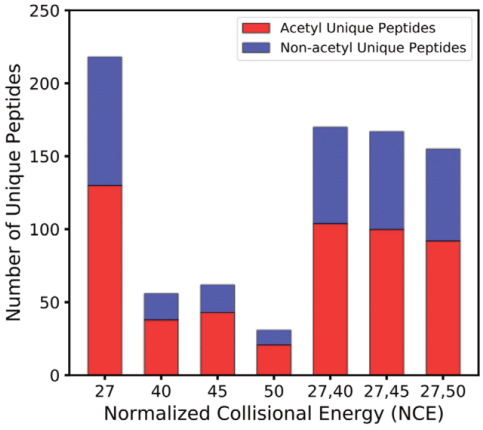


Figure 5. Unique peptides identified with different values of fixed and stepped NCE. Acetylated (red) and non-acetylated (blue) peptides.

the spectra produced by the stepped method, as the higher collisional energy results in greater neutral loss and low mass product ions. Applying this approach, we rescued the number of peptides identified while maintaining a strong immonium ion, such that 95.4% of acetylated PSMs presented a 126 ion within the 10 most abundant ions (compared to 80% from 27 NCE). Butyrylation

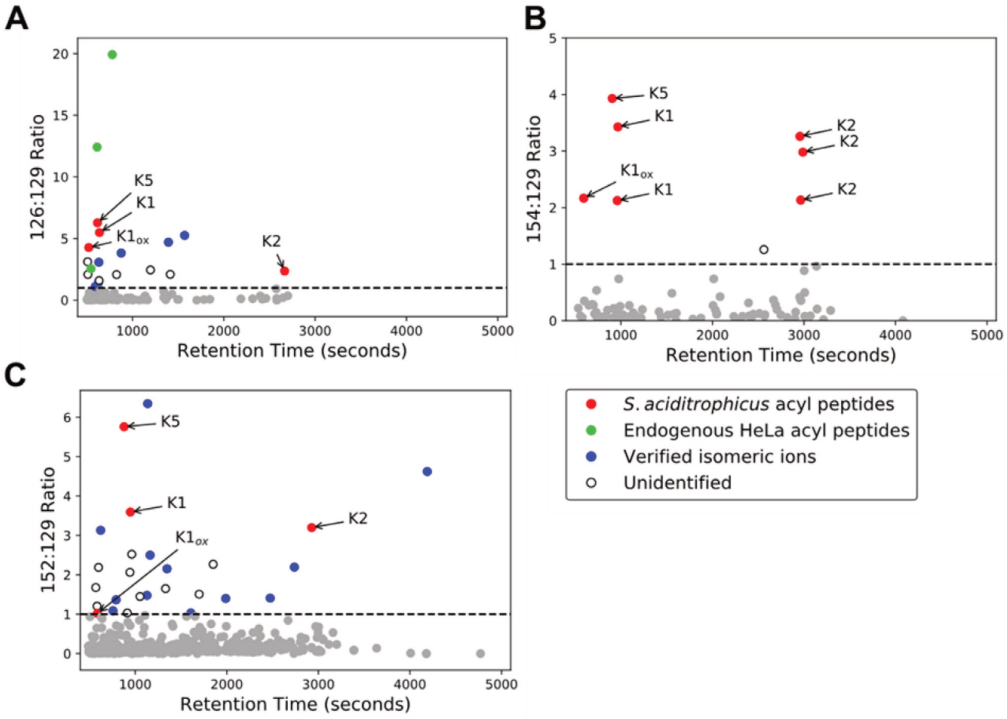


Figure 6. Immonium ion abundance ratio reveals low abundance synthetic acyl-peptides in a HeLa lysate background. Scatterplots of the immonium ion abundance ratio versus retention time for A) lysine acetylation ([126]/[129]), B) lysine butyrylation ([154]/[129]), and C) lysine crotonylation ([152]/[129]), respectively. Grey dots represent spectra with an abundance ratio less than one. K1, K2, and K5 correspond to the synthetic *S. aciditrophicus* sequences **K**STPEAMAK, F**K**DEIPVVIK, and STDP**K**GPSVR, respectively, where the bolded residues are acylated. Verified isomeric ions and unidentified ions are described in the text.

(m/z 154.1232) showed similar trends with respect to stepped NCE (Figure S3, Supporting Information).

3.5. Immonium Ions Identify Acyl Modifications in a Complex Lysate Using a Stepped Collision Energy

A question that remained with the stepped collision energy method was whether it would readily reveal immonium ions for acetyl-lysine and other acyl-lysines (e.g., butyryl- and crotonyl-lysine [m/z 152.107]) in a complex biological sample. To investigate this question, synthetic acylated peptides were spiked into a trypsinized HeLa lysate from which the respective immonium ion: m/z 129 abundance ratios were monitored. The spiked peptides (**K**STPEAMAK, F**K**DEIPVVIK, STDP**K**GPSVR) differ in the type and location of acylation. In addition to acetyl-lysine, we chose butyryl- and crotonyl-lysine because they are relevant to *S. aciditrophicus*, a syntrophic bacterium for which we expect extensive protein acylation due to high cellular concentrations acyl-CoA metabolites.[17] Our lab identified these sequences as highly modified in the course of *S. aciditrophicus* proteomic studies.[39] Plotted are the cyclized immonium: 129 ion relative abundance ratios for each MS^2 spectrum (**Figure 6**). Acylated peptides are clearly identified in the population of PSMs with [immonium]/[129] ratios exceeding 1. Isomers of the 126[24] and 152 ions are also present (Figure S4, Supporting Information) and some peptides containing these sequences were identified (Figure 6). The quality of some MS^2 spectra with abundance ratios exceeding 1 was insufficient to ascribe a sequence. These instances are labeled as unidentified in Figure 6.

3.6. Using Immonium Ions to Identify the Presence of Novel PTMs

S. aciditrophicus has unique metabolic pathways that allow for the formation of novel protein PTMs. Given that its short chain and aromatic fatty acid metabolism generates a variety of reactive acyl-CoA metabolites, we expect the *S. aciditrophicus* proteome to display unique acyl-lysine modifications. These acyl-CoA intermediates can spontaneously acylate lysine side chain amines under physiological conditions.[40] One acyl-CoA intermediate

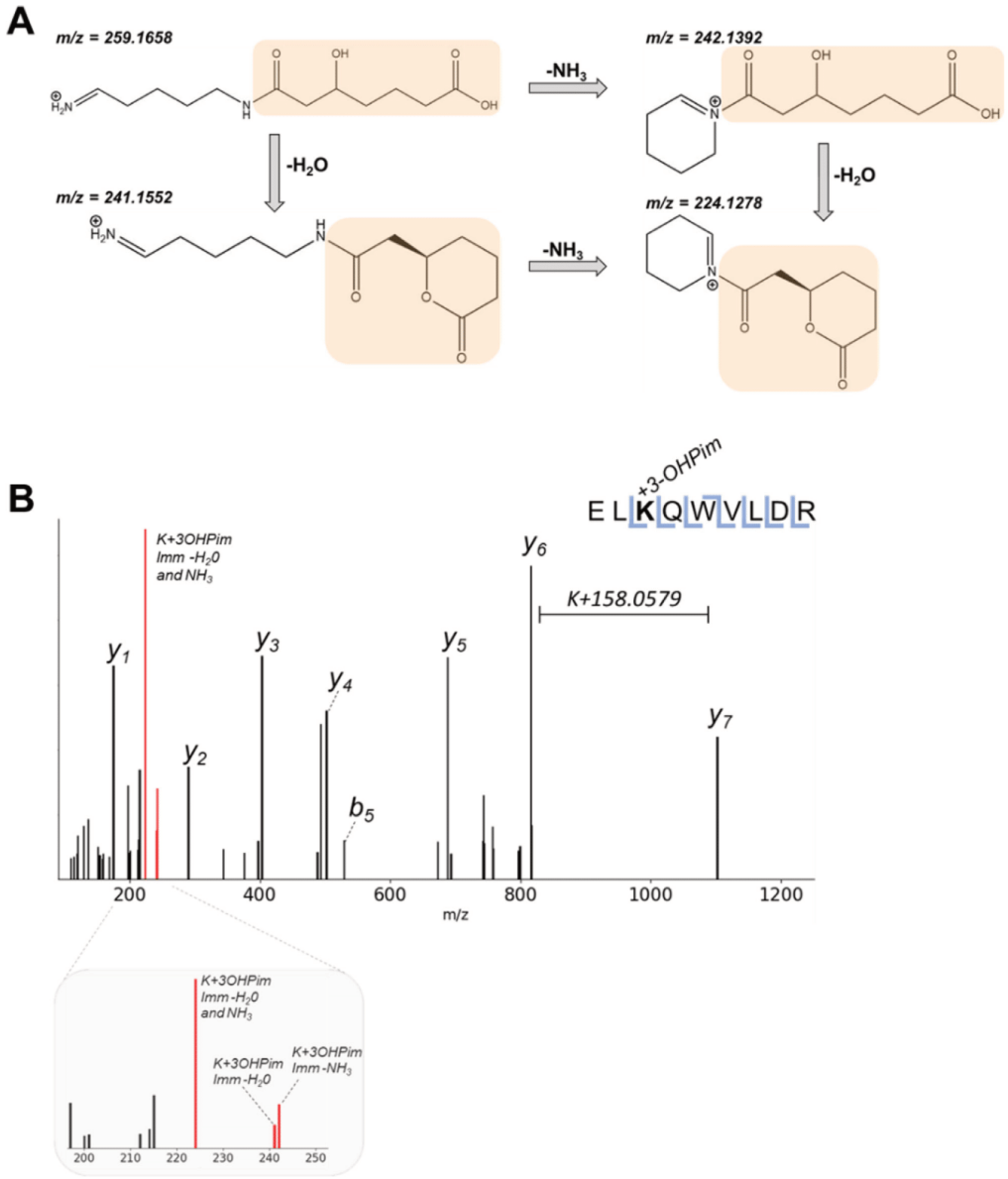


Figure 7. Novel modification (3-hydroxypimelylation) on benzoate-CoA ligase in *S. aciditrophicus* with its diagnostic ion. A) Immonium ion for 3-hydroxypimelylation (K+158.0579 Da) (highlighted in orange) undergoes facile neutral losses. B) MS/MS spectrum showing mass shift corresponding to 3-hydroxypimelylation and its low mass region enlarged to highlight immonium-related ions (colored in red).

within the benzoate degradation pathway is 3-hydroxypimelyl-CoA.[41] Interestingly, we noted that some mass spectra from the benzoate-cultivated *S. aciditrophicus* proteome presented mass shifts of K+158.0579 Da. To determine the validity of this novel modification, immonium-related ions were examined to verify that the shift was associated with a novel lysine PTM on a benzoate-CoA ligase peptide, rather than a misidentification.[42] Based on the immonium ion structure, one expects certain facile neutral losses (**Figure 7**) to be present, an NH_3 neutral loss (similar to that of the acetyl-lysine 126 ion), an H_2O loss, and the loss of both NH_3 and H_2O. Tandem mass spectra from two peptides containing the putative modification, benzoate-CoA ligase (RS03815/RS03820) and acetyl-CoA transferase (RS12490), presented ions corresponding to neutral losses of NH_3, H_2O, and NH_3+H_2O from the predicted immonium ions (Figure 7, Figure S5, Supporting Information); that is, 242.1392, 241.1552, and 224.1278 Da, respectively. These immonium-related ions validated the presence of a hydroxypimelyl-lysine, the first observation of this PTM. Although only two 3-hydroxypimelyl-lysine peptides were identified in this analysis, we have found many other 3-hydroxylpimelyl modifications from other *S. aciditrophicus* sample preparations that will be addressed in future studies.

4. Discussion

Immonium and immonium-related ions are excellent proxies for the presence of acetylation.[24,43] Despite their longtime use in global analyses, current instrumentation and methods have not been optimized for their use. While previous reports found that the specificity for acetylation of the m/z 126 ion was superior to that of m/z 143,[24] the former is, nevertheless, observed in mass spectra lacking any acetylation, as is confirmed in our analysis. Hence, the diagnostic ion's presence alone is not sufficient to claim modification with high confidence, or to restrict data acquisition or analysis to only acetylated peptides. Incorporating abundance ratios provides a nuanced way to utilize this marker. The [126]/[129] ratio is meaningful when identifying candidate PSMs that contain acetyl-lysine residues. Although some sequence contexts make it an imperfect criterion, ions likely containing acetylated lysine can be identified by the [126]/[129] ratio in MS^2 spectra prior to database searching. This can filter candidate spectra prior to sequence assignment or highlight modified precursors in the case of limited sequence information. This criterion can be also used in conjunction with other evidence for more rigorous identification. Likewise, incorporating a stepped collisional energy CID experiment may increase confidence in acyl-lysine assignments. Akin to oxonium ion analysis, using a stepped NCE strategy highlights the diagnostic ions of acyl-lysine while also optimizing sequence-related information in a data-dependent acquisition (DDA) method. Stepped NCE is particularly important for modifications present in the middle of a peptide, because generating these immonium ions require more collision energy.

While using immonium ions is appropriate for CID, transferring that use to other activation methods requires consideration of the mechanisms by which those methods fragment peptides. Another common fragmentation approach in the analysis of peptides and PTMs is electron transfer dissociation (ETD),[44–46] and interest in ultraviolet (UV) photodissociation (PD) is growing as well.[47] CID fragmentation yields many products, including the neutral losses that form immonium ions. ETD mostly cleaves along the Cα-N bond of the peptide backbone,[44] thereby making immonium ion formation unlikely. However, the production of immonium ions by ETD and PD has not been widely explored to date and requires further examination.

Future applications of stepped NCE with immonium ion abundance analyses can take two possible routes. One is to expose and verify novel PTMs. As the list of putative acyl modifications grows,[4,48,49] it becomes critical to obtain additional constraints (metrics) to validate the presence of increasingly complex modifications. Many acyl-modified peptides can be isobaric with di- or tri-peptide related ions or be indistinguishable from non-acylated ions on low resolution instruments.[42] For example, acetyl and propionyl modifications differ by a single methyl group, as do propionyl and butyryl modifications. The presence of a methylated residue or an amino acid that differs from another by a methyl group (Ser/Thr, Asn/Gln, Ala/Val, etc.) may make it impossible to distinguish the correct acyl modification.[50] Similarly, carbamylation has a 43.00582 Da mass shift, whereas acetylation combined with deamidation produce a 42.99458 Da shift. Depending on the mass of the ion and the resolving power of the instrument, these PTMs could be indistinguishable in the event of poor fragmentation and unincorporated diagnostic ions.

Stepped NCE can also be employed in conjunction with data-independent acquisition (DIA) approaches. Currently, DIA is greatly limited by the complexity of spectral deconvolution, particularly by fragment ion interference when co-eluting peptides have very similar masses.[51] Diagnostic ions, however, can increase the confidence that an acyl-lysine precursor is present in the convoluted spectra; thereby providing more information for elegant algorithms to consider when identifying and assigning PTMs. Furthermore, experiments targeting acyl modifications can exploit immonium and immonium-related ions in parallel reaction monitoring (PRM) or product ion scanning.

Increased understanding of acyl modifications has suggested some biological functions for these modifications.[52] Given the reactive nature of acyl-CoA species[17] and the new depths proteomics can reach, it should be expected that more acyl modifications will be found and may show biological significance in metabolic pathways across organisms. As datasets grow in size and complexity, confident assignments are essential. Properly incorporating immonium ions into the assignment of acyl-modifications will increase confidence in established PTM identifications and support/validate assignment of those yet to be discovered.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

J.M.M. and J.Y.F. contributed equally to this work. Funding from the Department of Energy Office of Science (BER) contract DE-FC-02-02ER63421 (to J.A.L.; UCLA/DOE Institute for Genomics and Proteomics), NIH Ruth L. Kirschstein National Research Service Award GM007185, and NSF Graduate Research Fellowship (to J.Y.F.; DGE-1650604) is acknowledged. J.M.M. was supported by a UCLA Molecular Biology Institute Whitcome

Fellowship. The authors would also like to thank Dr. Michael J. McInerny (University of Oklahoma) and Dr. Robert P. Gunsalus (UCLA) for the *S. aciditrophicus* samples used in the study.

Conflict of Interest

The authors declare no conflict of interest.

Keywords

acylation, immonium ions

Received: April 28, 2020
Revised: August 3, 2020
Published online:

[1] G. Walsh, R. Jefferis, *Nat. Biotechnol.* **2006**, *24*, 1241.
[2] K. W. Moremen, M. Tiemeyer, A. V. Nairn, *Nat. Rev. Mol. Cell Biol.* **2012**, *13*, 448.
[3] E. Guccione, S. Richard, *Nat. Rev. Mol. Cell Biol.* **2019**, *20*, 642.
[4] B. R. Sabari, D. Zhang, C. D. Allis, Y. Zhao, *Nat. Rev. Mol. Cell Biol.* **2017**, *18*, 90.
[5] S. J. Humphrey, D. E. James, M. Mann, *Tr. Endocrinol. Metab.* **2015**, *26*, 676.
[6] C. M. VanDrisse, J. C. Escalante-Semerena, *Annu. Rev. Microbiol.* **2019**, *73*, 111.
[7] M. C. N. Sack, T. Finkel, *Cold Spring Harb. Perspect. Biol.* **2012**, *4*, a013102.
[8] C. Carrico, J. G. Meyer, W. He, B. W. Gibson, E. Verdin, *Cell Metab.* **2018**, *27*, 497.
[9] E. Verdin, M. Ott, *Nat. Rev. Mol. Cell Biol.* **2015**, *16*, 258.
[10] B. M. Turner, *BioEssays* **2000**, *22*, 836.
[11] B. M. Turner, *Cell. Mol. Life Sci.* **1998**, *54*, 21.
[12] H. A. Crosby, E. K. Heiniger, C. S. Harwood, J. C. Escalante-Semerena, *Mol. Microbiol.* **2010**, *76*, 874.
[13] J. G. Gardner, F. J. Grundy, T. M. Henkin, J. C. Escalante-Semerena, *J. Bacteriol.* **2006**, *188*, 5460.
[14] M. Ott, M. Schnölzer, J. Garnica, W. Fischle, S. Emiliani, H. R. Rackwitz, E. Verdin, *Curr. Biol.* **1999**, *9*, 1489.
[15] B. K. Hansen, R. Gupta, L. Baldus, D. Lyon, T. Narita, M. Lammers, C. Choudhary, B. T. Weinert, *Nat. Commun.* **2019**, *10*, 1055.
[16] Y. Li, J. C. Silva, M. E. Skinner, D. B. Lombard, *Methods Mol. Biol.* **2013**, *1077*, 81.
[17] A. G. Trub, M. D. Hirschey, *Trends Biochem. Sci.* **2018**, *43*, 369.
[18] P. J. Boersema, S. Mohammed, A. J. R. Heck, *J. Mass Spectrom.* **2009**, *44*, 861.
[19] S. M. M. Sweet, C. M. Bailey, D. L. Cunningham, J. K. Heath, H. J. Cooper, *Mol. Cell. Proteomics* **2009**, *8*, 904.
[20] Q. Cao, X. Zhao, Q. Zhao, X. Lv, C. Ma, X. Li, Y. Zhao, B. Peng, W. Ying, X. Qian, *Anal. Chem.* **2014**, *86*, 6804.
[21] A. Halim, U. Westerlind, C. Pett, M. Schorlemer, U. Rüetschi, G. Brinkmalm, C. Sihlbom, J. Lengqvist, G. Larson, J. Nilsson, *J. Proteome Res.* **2014**, *13*, 6024.
[22] S. A. Carr, M. J. Huddleston, M. F. Bean, *Protein Sci.* **1993**, *2*, 183.
[23] J. Zaia, *Mass Spectrom. Rev.* **2004**, *23*, 161.
[24] M. B. Trelle, O. N. Jensen, *Anal. Chem.* **2008**, *80*, 3422.
[25] D. P. Zolg, M. Wilhelm, T. Schmidt, G. Médard, J. Zerweck, T. Knaute, H. Wenschuh, U. Reimer, K. Schnatbaum, B. Kuster, *Mol. Cell. Proteomics* **2018**, *17*, 1850.
[26] A. M. Falick, W. M. Hines, K. F. Medzihradszky, M. A. Baldwin, B. W. Gibson, *J. Am. Soc. Mass Spectrom.* **1993**, *4*, 882.
[27] C. W. Hung, A. Schlosser, J. Wei, W. D. Lehmann, *Anal. Bioanal. Chem.* **2007**, *389*, 1003.
[28] T. Yalcin, A. G. Harrison, *J. Mass Spectrom.* **1996**, *31*, 1237.
[29] J. Baeza, M. J. Smallegan, J. M. Denu, *ACS Chem. Biol.* **2015**, *10*, 122.
[30] J. Rappsilber, M. Mann, Y. Ishihama, *Nat. Protoc.* **2007**, *2*, 1896.
[31] J. Erde, R. R. Ogorzalek Loo, J. A. Loo, *Methods Mol. Biol.* **2017**, *1550*, 11.
[32] J. Erde, R. R. Ogorzalek Loo, J. A. Loo, *J. Proteome Res.* **2014**, *13*, 1885.
[33] T. Svinkina, H. Gu, J. C. Silva, P. Mertins, J. Qiao, S. Fereshetian, J. D. Jaffe, E. Kuhn, N. D. Udeshi, S. A. Carr, *Mol. Cell. Proteomics* **2015**, *14*, 2429.
[34] W. D. van Dongen, H. F. M. Ruijters, H.-J. Luinge, W. Heerma, J. Haverkamp, *J. Mass Spectrom.* **1996**, *31*, 1156.
[35] E. A. Kapp, F. Schütz, G. E. Reid, J. S. Eddes, R. L. Moritz, R. A. J. O'Hair, T. P. Speed, R. J. Simpson, *Anal. Chem.* **2003**, *75*, 6251.
[36] J. A. Loo, C. G. Edmonds, R. D. Smith, *Anal. Chem.* **1993**, *65*, 425.
[37] L. A. Breci, D. L. Tabb, J. R. Yates, V. H. Wysocki, *Anal. Chem.* **2003**, *75*, 1963.
[38] Y. Huang, J. M. Triscari, G. C. Tseng, L. Pasa-Tolic, M. S. Lipton, R. D. Smith, V. H. Wysocki, *Anal. Chem.* **2005**, *77*, 5800.
[39] H. Nguyen, M. McInerney, R. Gunsalus, J. A. Loo, R. R. O. Loo, presented at 64th ASMS Conf. Mass Spectrom. Allied Top, San Antonio, TX, June 2016.
[40] G. R. Wagner, R. M. Payne, *J. Biol. Chem.* **2013**, *288*, 29036.
[41] M. S. Elshahed, V. K. Bhupathiraju, N. Q. Wofford, M. A. Nanny, M. J. McInerney, *Appl. Environ. Microbiol.* **2001**, *67*, 1728.
[42] M. S. Kim, J. Zhong, A. Pandey, *Proteomics* **2016**, *16*, 700.
[43] E. S. Nakayasu, S. Wu, M. A. Sydor, A. K. Shukla, K. K. Weitz, R. J. Moore, K. K. Hixson, J.-S. Kim, V. A. Petyuk, M. E. Monroe, L. Pasa-Tolic, W.-J. Qian, R. D. Smith, J. N. Adkins, C. Ansong, *Int. J. Proteomics* **2014**, *2014*, 730725.
[44] J. E. P. Syka, J. J. Coon, M. J. Schroeder, J. Shabanowitz, D. F. Hunt, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 9528.
[45] L. M. Mikesh, B. Ueberheide, A. Chi, J. J. Coon, J. E. P. Syka, J. Shabanowitz, D. F. Hunt, *Biochim. Biophys. Acta - Proteins Proteomics* **2006**, *1764*, 1811.
[46] A. Chi, C. Huttenhower, L. Y. Geer, J. J. Coon, J. E. P. Syka, D. L. Bai, J. Shabanowitz, D. J. Burke, O. G. Troyanskaya, D. F. Hunt, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 2193.
[47] M. A. Halim, L. MacAleese, J. Lemoine, R. Antoine, P. Dugourd, M. Girod, *J. Am. Soc. Mass Spectrom.* **2018**, *29*, 270.
[48] D. Zhang, Z. Tang, H. Huang, G. Zhou, C. Cui, Y. Weng, W. Liu, S. Kim, S. Lee, M. Perez-Neut, J. Ding, D. Czyz, R. Hu, Z. Ye, M. He, Y. G. Zheng, H. A. Shuman, L. Dai, B. Ren, R. G. Roeder, L. Becker, Y. Zhao, *Nature* **2019**, *574*, 575.
[49] H. Huang, D. Zhang, Y. Wang, M. Perez-Neut, Z. Han, Y. G. Zheng, Q. Hao, Y. Zhao, *Nat. Commun.* **2018**, *9*, 3374.
[50] S. Lee, M. Tan, L. Dai, O. K. Kwon, J. S. Yang, Y. Zhao, Y. Chen, *J. Proteome Res.* **2013**, *12*, 1007.
[51] L. C. Gillet, P. Navarro, S. Tate, H. Röst, N. Selevsek, L. Reiter, R. Bonner, R. Aebersold, *Mol. Cell. Proteomics* **2012**, *11*, O111.016717.
[52] Y. Xiong, K. L. Guan, *J. Cell Biol.* **2012**, *198*, 155.

Supplemental Tables and Figures

Leveraging immonium ions for targeting acyl-lysine modifications in proteomic datasets

John M. Muroski[1*], Janine Y. Fu[1*], Hong Hanh Nguyen[1#], Rachel R. Ogorzalek Loo[2,3,4†], Joseph A. Loo[1,2,3,4†]

1 Department of Chemistry and Biochemistry, University of California, Los Angeles, CA, USA

2 David Geffen School of Medicine, Department of Biological Chemistry, University of California, Los Angeles, CA, USA

3 UCLA-DOE Institute, University of California, Los Angeles, CA, USA

4 UCLA Molecular Biology Institute, University of California, Los Angeles, CA, USA

* Equal contribution

Current address: TRANSMED Co. Ltd., Ho Chi Minh City, Vietnam

[†]Corresponding authors (RLoo@mednet.ucla.edu, JLoo@chem.ucla.edu)

Table S1. Distribution of PSMs containing the *m/z* 101 immonium ion from acetylated BSA (27 NCE) and in Jurkat E6-1 cells (25 NCE) (Svinkina *et al.*)[33]

		PSMs with 101 ion		PSMs without 101 ion	
Acetyl-BSA	Non-acetyl	59	(25%)	177	(75%)
	Acetyl	58	(17.3%)	277	(82.7%)
Svinkina et al.	Non-acetyl	201	(3.3%)	5834	(96.7%)
	Acetyl	151	(3.3%)	4449	(96.7%)

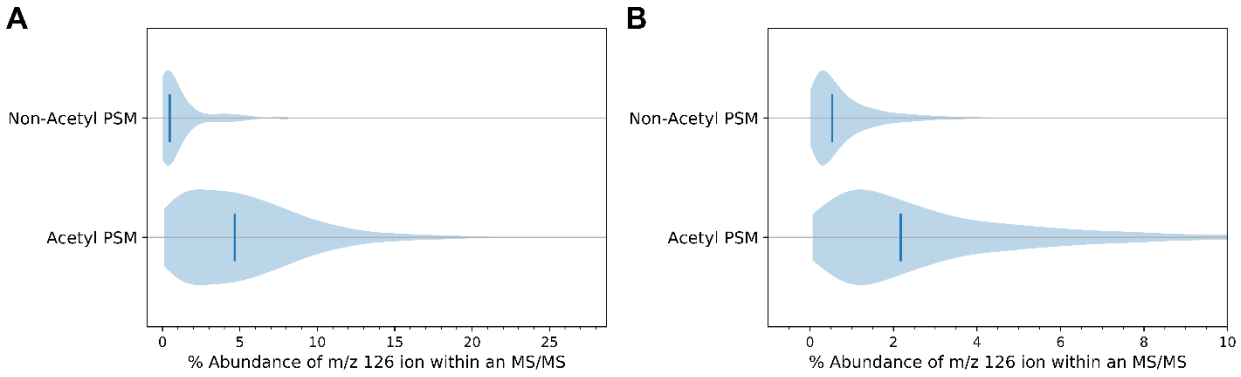


Figure S1. Percentage of total MS^2 ion signal for *m/z* 126. A) Violin plots of the relative signal from *m/z* 126 for acetylated and non-acetylated PSMs of acetyl-BSA. Vertical lines denote the 0.5% and 4.7% median percent intensity from non-acetyl and acetyl PSMs, respectively. B) Violin plots of the relative signal from *m/z* 126 for all PSMs in the Jurkat E6-1 cell dataset of Svinkina *et al.*[33] Vertical lines denote the 0.5% and 2.2% medians for non-acetyl and acetyl PSMs, respectively.

Table S2. Distribution of PSMs containing 154 and 129 immonium ions in peptides from butyrylated BSA (27 NCE).

		PSMs with 154 ion		PSMs without 154 ion		PSMs with 129 ion		PSMs without 129 ion	
Butyryl-BSA	Non-butyryl	4	(18.2%)	18	(81.8%)	15	(68.2%)	7	(31.8%)
	Butyryl	27	(96.4%)	1	(3.7%)	16	(57.1%)	12	(42.9%)

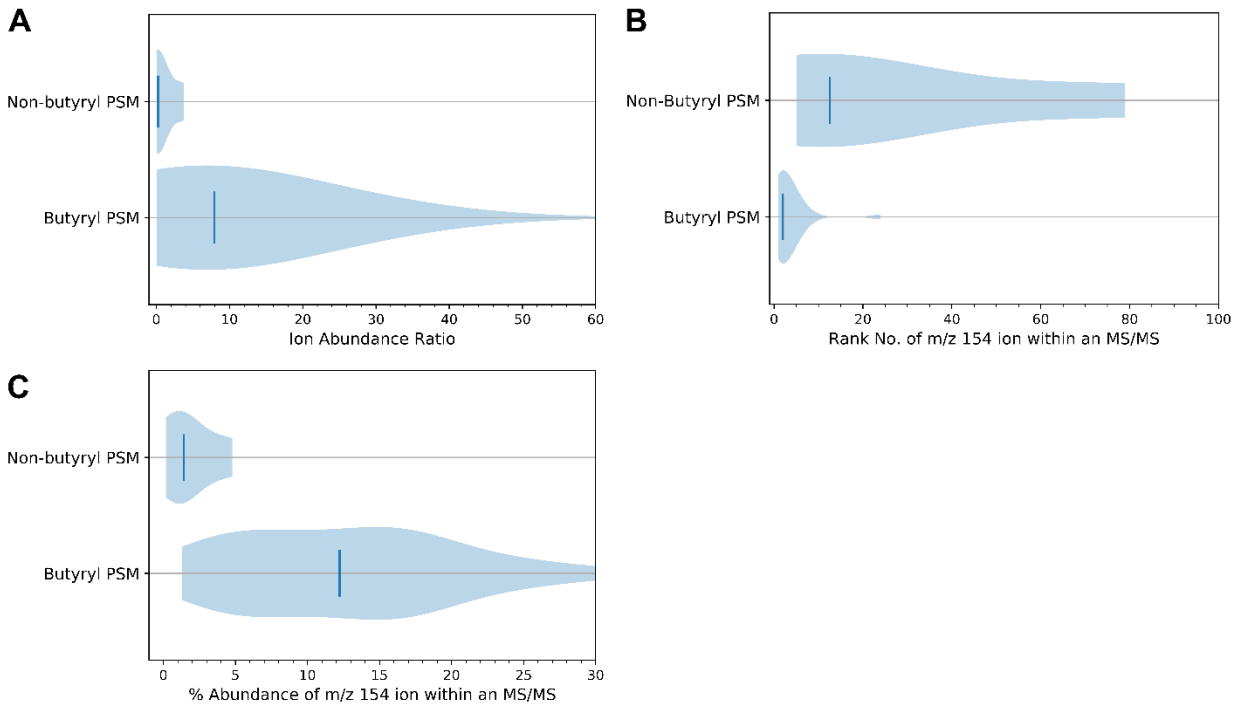


Figure S2. The 154 ion and [154]/[129] abundance ratios as markers for butyryl-lysine at 27 NCE. A) Violin plots of the [154]/[129] ion abundance ratio for all butyryl-BSA PSMs. Vertical lines denote the 0.3 and and 8.0 median ion abundance ratios for non-butyryl and butyryl PSMs, respectively. B) The abundance rank of *m/z* 154 relative to other product ions in each spectrum. Vertical lines denote the median ranks, 13 and 2, for non-butyryl and butyryl PSMs, respectively. C) The percentage of total product ion signal from *m/z* 154. The vertical lines denote the median percent abundance, 1.4% and 12.2%, for non-butyryl and butyryl PSMs, respectively.

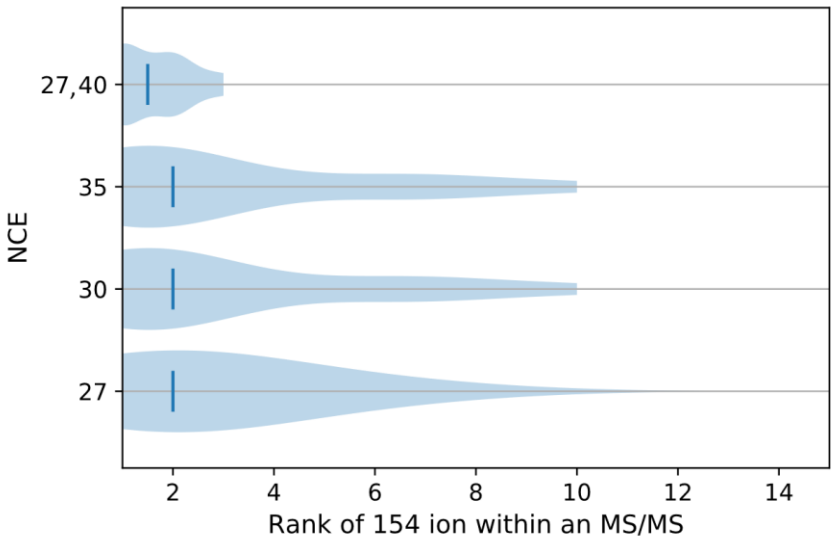


Figure S3. Abundance rank of the 154 ion depends on NCE setting. Violin plots of the 154 ion's abundance rank in a spectrum for butyryl PSMs. Vertical lines denote the median rankings: 2 for fixed NCE= 27, 30, and 35 and 1 for stepped NCE= 27,40. No butyryl PSMs were identified at 40, 45, and 50 NCE.

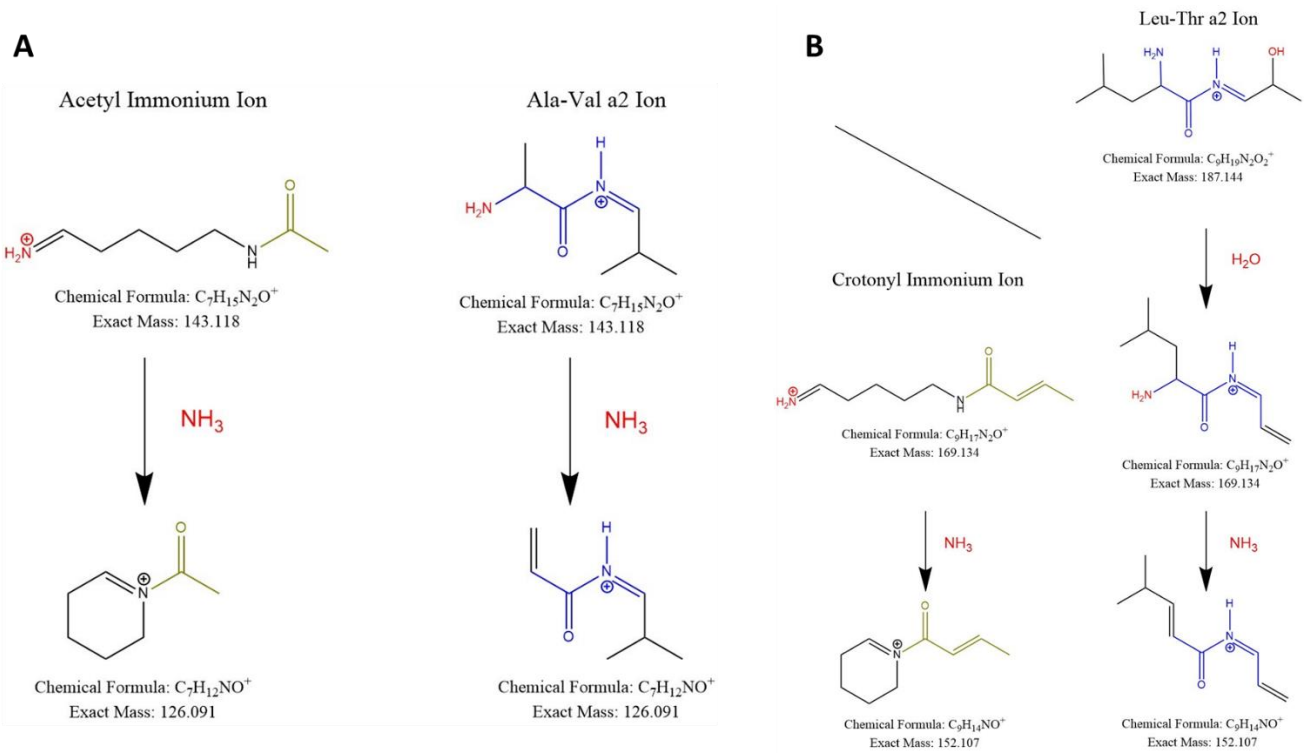


Figure S4. Verified isomers of acyl-lysine marker ions. Proposed ions that are isobaric to the acyl-lysine immonium ions depicted in Figure 6. A) Pathway producing Ala-Val products isobaric with the acetyl-lysine immonium ion. Not shown are Xle-Gly products that are also isobaric. B) Pathway producing Xle-Thr products isobaric with the crotonyl-lysine immonium ion.

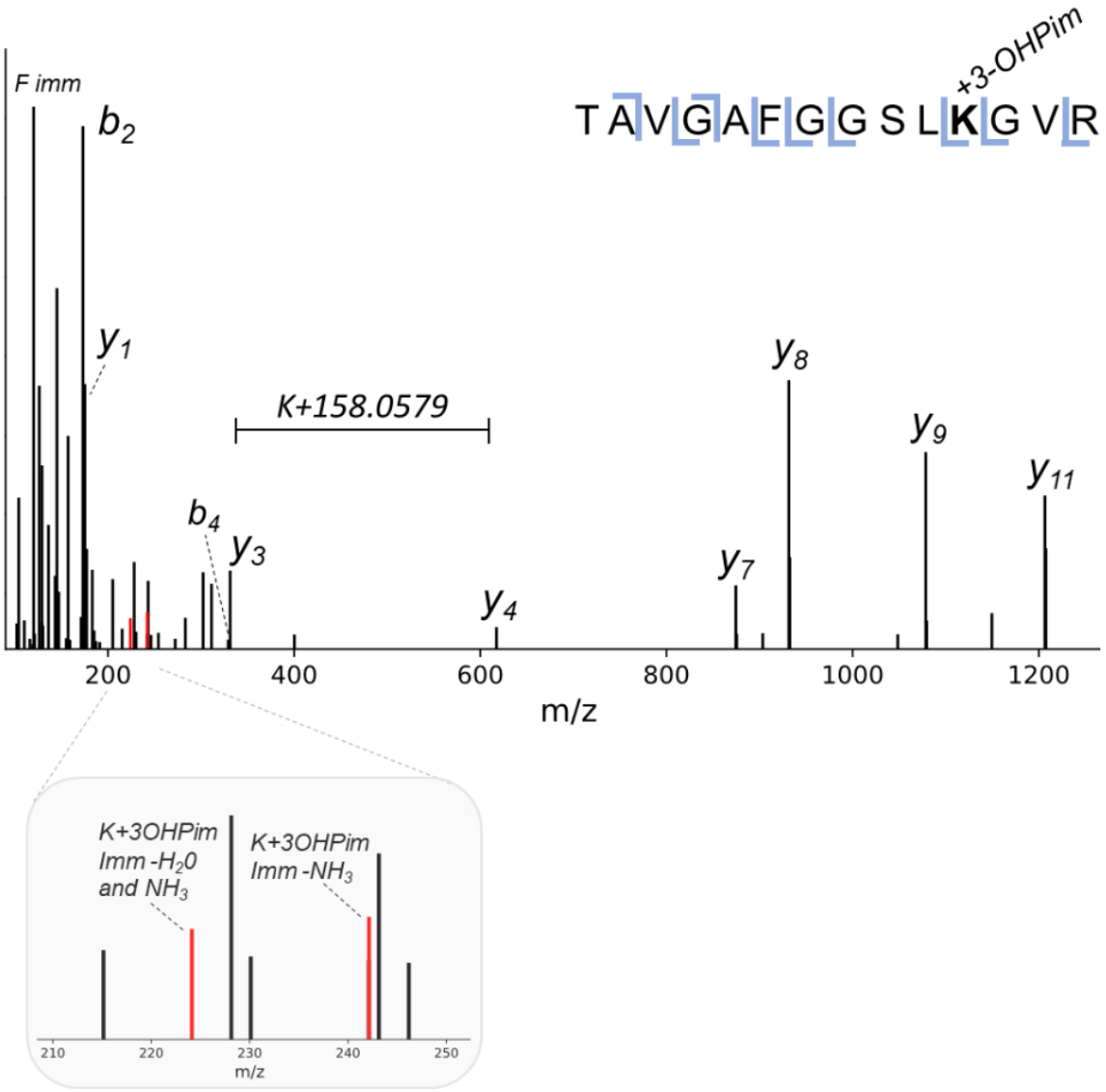


Figure S5. Novel modification (3-hydroxypimelylation) on acetyl-CoA acetyltransferase in *S. aciditrophicus* with its diagnostic ion. MS/MS spectrum showing mass shift corresponding to 3-hydroxypimelylation and its low mass region enlarged to highlight immonium-related ions (colored in red).

CHAPTER 3

The acyl-proteome of *Syntrophus aciditrophicus* reveals metabolic relationship with benzoate degradation

The following is a manuscript submitted for publication.

The acyl-proteome of Syntrophus aciditrophicus reveals metabolic relationship with benzoate degradation

John M. Muroski[1], Janine Y. Fu[1], Hong Hanh Nguyen[1], Neil Q. Wofford[2], Housna Mouttaki[2], Kimberly L. James[2], Michael J. McInerney[2], Robert P. Gunsalus[3,5,6], Joseph A. Loo[1,4,5,6], Rachel R. Ogorzalek Loo[1,5,6]

1 Department of Chemistry and Biochemistry, University of California, Los Angeles, CA, USA

2 Department of Microbiology and Plant Biology, University of Oklahoma, Norman, Oklahoma, USA

3 Department of Microbiology, Immunology and Molecular Genetics, University of California, Los Angeles, Los Angeles, California, USA

4 David Geffen School of Medicine, Department of Biological Chemistry, University of California, Los Angeles, CA, USA

5 UCLA-DOE Institute, University of California, Los Angeles, CA, USA

6 UCLA Molecular Biology Institute, University of California, Los Angeles, CA, USA

Running Title

The acyl-proteome of Syntrophus aciditrophicus

Abbreviations

AGC – automatic gain control
ATP – adenosine triphosphate
BCL – benzoate-CoA ligase
BLAST- basic local alignment search tool
CoA – Coenzyme A
COG – clusters of orthologous groups
DDA – data-dependent acquisition
eFASP – enhanced filter-assisted sample preparation
ESI – electrospray ionization
$FAD/FADH_2$ – flavin adenine dinucleotide (oxidized/reduced)
GO – gene ontology
HCD – higher-energy collision dissociation
HILIC – hydrophilic interaction liquid chromatography
KEGG – Kyoto Encyclopedia of Genes and Genomes
MALDI – matrix-assisted laser desorption/ionization
NAD^+/NADH – nicotinamide adenine dinucleotide (oxidized/reduced)
2D-PAGE – two-dimensional polyacrylamide gel electrophoresis
PTM – post-translational modification
RACS – reactive acyl-Coenzyme A species
TOF – time-of-flight

Abstract

Syntrophus aciditrophicus is a model syntrophic bacterium that degrades fatty and aromatic acids into acetate, CO_2, formate and H_2 that are utilized by methanogens and other hydrogen-consuming microbes. The degradation of benzoate by *S. aciditrophicus* proceeds by a multi-step pathway that involves many reactive acyl-Coenzyme A species (RACS) as intermediates which can potentially result in N^{ε}-acylation of lysine residues in proteins. Herein, we investigate post-translational modifications in the *S. aciditrophicus* proteome to identify and characterize a variety of acyl-lysine modifications that correspond to RACS present in the benzoate degradation pathway. Modification levels are sufficient to support post-translational modification analyses without antibody enrichment, enabling the study of a range of acylations located, presumably, on the most extensively acylated proteins. Seven types of acyl modifications were identified throughout the proteome, six of which correspond directly to RACS that are intermediates in the benzoate degradation pathway. Benzoate–degrading proteins are heavily represented among acylated proteins. The presence of functional deacylase enzymes in *S. aciditrophicus* indicates a potential regulatory system/mechanism by which these bacteria modulate acylation. Uniquely, N^{ε}-acyl-lysine RACS are highly abundant in these syntrophic bacteria, raising the compelling possibility of enzyme modulation during benzoate degradation in this, and potentially, other syntrophic bacteria. Our results outline candidates to further study the impact of acylations within syntrophic systems.

Introduction

Syntrophic bacteria serve critical roles in bioremediation and carbon cycling in anaerobic environments (1–6). They degrade a broad range of aliphatic and aromatic acids to hydrogen, formate, CO_2 and acetate in co-operation with hydrogen- and/or formate-consuming microbes such as the methanogenic archaea that generate methane (2, 3). Here, methanogens and/or other hydrogen- and/or formate-consuming partners are required to maintain low hydrogen and formate levels, such that syntrophic substrate degradation can occur spontaneously (4). This obligate partnership between different members of the microbial community exists because anaerobic degradation is thermodynamically unfavorable when hydrogen or formate levels are high (3, 5). Direct electron transfer between syntrophic microorganisms and their partner microorganisms is also possible (7). Importantly, the degradation of syntrophic substrates requires that multiple enzymatically catalyzed reactions be performed on many acyl-coenzyme A (CoA) intermediates. Even at low hydrogen and formate levels, the necessary acyl-CoA oxidations occurring during syntrophic metabolism are weakly exergonic with free-energy changes close to thermodynamic equilibrium (4). The energetic challenges of living at the edge of thermodynamic feasibility make syntrophic microbes interesting models for exploring energy conservation and metabolic regulation (6).

Syntrophus aciditrophicus is an anaerobic, Gram-negative bacterium that degrades fatty, aromatic and alicyclic (cyclohexane-1-carboxylate) acids to acetate, CO_2, formate and hydrogen when grown in coculture with hydrogen- and/or formate-consuming partner microorganisms. *S. aciditrophicus* can grow in pure culture with crotonate or benzoate (8–10) and it can also use benzoate as an electron acceptor forming cyclohexane-1-carboxylate (11). Previous work has elucidated the pathway for crotonate, benzoate and cyclohexane-1-carboxylate metabolism (12–14) and found that *S. aciditrophicus* uses many of the same enzymes to degrade and to synthesize benzoate and cyclohexane carboxylate (10, 12).

Due to the challenges of isolating, manipulating, and cultivating syntrophic microorganisms, questions still remain regarding the regulatory mechanisms involved in the important anaerobic process. The oxidation of benzoate and other substrates generates reduced cofactors (NADH and $FADH_2$) that, if unregulated, could cause stress and cellular dysfunction (15). Intermediates of substrate metabolism may also stress cells. In *S. aciditrophicus*, the various acid substrates are activated through their respective CoA derivative, which serves as a scaffold through subsequent conversions, until acetate, CO_2, hydrogen and formate are released (6, 16). The intermediates of these pathways are a series of reactive acyl-CoA species (RACS), which have the potential to modify the nucleophilic side chain of lysine (17). Given the multiple stressors that *S. aciditrophicus* cells experience in carbon metabolism and that its enzymes can operate in either direction, depending on the substrate and environmental conditions (10, 12), it is suspected that previously undescribed pathway regulation is involved. The reversibility of syntrophic metabolism and its correspondingly low energy yields makes us consider that post-translational modifications (PTMs), specifically acylations associated with RACS intermediates, may play a role modulating degradations and syntheses in the cell.

Identifying and characterizing PTMs is most often done using mass-spectrometry based protein analysis. At the systems-level, proteomics can capture PTMs occurring across all cellular proteins. However, a system that contains a wide range of acylations brings analytical complications, including potential sequence misidentifications due to isobaric or isomeric combinations (18, 19). Such ambiguities must be addressed to confidently ascribe sequences, modifications, and modified residues in our data. The increase in theoretical search space and scoring thresholds needed to satisfy false-discovery rate cutoffs in database searching are also a consideration for proteomes possessing a broad variety of acylations (20).

Here, we take a systems level approach to identify post-translational acylations in the *S. aciditrophicus* proteome associated with syntrophic benzoate metabolism. Without employing PTM-

specific enrichment procedures, we have identified a wide range of acylations in *S. aciditrophicus* by targeting acyl-CoAs in our database searches that are known to be involved in fatty and aromatic acid degradation in this and other microbes (16, 21). The use of marker ions diagnostic of lysine acylation increased confidence in the identified post-translational modifications (22). We also identified acylations that are previously undocumented across the bacterial domain. An understanding of how these modifications are regulated is currently unknown, but we searched for proteins in the *S. aciditrophicus* genome that have promiscuous deacylase activity in bacterial systems, *e.g.*, sirtuins (23, 24). To this end, we identified one functional sirtuin homolog with deacylase activity among the two present in the *S. aciditrophicus* genome.

Experimental

Media, Cultivation and Cell Harvesting

The pure cultures of *Syntrophus aciditrophicus* strain SB (DSM 26646) was grown in 500-ml Schott bottles with 250 ml of a basal medium (25) with 20 mM crotonate or with 10 mM crotonate plus 2 mM benzoate. *S. aciditrophicus* was grown in coculture with *Methanospirillim hungatei* JF1 (ATCC 27890) in 2-liter Schott bottles with 1 L of Tanner's mineral medium with 14 mM benzoate. Tanner's mineral medium was used for large culture volumes to avoid chemical precipitants that form in large volumes of the basal medium. Tanner's mineral medium (26) contained (per liter): 10 ml of Tanner's minerals, 5 ml of Tanner's metals, 10 ml of Tanner's vitamins, 1 ml of 0.1% of resazurin, 3.75 g $NaHCO_3$, and 20 ml of cysteine-sulfide (2.5 % each) solution (27). The pH of both media was adjusted with NaOH and HCl to pH 7.1 to 7.3. Anaerobic procedures of Balch and Wolfe (28) were used for the preparation of anaerobic medium and solutions and for the inoculation and sampling of cultures. The headspace was pressurized with N_2/CO_2 gas mixture (80%:20% v/v) to 70 kilopascal and the cultures were incubated at 37°C without shaking.

Cultures were grown to mid-log phase and harvested by centrifugation (8,000 x g for 20 minutes at 4°C) under strict anaerobic conditions. The cell pellet was washed twice by resuspending the pellet in 50 mM anoxic potassium phosphate buffer (pH 7.5) and centrifuging as above. The final cell pellet was resuspended in the anoxic potassium phosphate buffer, transferred to cryovials and stored in -80°C (29). Culture manipulations were performed in an anaerobic Coy chamber and all centrifuge steps were done with sealed, anoxic, centrifuge tubes (28).

2D-PAGE of S. aciditrophicus

2D-PAGE were performed and analyzed as described in the Supplemental Materials section by James *et al* (12).

S. aciditrophicus Culture Peptide Preparation

The coculture of *Syntrophus aciditrophicus* S2A along with *Methanospirillum hungatei* JF1 was grown with benzoate as the substrate as previously described (12, 29). For sample workup, frozen cell pellets were resuspended in buffer containing 4.0% (v/v) ammonium lauryl sulfate (ALS), 0.1% sodium deoxycholate (w/v), 5 mM tris(2-carboxyethyl)phosphine (TCEP) and 100 mM ammonium bicarbonate. Peptides were then prepared via the enhanced filter-aided sample preparation (eFASP) method, similar to that described by Erde, *et al.* (30, 31). Briefly, after cell lysis the lysate was exchanged into a buffer consisting of 8M urea, 0.1% (w/v) sodium deoxycholate, and 0.1% (w/v) *n*-octyl glucoside using a 10kDa Microcon ultrafiltration unit (Millipore). Proteins were alkylated with 17 mM iodoacetamide for an hour at room temperature and were buffer exchanged into 0.1% (w/v) sodium deoxycholate, and 0.1% (w/v) n-octyl glucoside in 100mM ABC. Trypsin (1:100 w/w) was added and incubated overnight at 37°C. Ethyl acetate extraction was used to remove detergents, as previously described (30, 31).

Resulting peptides were separated off-line *via* hydrophilic interaction chromatography (HILIC). Fifty µg of peptides (as assayed by Pierce™ Quantitative Fluorometric Peptide Assay) were deposited onto a BioPureSPN MACRO PolyHYDROXYETHYL A (The Nest Group) in 90% acetonitrile and 150 mM ammonium formate, pH 3. Peptides were eluted into six fractions with ammonium formate buffers sequentially decreasing stepwise in organic content: 80% acetonitrile, 78% acetonitrile, 74% acetonitrile, 71% acetonitrile, 40% acetonitrile, and 35.5% acetonitrile. Fractions 1 and 6 were combined for analysis. No antibody affinity treatments were performed to enrich for acylated peptides.

Liquid Chromatography and Tandem Mass Spectrometry

Peptides were mass-measured and fragmented using high-pressure liquid chromatography tandem mass spectrometry (LC-MS/MS). Using an EASY nLC1000 (Thermo Scientific), 200 ng of peptides were loaded onto an Acclaim PepMap100 C18 trap column (Thermo Scientific, Product #16-494-6, 75 µm × 2 cm, 100 Å) and separated on an Acclaim PepMap RSLC C18 analytical column (Thermo Scientific, Product #03-251-873, 75 µm × 25 cm, 100 Å). Buffer A (0.1% formic acid) and buffer B (0.1% formic acid in 100% acetonitrile) were mixed and delivered at 300 nL min^{-1} as the gradient: 3-20% B in 62 minutes, 20-30% B in 31 minutes, 30-50% B in 5 minutes, and 50-80% B in 2 minutes.

A QExactive (Thermo Fisher Scientific, San Jose, CA, USA) mass spectrometer was operated using data-dependent acquisition (DDA) mode. MS scans (*m/z* 300-1800) were acquired at 70,000 resolution, with an automatic gain control (AGC) target of 1E6 and a maximum fill time of 100 ms. The 10 most abundant precursor ions were dissociated sequentially using higher-energy collisional dissociation (HCD) at a normalized collisional energy of 27 (unless otherwise indicated), and MS/MS spectra were acquired at 17,500 resolution with an AGC target of 1E5 at a maximum fill time of 80 ms. The 2D-gel mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [1] partner repository with the dataset identifier PXD025631. The shotgun mass spectrometry

proteomics data been deposited to the ProteomeXchange Consortium via the PRIDE (32) partner repository with the dataset identifier PXD025603.

Database Searching and Data Analysis

QExactive *.RAW files were analyzed using ProteomeDiscoverer (version 1.4) employing Matrix Science's Mascot search algorithm (33). A database concatenating UniProt *S. aciditrophicus* and *M. hungatei* protein sequences (as of July 8, 2019), to the sequences of common contaminants was employed for searches. Parameters for the Mascot search were: enzyme name, trypsin; maximum missed cleavage sites, 2; precursor mass tolerance, 10 ppm; fragment mass tolerance, 0.02 Da; and variable methionine oxidation and cysteine carbamidomethylation. Searches also considered variable acyl modifications on lysines, selected from **Table 1**. These modifications were selected from RACS identified in aromatic acid degrading pathways in a variety of bacteria. Spectra matched to acylated peptides with Mascot ion scores ≥ 20 were subjected to manual examination. A score of 20 is associated with an expectation value of ≥ 0.1, chosen to cast the widest net of spectra possible that can correspond to modification.

Additional Bioinformatic Analyses

The Kyoto Encyclopedia of Genes and Genomes (KEGG) database was used to annotate relevant pathways associated with the acylated proteins identified. Using KEGG Mapper, UniProt accessions were converted to KEGG identifiers and then mapped to the KEGG pathway database (34). To identify GO/KEGG pathways frequently represented by acylated proteins, functional enrichment was performed by the StringApp (Version 1.6.0) in Cytoscape (Version 3.8.2) and *p*-values were corrected for multiple testing within each category using the Benjamini–Hochberg procedure (35).

To search for sequence patterns associated with acylated peptides, the motif-x algorithm (36) integrated into the MoMo Modification Motifs program (Version 5.3.0) on the MEME suite platform (37) was used to identify enriched motifs from acylated sequences of acyl-21-mers (10 amino acids upstream and downstream of the corresponding acylation site) (38). The *S. aciditrophicus* Uniprot proteome sequences were used as the background database. The other parameters were set to default values: *p*-value threshold: 0.000001, minimum number of occurrences for residue/position pair: 10.

All acylated proteins identified were searched against the STRING database (Version 11.0) to reveal potential protein-protein interactions (39). Only protein interactions found among the acylated proteins were selected, removing any external candidate interactions and only interactions with at least medium confidence (>0.4) were included. The interaction network was visualized in Cytoscape (Version 3.8.2).

Cloning, expression and purification of Syn_00042 and Syn_01020

The two *S. aciditrophicus* sirtuin candidate genes annotated as sir2 family proteins were identified in the genome by homology search, obtained via PCR of genomic DNA, cloned into plasmid pMAPLE21 and expressed in *E. coli* strain BL21 as previously described by Arbing *et al.* (40). The recombinant proteins were purified by Ni-affinity chromatography followed by a size exclusion (Superdex 75) and a Q-Sepharose anion exchange column steps. Purity was greater than 95% as visualized by SDS-PAGE. The apparent protein sizes were consistent with the predicted gene/protein sizes.

Sirtuin Assay (Anhydride)

Acyl-insulins were prepared from the corresponding acyl-anhydrides; *i.e.*, glutaric anhydride was used for glutaryl-lysine, using methods adapted from Baeza *et al.* (41). About 25 µmol of anhydride was added to 100 µL of a 1 mg/mL solution of human insulin (Alfa Aesar, J67626) in 100 mM ammonium bicarbonate. After incubating at 4°C for 20 minutes, the solution pH was readjusted to ~8 using an

ammonium hydroxide solution. Anhydride addition and incubation followed by pH adjustment were repeated twice more. Ostensible *O*-acylation was reversed by adding 50% w/v of hydroxylamine hydrochloride in H_2O (adjusted with ammonium hydroxide to pH 7-8). Following overnight, room temperature incubation, the modified products were buffer-exchanged into 100mM ammonium bicarbonate using 3kD MWCO Amicon spin filters (Millipore).

For the matrix-assisted laser desorption/ionization (MALDI)-based activity assay, approximately a 27 µM solution of acyl-modified insulin was mixed with 0.24 µM solution of recombinantly expressed SYN_00042 gene product and excess oxidized nicotinamide adenine dinucleotide (NAD^+) in 100 mM ammonium bicarbonate. The solution was incubated at 37°C for 2 hours. Samples were spotted on a MALDI sample stage mixed 1:1 with a solution of saturated sinapinic acid matrix dissolved in 50% acetonitrile and 0.1% trifluoroacetic acid. MALDI mass spectra were obtained with an Applied BioSystems Voyager-DE STR time-of-flight (TOF) mass spectrometer.

Time-series measurements of glutaryl-insulin employed a Synapt G2-Si ESI-quadrupole TOF for intact mass measurements. The assay was performed at 37°C for the period of time indicated and quenched with 50% acetonitrile. The glutaryl-insulin was diluted to a final concentration of 10 µM. Twenty µM ubiquitin was added to the solution mixture prior to mass measurement as an internal standard. The solution was added to the mass spectrometer via direct infusion.

Experimental design and statistical rationale

S. aciditrophicus cells were grown in pure culture on crotonate as a sole carbon source or with crotonate/benzoate to analyze via 2D-PAGE. Cells were cocultured with *M. hungatei and* grown on benzoate for mass-spectrometry proteomic analysis. Three samples from each culture condition were grown and, for each biological sample, three technical replicates were analyzed. As a priority was to maximize the identification of potentially low stoichiometry post-translational modifications, offline

HILIC was used to increase the depth of the identifiable proteome. Resulting data was searched for the acyl modifications shown in **Table 1**. To identify PTMs, mass spectra matched to acylated peptides with Mascot ion scores ≥ 20 were subjected to manual examination. Modifications were considered present if the tandem mass spectra included acyl-lysine-associated immonium ions (22).

Results

2D-Gel Analysis of S. aciditrophicus Acylated Proteins

S. aciditrophicus cell lysates from axenic cultivation on benzoate supplemented with crotonate (9) were harvested and subjected to two-dimensional polyacrylamide gel electrophoresis (2D-PAGE, **Figure 1**) (42, 43). Spots were excised, digested with trypsin in-gel, and identified using tandem mass spectrometry (12) (**Table 2**). These analyses revealed many instances where the same protein was identified in multiple spots at the same molecular size, but different isoelectric points. Classically, this pattern suggests a heterogeneous cluster of PTMs on a protein, often seen with glycoproteins (44). However, evidence of in *S. aciditrophicus* glycoproteins has not been found. Some of these proteins were identified as housekeeping proteins (**Table 2**), while others have roles in the catabolism of benzoate and other aromatic and fatty acids. Given the high abundance of RACS in these pathways, acyl-lysine modifications were considered possible. Acylation will neutralize the positive charge on lysine, or, in the cases of acidic RACS, such as glutaryl-CoA, will switch lysine to a negatively charged site. Both types of acylation reduce protein isoelectric points, shifting migration towards the anode (the left in standard 2D gel images); hence, a distribution of many and differently charged acylations would give a pattern similar to that seen in **Figure 1**. These modified proteins migrate as multiple distinct species.

Shotgun Proteomics Identifies Many Acyl Modifications in the S. aciditrophicus Proteome

Three biological replicates of *S. aciditrophicus/M. hungatei* cocultures utilizing benzoate as the carbon source were analyzed with three technical replicates per biological sample of LC-MS/MS data-dependent acquisition (DDA) (**supplemental Table S1,S2**). One biological replicate incorporated stepped collision energies to confirm acyl-lysine modifications from the diagnostic immonium ions generated (22). Speculating that peptides acylated by reactive intermediates from benzoate degradation might be observable, we sought evidence for them by including acylation mass shifts in database searches. **Table 1** illustrates the acylations that were considered including known intermediates for *S. aciditrophicus*, for related organisms, and other common acylations. A total of 125 sites were identified in sixty different proteins containing one of 7 different types of acylations (**supplemental Table S3**). The 3-hydroxypimelylation and acetylation species predominated in both the total numbers of proteins and sites. All of PTMs found in each biological replicate are listed in **Table 3**, along with information about the supporting immonium ions. Acylations corresponding to seven of the searched intermediates in *S. aciditrophicus* were found in proteins containing lysine-modified peptides (**Table 3**). Biological replicates demonstrated that the modifications were reproducibly detected in independent samples (**supplemental Fig S2**, **supplemental Table S3**).

Complex proteomic datasets can be subject to misidentification of PTMs due to isobaric peptides, spectral complexity, chimeric precursors, and spectra lacking sufficient sequence-related ions (18). The large search space traversed when seeking evidence of multiple, variable modifications from shotgun datasets demands that the modified peptides be assigned confidently. To meet these challenges, we relied upon the presence of immonium ions specific for each putative acyl modification. Immonium ions and immonium-like ions are strong indicators of acyl modification (22). Proportions of putative acyl-modified spectra displaying diagnostic immonium ions are presented in **Table 3**. We also examined sequences upstream and downstream of each acylated residue to determine if there were sequence

preferences for these modifications and, indeed, we found that nearby glycines favored lysine acetylation (**supplemental Fig S3**).

Pathway Analysis of Modified Proteins Gives Insight into Metabolism

A functional annotation analysis classified proteins by their KEGG pathways and Gene Ontology (GO). The GO functional classification was grouped into three categories: biological process, molecular function, and cellular component. Most acylated proteins mapped onto KEGG pathways had roles in metabolic processes, as expected. Benzoate degradation, oxidative phosphorylation, and carbon metabolism were highly enriched (**Fig 2A**). GO enrichment analyses of all acylated proteins revealed biological processes primarily related to the synthesis of metabolites, as well as energy production (**Figure 2B**)**.** Processes related to nucleotide metabolism were also highly enriched, in line with recent reports that nucleotide-binding regions increase acylation of proximate lysines (45). Altogether, the findings suggest that these acylated proteins are involved in degrading aromatic and fatty acids and carbon metabolism, in general.

Acyl-lysine modifications have been shown to impact protein-protein interactions (PPI) and affect the formation of enzyme complexes, altering their functional role in cellular physiology. To determine the connections among the acylated proteins and to elucidate their functional PPI networks, we generated an interaction network based on the STRING database, visualized through Cytoscape (**Figure 2C**). Within the network there are two highly interconnected clusters, benzoate degradation and oxidative phosphorylation (**Figure 2C**)**.** These strong physiological interactions among all modified protein led us to further investigate the role of protein acylation in the degradation of aromatic compounds.

Benzoate Degradation Enzymes are Heavily Acylated

To view specifically how each type of acylation correlated to functional pathways, proteins were grouped based on their KEGG function to determine the degree and significance of each modification. Proteins across a variety of pathways are modified by different acyl groups at varying levels, summarized by the heat map in **Figure 3A**. Acetylation is found in the widest range of functional pathways, with the highest number of acetyl modifications occurring in proteins related to the biosynthesis of secondary metabolites. Other acyl modifications are enriched in the benzoate degradation pathway proteins, corresponding to the reactive acyl-CoA intermediates in this pathway (16). These findings imply that high local concentrations of intermediates may induce spontaneous modification.

The largest number and widest variety of acyl modifications were found in the benzoate degradation pathway (**Fig 3B**). Ten different proteins in this pathway were identified bearing several different acyl modifications. Acylated proteins included benzoate-CoA ligase, the initial step in the pathway. Several of the other acylated protein are members of gene clusters involved in benzoic acid metabolism (*bam* genes) that are found across a number of anaerobic delta-proteobacteria (46, 47). Three genes in this pathway, *bamR* (SYN_01653), *bamQ* (SYN_01655), and *bamA* (SYN_01654), which successively catalyze the reduction of cyclohexa-1,5-diene-1-carboxyl-CoA to 3-hydroxypimelyl-CoA, all display modifications. All three gene products are acetylated, *bamQ* and *bamA* are glutarylated and 3-hydroxypimelylated, and *bamA* is also 3-hydroxybutyrylated and benzoylated. Interestingly, all five modifications identified on *bamA* are found on the same site, lysine 261 (**supplemental Fig S4**). Peptides containing these lysine residues were also found unmodified.

S. aciditrophicus uses acetyl-CoA synthetase (*acs1*, SYN_02635) to make ATP from acetyl-CoA. Other bacteria, in contrast, employ phosphate acetyltransferase and acetate kinase, enzymes that are absent in the *S. aciditrophicus* genome. Using *acs1* to produce ATP is also remarkable because acetyl-CoA synthetases were previously thought to function only in activating acetate to acetyl-CoA, not in the reverse direction (ATP-forming direction) (29). The rates at which ATP is produced when acetyl-CoA is

the limiting substrate differ between purified and recombinantly expressed Asc1 (V_{max} of 7.5 and 1.2, respectively and K_m of 0.41 and 1.34 mM, respectively) (29), suggesting that there are factors associated with the *in vivo* protein that are not recapitulated when the enzyme is recombinantly expressed. PTMs may be one such factor. Indeed, six different sites on this protein are seen in acylated forms, all of which included acetylation, except for one site that was only butyrylated.

The proteins involved in energy conservation were also heavily acetylated. ATP synthase, Atp1, is a large complex that contains many components. Acetylation sites were identified on several subunits of Atp1, namely on the alpha (SYN_00546), beta (SYN_00544), and gamma (SYN_00545) chains (**supplemental Table S3**). Two homologs of ATP synthase B (the beta subunit) were also acetylated (SYN_00548 and SYN_00549).

Benzoate-CoA Ligase Reveals a Highly Modified Residue in a Conserved Region

Five of the seven acyl modifications identified in *S. aciditrophicus* proteins appeared in the sequence TATGKIQR, present in paralogs Bcl1 (SYN_2898) and Bcl2 (SYN_2896), which differ in the activity towards different fatty and aromatic acids. The modified lysine is critical to benzoate-CoA ligase (BCL) function (48, 49), catalyzing the first step of benzoate degradation. Immonium ions present in the spectra shown in **Figure 4A** confirm that the unique mass shifts do, indeed, result from acylation and are not induced by misidentifications that may result from additive side reactions that occur during sample processing; e.g., formylation from exposure to high concentrations of formic acid or 12-Da mass shifts from exposure to formaldehyde (50, 51). This modification site was further investigated to determine what functional role, if any, the PTMs play. Basic Local Alignment Search Tool (BLAST) sequence alignments against closely related BCLs (48) indicated that the lysine falls in a highly conserved region (**Figure 4B**). Previous *Burkholderia xenovorans* structural studies of benzoate-bound BCL (Bxe_A1419), a homolog with 46% sequence identity to Bcl1 and 44% to Bcl2, indicated that this conserved lysine is located near

the benzoate carboxyl (**Figure 4C**) and is hypothesized to coordinate to benzoate during enzymatic activity (48). Acylation both neutralizes the ε-amino side chain and adds a bulky group disrupting coordination to the carboxylate. The *Rhodopseudomonas palustris* BCL (BadA), with 43% sequence identity to Bcl1 and 42% to Bcl2, has also been found acetylated at that residue (K512); *in vitro* studies established that acetylation abolishes enzymatic activity (49). This behavior suggests a possible avenue of enzyme and, in turn, pathway feedback inhibition.

S. aciditrophicus Sirtuins Deacylate Promiscuously

Given the ubiquity of acyl modifications in *S. aciditrophicus*, controlling and/or regulating acyl modifications would seem to be advantageous and critical. Sirtuin proteins have been shown to possess generalized deacylase activity, often promiscuous with respect to both sequence context and acyl modification (23, 52, 53). Bacterial systems that contain sirtuin homologs are typically denoted as *cobB* or as SIR2 family (24, 54). To determine if a deacylase capabilities are present, the *S. aciditrophicus* genome was searched with BLAST to determine if *E. coli cobB* homologs were present, and SYN_00042 and SYN_01020 were found (**supplemental Fig S5**). SYN_00042 and SYN_01020 were recombinantly expressed to assay activity by mass spectrometry (Materials and Methods).

Insulin, a dual-chain protein with one lysine and two free NH_2-termini, was synthetically acylated with acyl anhydrides. Acylated insulin was incubated with recombinantly produced SYN_00042 gene product and its cofactor, NAD^+. MALDI spectra of glutarylated insulin showed a 114 Da decrease in mass after incubation with the SYN_00042 gene product (**Figure 5A**). Several acyl modifications were tested to assess the sirtuin's specificity. The SYN_00042 gene product had deacylase activity against butyryl and succinyl acylations as well (**supplemental Fig S5**), but did not appear to show activity on shorter acetyl or propionyl chains. A quantitative assay of SYN_00042 gene product deacylase activity on glutaryl-insulin, performed by ESI-MS, clearly shows time-dependent de-glutarylase activity (**Figure 5B**).

SYN_01020 was expressed in an attempt to assess its function; however, the SYN_01020 gene product did not display reproducible activity.

Discussion

PTMs can modulate protein function in response to cellular or environmental changes and may occur spontaneously due to cellular conditions or simply over time (17, 55–57). Acylation is one class of modification that can be spontaneously induced in the presence of acyl phosphate or RACS, or can be enzyme-mediated by a lysine acyltransferase (KAT) (58). Regardless of its origin, acylation can affect enzymatic activity profoundly. *In vitro* work has shown that modifying lysine side chains near catalytic regions of bacterial enzymes can directly alter function (59–61). Work in other systems has shown more subtle but equally significant acylation effects: acetylation can disrupt enzyme complexes, thereby altering activity (62–65). Acetyl-lysine has been identified as a ubiquitous modification in bacteria (66). Most acetylated proteins in our system were involved in the synthesis of secondary (**Figure 3A**), which makes sense given that acetyl-CoA is an intermediate metabolite in many of those pathways. Benzoyl- (52), glutaryl -(67), hydroxybutyryl - (68, 69), crotonyl- (70, 71), and butyryl-lysine (72, 73) have shown function in some systems under certain conditions, although the studies have been less extensive than those for acetyl-lysine, and few, if any, bacterial studies have reported their occurrence. A 3-hydroxypimelylation has previously been reported in *S. aciditrophicus* by our laboratory (22). Detecting these modifications across biological replicates without pre-enrichment suggests that they are uniquely prevalent and abundant in this system. While butyryl-CoA is not expected to be an intermediate in the degradation pathway, *S. aciditrophicus* has genes for butyrate dehydrogenase activity (3, 8) that could synthesize butyryl-CoA upon buildup of crotonyl-CoA or reduced cofactors (8, 12).

To begin to fully understand how these modifications affect global cellular function requires a comprehensive and unbiased catalog of acylations. Recording a biological system's acyl modifications

comprehensively presents many challenges. Low stoichiometries of acyl modifications in many model systems have required enrichment to identify and characterize acylated peptides (67, 74–82). Pan-specific antibodies are valuable for enriching specifically-modified peptides, but have some limitations. Antibody cross-reactivity may lessen the modification specificity of enriched peptides, a problem for western blotting, but not mass spectrometry. Sequence-dependent binding efficiencies can bias results and challenge quantification strategies. Our syntrophic system provides a unique opportunity to investigate the array of acylations without enrichment and bypass potentially limiting technical challenges that are used in other systems. The experimental approach described herein using RACS as a predictor for lysine acylations is complementary to antibody enrichment in that it is blind to the modification type (e.g., specific antibodies used) and it utilizes diagnostic marker ions to validate putative modifications whose uncertainty may increase from the enlarged search space.

Many of the *S. aciditrophicus* proteins decorated with acyl modifications function to degrade benzoate and could result in cellular stress, such as carbon or reductive stress, if left unregulated. Modulating enzymatic activity through protein acylation would be a fitting and elegant mechanism of global metabolic regulation, given the low cellular energy availability and the high acyl-CoA intermediate availability, a feature also evidenced by the unusual ability of this species to use acetyl-CoA to form ATP (29). The benzoate degradation pathway is a likely candidate for this level of regulation as its constituents frequently interact with RACS. While we have identified a wide range of acyl modifications in this system, the biological role that these modifications play has yet to be explored. We can, however, provide some insight into the functional effect of certain acylation sites by drawing on similar modifications in other systems. BCL acylation, specifically acetylation has been identified as a means of negative feedback inhibition; stopping benzoate-CoA ligase activity in *R. palustrus* by directly inhibiting enzymatic catalysis (49). The large number of acyl modifications we report at this site may indicate that when its respective acyl-intermediate builds, the PTM can act as a brake to slow or stop degradation,

thereby acting as a negative feedback inhibitor to mitigate carbon and reductive stress. BCL also consumes ATP, creating a direct link between the energetic state of the cell and aromatic degradation. Additionally, the multiple sites of acetylation found on ATP synthase subunits further hint at a link between bioenergetics and acylation as it too demonstrates a relationship between the reactive metabolites and oxidative phosphorylation.

Beyond the energetics of the system the cellular reductive state may also be regulated by the acylation of enzymes. Benzoate degradation generates NADH and $FADH_2$ and their build up may stress the cell. A buildup of intermediates that subsequently modify and slow the rate of enzymatic catalysis would safeguard against the reductive stress that has been shown to be damaging in other organisms (15, 83, 84). This link is further enforced by the presence of sirtuins, whose activity rely upon the oxidized cofactor NAD^+ for activity (85), innately linking the extent of protein acylation to cellular redox state. Interestingly, the sirtuin assayed in this system appeared to have a bias for longer chain acylations, particularly glutarylation, a trait that is shared with Sirt4 in mammalian systems (86, 87).

The PTMs identified derive from acyl-CoA intermediates at steps which require the loss of 2 [H] (release of two reducing equivalents) (16). Hydrogen buildup is a limiting factor in the syntrophic metabolism, as is evidenced by the need for a hydrogen scavenging partner (4). It is therefore expected that some metabolites would accumulate under high hydrogen concentrations, providing another link between the metabolic conditions of the cell and the observed acyl modifications. Interestingly, the lower benzoate pathway (**Fig 2C**) where 3-hydroxypimelyl-CoA is degraded is conserved across many anaerobes that degrade benzoate and other aromatic compounds (88, 89). The acyl modifications we identified on these proteins suggest that the modifications are likely present in other bacteria sharing the same or similar metabolic pathways, especially under cellular conditions that induce an accumulation of intermediates. *S. aciditrophicus* has demonstrated a wide array of acyl modifications within its proteome. Further investigation into the function of these acylations presents an opportunity

to understand how these essential environmental microbes regulate their metabolism. Other syntrophs also contain critical pathways with RACS intermediates (90–92) and a wider investigation of other syntrophs may further reveal the role RACS play in bacterial metabolism.

Acknowledgements

Funding from the Department of Energy Office of Science (BER) contract DE-FC-02-02ER63421 (to J.A.L. and R.P.G; UCLA/DOE Institute for Genomics and Proteomics), NIH Ruth L. Kirschstein National Research Service 18 Award (to J.Y.F.; GM007185), NSF Award 1911781 to R.P.G and M.J. M, and NSF Graduate Research Fellowship (to J.Y.F.; DGE-1650604) is acknowledged. acknowledged. We are grateful for the 2D-PAGE and mass spectrometry analyses performed by Yanan Yang. We would like to thank Mark Arbing and Annie Shin of the UCLA-DOE Institute Protein Expression Technology Center (supported by grant 447147-EL-21341 from the US Department of Energy) for expression and purification of *Syntrophus aciditrophicus* SYN_00042 and SYN_1020.

Author contributions

JMM, HHN, MJM designed research; JMM, JYF, HHN, HM, NQW, KLJ performed research, JMM, JYF, HM, KLJ, NQW MJM, RPG, RROL analyzed data; and JMM, JYF, MJM, RPG, JAL, RROL wrote and edited the paper.

Conflict of Interest Statement

The authors declare no conflict of interest.

Citations

1. Stams, A. J. M., Sousa, D. Z., Kleerebezem, R., and Plugge, C. M. (2012) Role of syntrophic microbial communities in high-rate methanogenic bioreactors. *Water Sci. Technol.* **66**, 352–362
2. McInerney, M. J., Sieber, J. R., and Gunsalus, R. P. (2009) Syntrophy in anaerobic global carbon cycles. *Curr. Opin. Biotechnol.* **20**, 623–632
3. McInerney, M. J., Rohlin, L., Mouttaki, H., Kim, U., Krupp, R. S., Rios-Hernandez, L., Sieber, J., Struchtemeyer, C. G., Bhattacharyya, A., Campbell, J. W., and Gunsalus, R. P. (2007) The genome of Syntrophus aciditrophicus: life at the thermodynamic limit of microbial growth. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 7600–5
4. Schink, B. (1997) Energetics of syntrophic cooperation in methanogenic degradation. *Microbiol. Mol. Biol. Rev.* **61**, 262–80
5. Stams, A. J. M., and Plugge, C. M. (2009) Electron transfer in syntrophic communities of anaerobic bacteria and archaea. *Nat. Rev. Microbiol.* **7**, 568–577
6. Jackson, B. E., and McInerney, M. J. (2002) Anaerobic microbial metabolism can proceed close to thermodynamic limits. *Nature*. **415**, 454–456
7. Walker, D. J. F., Nevin, K. P., Holmes, D. E., Rotaru, A. E., Ward, J. E., Woodard, T. L., Zhu, J., Ueki, T., Nonnenmann, S. S., McInerney, M. J., and Lovley, D. R. (2020) Syntrophus conductive pili demonstrate that common hydrogen-donating syntrophs can have a direct electron transfer option. *ISME J.* **14**, 837–846
8. Jackson, B. E., Bhupathiraju, V. K., Tanner, R. S., Woese, C. R., and McInerney, M. J. (1999) Syntrophus aciditrophicus sp. nov., a new anaerobic bacterium that degrades fatty acids and benzoate in syntrophic association with hydrogen- using microorganisms. *Arch. Microbiol.* **171**, 107–114

9. Elshahed, M. S., and McInerney, M. J. (2001) Benzoate Fermentation by the Anaerobic Bacterium Syntrophus aciditrophicus in the Absence of Hydrogen-Using Microorganisms. *Appl. Environ. Microbiol.* **67**, 5520–5525

10. Mouttaki, H., Nanny, M. A., and McInerney, M. J. (2007) Cyclohexane carboxylate and benzoate formation from crotonate in Syntrophus aciditrophicus. *Appl. Environ. Microbiol.* **73**, 930–938

11. Mouttaki, H., Nanny, M. A., and McInerney, M. J. (2008) Use of benzoate as an electron acceptor by Syntrophus aciditrophicus grown in pure culture with crotonate. *Environ. Microbiol.* **10**, 3265–3274

12. James, K. L., Kung, J. W., Crable, B. R., Mouttaki, H., Sieber, J. R., Nguyen, H. H., Yang, Y., Xie, Y., Erde, J., Wofford, N. Q., Karr, E. A., Loo, J. A., Ogorzalek Loo, R. R., Gunsalus, R. P., and McInerney, M. J. (2019) Syntrophus aciditrophicus uses the same enzymes in a reversible manner to degrade and synthesize aromatic and alicyclic acids. *Environ. Microbiol.* **21**, 1833–1846

13. Peters, F., Shinoda, Y., McInerney, M. J., and Boll, M. (2007) Cyclohexa-1,5-diene-1-carbonyl-coenzyme A (CoA) hydratases of Geobacter metallireducens and Syntrophus aciditrophicus: Evidence for a common benzoyl-CoA degradation pathway in facultative and strict anaerobes. *J. Bacteriol.* **189**, 1055–1060

14. Kuntze, K., Shinoda, Y., Moutakki, H., McInerney, M. J., Vogt, C., Richnow, H. H., and Boll, M. (2008) 6-Oxocyclohex-1-ene-1-carbonyl-coenzyme A hydrolases from obligately anaerobic bacteria: Characterization and identification of its gene as a functional marker for aromatic compounds degrading anaerobes. *Environ. Microbiol.* **10**, 1547–1556

15. Mavi, P. S., Singh, S., and Kumar, A. (2020) Reductive Stress: New Insights in Physiology and Drug Tolerance of Mycobacterium. *Antioxidants Redox Signal.* **32**, 1348–1366

16. Elshahed, M. S., Bhupathiraju, V. K., Wofford, N. Q., Nanny, M. A., and McInerney, M. J. (2001) Metabolism of Benzoate, Cyclohex-1-ene Carboxylate, and Cyclohexane Carboxylate by

"Syntrophus aciditrophicus" Strain SB in Syntrophic Association with H2-Using Microorganisms. *Appl. Environ. Microbiol.* **67**, 1728–1738

17. Trub, A. G., and Hirschey, M. D. (2018) Reactive Acyl-CoA Species Modify Proteins and Induce Carbon Stress. *Trends Biochem. Sci.* **43**, 369–379

18. Kim, M. S., Zhong, J., and Pandey, A. (2016) Common errors in mass spectrometry-based analysis of post-translational modifications. *Proteomics*. **16**, 700–714

19. Lee, S., Tan, M., Dai, L., Kwon, O. K., Yang, J. S., Zhao, Y., and Chen, Y. (2013) MS/MS of synthetic peptide is not sufficient to confirm new types of protein modifications. *J. Proteome Res.* **12**, 1007–1013

20. Fu, Y. (2012) Bayesian false discovery rates for post-translational modification proteomics. *Stat. Interface*. **5**, 47–59

21. Harwood, C. S., Burchhardt, G., Herrmann, H., and Fuchs, G. (1998) Anaerobic metabolism of aromatic compounds via the benzoyl-CoA pathway. *FEMS Microbiol. Rev.* **22**, 439–458

22. Muroski, J. M., Fu, J. Y., Nguyen, H. H., Ogorzalek Loo, R. R., and Loo, J. A. (2021) Leveraging Immonium Ions for Targeting Acyl-Lysine Modifications in Proteomic Datasets. *Proteomics*. **21**, 2000111

23. Bheda, P., Jing, H., Wolberger, C., and Lin, H. (2016) The Substrate Specificity of Sirtuins. *Annu. Rev. Biochem.* **85**, 405–429

24. Zhao, K., Chai, X., and Marmorstein, R. (2004) Structure and Substrate Binding Properties of cobB, a Sir2 Homolog Protein Deacetylase from Escherichia coli. *J. Mol. Biol.* **337**, 731–741

25. Sieber, J. R., Le, H. M., and Mcinerney, M. J. (2014) The importance of hydrogen and formate transfer for syntrophic fatty, aromatic and alicyclic metabolism. *Environ. Microbiol.* **16**, 177–188

26. Tanner, R. S. (2007) Cultivation of Bacteria and Fungi. in *Manual of Environmental Microbiology*, 3rd Ed. (Hurst, C., Crawford, R., Garland, J., Lipson, D., Mills, A., and Stetzenbach, L. eds), pp. 69–

78, American Society of Microbiology, Washington, DC, 10.1128/9781555815882.ch6

27. McInerney, M. J., Bryant, M. P., and Pfennig, N. (1979) Anaerobic bacterium that degrades fatty acids in syntrophic association with methanogens. *Arch. Microbiol.* **122**, 129–135

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

29. James, K. L., Ríos-Hernández, L. A., Wofford, N. Q., Mouttaki, H., Sieber, J. R., Sheik, C. S., Nguyen, H. H., Yang, Y., Xie, Y., Erde, J., Rohlin, L., Karr, E. A., Loo, J. A., Loo, R. R. O., Hurst, G. B., Gunsalus, R. P., Szweda, L. I., and McInerney, M. J. (2016) Pyrophosphate-dependent ATP formation from acetyl coenzyme a in Syntrophus aciditrophicus, a new twist on ATP formation. *MBio*. 10.1128/mBio.01208-16

30. Erde, J., Loo, R. R. O., and Loo, J. A. (2014) Enhanced FASP (eFASP) to increase proteome coverage and sample recovery for quantitative proteomic experiments. *J. Proteome Res.* **13**, 1885–1895

31. Erde, J., Loo, R. R. O., and Loo, J. A. (2017) Improving proteome coverage and sample recovery with enhanced FASP (eFASP) for quantitative proteomic experiments. in *Methods in Molecular Biology*, pp. 11–18, Humana Press Inc., **1550**, 11–18

32. Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D. J., Inuganti, A., Griss, J., Mayer, G., Eisenacher, M., Pérez, E., Uszkoreit, J., Pfeuffer, J., Sachsenberg, T., Yilmaz, Ş., Tiwary, S., Cox, J., Audain, E., Walzer, M., Jarnuczak, A. F., Ternent, T., Brazma, A., and Vizcaíno, J. A. (2019) The PRIDE database and related tools and resources in 2019: Improving support for quantification data. *Nucleic Acids Res.* **47**, D442–D450

33. Perkins, D. N., Pappin, D. J. C., Creasy, D. M., and Cottrell, J. S. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. in *Electrophoresis*,

pp. 3551–3567, Wiley-VCH Verlag, **20**, 3551–3567

34. Kanehisa, M., and Sato, Y. (2020) KEGG Mapper for inferring cellular functions from protein sequences. *Protein Sci.* **29**, 28–35

35. Doncheva, N. T., Morris, J. H., Gorodkin, J., and Jensen, L. J. (2019) Cytoscape StringApp: Network Analysis and Visualization of Proteomics Data. *J. Proteome Res.* **18**, 623–632

36. Schwartz, D., and Gygi, S. P. (2005) An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. *Nat. Biotechnol.* **23**, 1391–1398

37. Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., Ren, J., Li, W. W., and Noble, W. S. (2009) MEME Suite: Tools for motif discovery and searching. *Nucleic Acids Res.* 10.1093/nar/gkp335

38. Cheng, A., Grant, C. E., Noble, W. S., and Bailey, T. L. (2019) MoMo: discovery of statistically significant post-translational modification motifs. *Bioinformatics*. **35**, 2774–2782

39. Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N. T., Morris, J. H., Bork, P., Jensen, L. J., and Von Mering, C. (2019) STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613

40. Arbing, M. A., Chan, S., Harris, L., Kuo, E., Zhou, T. T., Ahn, C. J., Nguyen, L., He, Q., Lu, J., Menchavez, P. T., Shin, A., Holton, T., Sawaya, M. R., Cascio, D., and Eisenberg, D. (2013) Heterologous Expression of Mycobacterial Esx Complexes in Escherichia coli for Structural Studies Is Facilitated by the Use of Maltose Binding Protein Fusions. *PLoS One*. **8**, e81753

41. Baeza, J., Smallegan, M. J., and Denu, J. M. (2015) Site-Specific Reactivity of Nonenzymatic Lysine Acetylation. *ACS Chem. Biol.* **10**, 122–128

42. O'Farrell, P. H. (1975) High resolution two dimensional electrophoresis of proteins. *J. Biol. Chem.* **250**, 4007–4021

43. Ogorzalek Loo, R. R., Cavalcoli, J. D., VanBogelen, R. A., Mitchell, C., Loo, J. A., Moldover, B., and Andrews, P. C. (2001) Virtual 2-D gel electrophoresis: Visualization and analysis of the E. coli proteome by mass spectrometry. *Anal. Chem.* **73**, 4063–4070

44. Kleinert, P., Kuster, T., Arnold, D., Jaeken, J., Heizmann, C. W., and Troxler, H. (2007) Effect of glycosylation on the protein pattern in 2-D-gel electrophoresis. *Proteomics*. **7**, 15–22

45. James, A. M., Smith, A. C., Ding, S., Houghton, J. W., Robinson, A. J., Antrobus, R., Fearnley, I. M., and Murphy, M. P. (2020) Nucleotide-binding sites can enhance N-acylation of nearby protein lysine residues. *Sci. Rep.* **10**, 1–13

46. Wischgoll, S., Heintz, D., Peters, F., Erxleben, A., Sarnighausen, E., Reski, R., Van Dorsselaer, A., and Boll, M. (2005) Gene clusters involved in anaerobic benzoate degradation of Geobacter metallireducens. *Mol. Microbiol.* **58**, 1238–1252

47. Carmona, M., Zamarro, M. T., Blázquez, B., Durante-Rodríguez, G., Juárez, J. F., Valderrama, J. A., Barragán, M. J. L., García, J. L., and Díaz, E. (2009) Anaerobic Catabolism of Aromatic Compounds: a Genetic and Genomic View. *Microbiol. Mol. Biol. Rev.* **73**, 71–133

48. Bains, J., and Boulanger, M. J. (2007) Biochemical and Structural Characterization of the Paralogous Benzoate CoA Ligases from Burkholderia xenovorans LB400: Defining the Entry Point into the Novel Benzoate Oxidation (box) Pathway. *J. Mol. Biol.* **373**, 965–977

49. Crosby, H. A., Heiniger, E. K., Harwood, C. S., and Escalante-Semerena, J. C. (2010) Reversible Nε-lysine acetylation regulates the activity of acyl-CoA synthetases involved in anaerobic benzoate catabolism in Rhodopseudomonas palustris. *Mol. Microbiol.* **76**, 874–888

50. Du, Y., Wang, F., May, K., Xu, W., and Liu, H. (2012) Determination of deamidation artifacts introduced by sample preparation using 18O-labeling and tandem mass spectrometry analysis. *Anal. Chem.* **84**, 6355–6360

51. Tang, L., Wu, Z., Wang, J., and Zhang, X. (2020) Formaldehyde Derivatization, an Unexpected Side

Reaction During Filter-Aided Sample Preparation. *Anal. Chem.* 10.1021/acs.analchem.0c01981

52. Huang, H., Zhang, D., Wang, Y., Perez-Neut, M., Han, Z., Zheng, Y. G., Hao, Q., and Zhao, Y. (2018) Lysine benzoylation is a histone mark regulated by SIRT2. *Nat. Commun.* **9**, 1–11

53. Mathias, R. A., Greco, T. M., Oberstein, A., Budayeva, H. G., Chakrabarti, R., Rowland, E. A., Kang, Y., Shenk, T., and Cristea, I. M. (2014) Sirtuin 4 is a lipoamidase regulating pyruvate dehydrogenase complex activity. *Cell*. **159**, 1615–1625

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

55. Lindner, H., and Helliger, W. (2001) Age-dependent deamidation of asparagine residues in proteins. *Exp. Gerontol.* **36**, 1551–1563

56. Aswad, D. W., Paranandi, M. V., and Schurter, B. T. (2000) Isoaspartate in peptides and proteins: Formation, significance, and analysis. *J. Pharm. Biomed. Anal.* **21**, 1129–1136

57. Dalle-Donne, I., Aldini, G., Carini, M., Colombo, R., Rossi, R., and Milzani, A. (2006) Protein carbonylation, cellular dysfunction, and disease progression. *J. Cell. Mol. Med.* **10**, 389–406

58. Christensen, D. G., Baumgartner, J. T., Xie, X., Jew, K. M., Basisty, N., Schilling, B., Kuhn, M. L., and Wolfe, A. J. (2019) Mechanisms, detection, and relevance of protein acetylation in prokaryotes. *MBio*. 10.1128/mBio.02708-18

59. Christensen, D. G., Xie, X., Basisty, N., Byrnes, J., McSweeney, S., Schilling, B., and Wolfe, A. J. (2019) Post-translational Protein Acetylation: An elegant mechanism for bacteria to dynamically regulate metabolic functions. *Front. Microbiol.* **10**, 1604

60. Garrity, J., Gardner, J. G., Hawse, W., Wolberger, C., and Escalante-Semerena, J. C. (2007) N-lysine propionylation controls the activity of propionyl-CoA synthetase. *J. Biol. Chem.* **282**, 30239–30245

61. Gardner, J. G., Grundy, F. J., Henkin, T. M., and Escalante-Semerena, J. C. (2006) Control of Acetyl-

Coenzyme A Synthetase (AcsA) Activity by Acetylation/Deacetylation without NAD+ Involvement in Bacillus subtilis. *J. Bacteriol.* **188**, 5460–5468

62. Kitata, R. B., Dimayacyac-Esleta, B. R. T., Choong, W.-K., Tsai, C.-F., Lin, T.-D., Tsou, C.-C., Weng, S.-H., Chen, Y.-J., Yang, P.-C., Arco, S. D., Nesvizhskii, A. I., Sung, T.-Y., and Chen, Y.-J. (2015) Mining Missing Membrane Proteins by High-pH Reverse-Phase StageTip Fractionation and Multiple Reaction Monitoring Mass Spectrometry. *J. Proteome Res.* **14**, 3658–69

63. Zhu, Y., Zou, X., Dean, A. E., Brien, J. O., Gao, Y., Tran, E. L., Park, S. H., Liu, G., Kieffer, M. B., Jiang, H., Stauffer, M. E., Hart, R., Quan, S., Satchell, K. J. F., Horikoshi, N., Bonini, M., and Gius, D. (2019) Lysine 68 acetylation directs MnSOD as a tetrameric detoxification complex versus a monomeric tumor promoter. *Nat. Commun.* 10.1038/s41467-019-10352-4

64. Zhang, X., Yuan, Z., Zhang, Y., Yong, S., Salas-Burgos, A., Koomen, J., Olashaw, N., Parsons, J. T., Yang, X. J., Dent, S. R., Yao, T. P., Lane, W. S., and Seto, E. (2007) HDAC6 Modulates Cell Motility by Altering the Acetylation Level of Cortactin. *Mol. Cell*. **27**, 197–213

65. Tang, Y., Zhao, W., Chen, Y., Zhao, Y., and Gu, W. (2008) Acetylation Is Indispensable for p53 Activation. *Cell*. **133**, 612–626

66. VanDrisse, C. M., and Escalante-Semerena, J. C. (2019) Protein Acetylation in Bacteria. *Annu. Rev. Microbiol.* **73**, 111–132

67. Tan, M., Peng, C., Anderson, K. A., Chhoy, P., Xie, Z., Dai, L., Park, J., Chen, Y., Huang, H., Zhang, Y., Ro, J., Wagner, G. R., Green, M. F., Madsen, A. S., Schmiesing, J., Peterson, B. S., Xu, G., Ilkayeva, O. R., Muehlbauer, M. J., Braulke, T., Mühlhausen, C., Backos, D. S., Olsen, C. A., McGuire, P. J., Pletcher, S. D., Lombard, D. B., Hirschey, M. D., and Zhao, Y. (2014) Lysine glutarylation is a protein posttranslational modification regulated by SIRT5. *Cell Metab.* **19**, 605–617

68. Liu, K., Li, F., Sun, Q., Lin, N., Han, H., You, K., Tian, F., Mao, Z., Li, T., Tong, T., Geng, M., Zhao, Y., Gu, W., and Zhao, W. (2019) p53 β-hydroxybutyrylation attenuates p53 activity. *Cell Death Dis.*

10, 1–13

69. Xie, Z., Zhang, D., Chung, D., Tang, Z., Huang, H., Dai, L., Qi, S., Li, J., Colak, G., Chen, Y., Xia, C., Peng, C., Ruan, H., Kirkey, M., Wang, D., Jensen, L. M., Kwon, O. K., Lee, S., Pletcher, S. D., Tan, M., Lombard, D. B., White, K. P., Zhao, H., Li, J., Roeder, R. G., Yang, X., and Zhao, Y. (2016) Metabolic Regulation of Gene Expression by Histone Lysine β-Hydroxybutyrylation. *Mol. Cell*. **62**, 194–206

70. Sun, C. F., Xu, W. F., Zhao, Q. W., Luo, S., Chen, X. A., Li, Y. Q., and Mao, X. M. (2020) Crotonylation of key metabolic enzymes regulates carbon catabolite repression in Streptomyces roseosporus. *Commun. Biol.* **3**, 1–14

71. Sabari, B. R., Tang, Z., Huang, H., Yong-Gonzalez, V., Molina, H., Kong, H. E., Dai, L., Shimada, M., Cross, J. R., Zhao, Y., Roeder, R. G., and Allis, C. D. (2015) Intracellular Crotonyl-CoA Stimulates Transcription through p300-Catalyzed Histone Crotonylation. *Mol. Cell*. **58**, 203–215

72. Chen, Y., Sprung, R., Tang, Y., Ball, H., Sangras, B., Kim, S. C., Falck, J. R., Peng, J., Gu, W., and Zhao, Y. (2007) Lysine propionylation and butyrylation are novel post-translational modifications in histones. *Mol. Cell. Proteomics*. **6**, 812–819

73. Xu, J. Y., Xu, Z., Liu, X. X., Tan, M., and Ye, B. C. (2018) Protein Acetylation and Butyrylation Regulate the Phenotype and Metabolic Shifts of the Endospore-forming Clostridium acetobutylicum. *Mol. Cell. Proteomics*. **17**, 1156–1169

74. Weinert, B. T., Satpathy, S., Hansen, B. K., Lyon, D., Jensen, L. J., and Choudhary, C. (2017) Accurate quantification of site-specific acetylation stoichiometry reveals the impact of Sirtuin deacetylase CobB on the E. coli acetylome. *Mol. Cell. Proteomics*. **16**, 759–769

75. Hansen, B. K., Gupta, R., Baldus, L., Lyon, D., Narita, T., Lammers, M., Choudhary, C., and Weinert, B. T. (2019) Analysis of human acetylation stoichiometry defines mechanistic constraints on protein regulation. *Nat. Commun.* 10.1038/s41467-019-09024-0

76. Weinert, B. T., Iesmantavicius, V., Moustafa, T., Schölz, C., Wagner, S. A., Magnes, C., Zechner, R., and Choudhary, C. (2014) Acetylation dynamics and stoichiometry in Saccharomyces cerevisiae. *Mol. Syst. Biol.* 10.1002/msb.134766

77. Sadhukhan, S., Liu, X., Ryu, D., Nelson, O. D., Stupinski, J. A., Li, Z., Chen, W., Zhang, S., Weiss, R. S., Locasale, J. W., Auwerx, J., and Lin, H. (2016) Metabolomics-assisted proteomics identifies succinylation and SIRT5 as important regulators of cardiac function. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 4320–4325

78. Nishida, Y., Rardin, M. J., Carrico, C., He, W., Sahu, A. K., Gut, P., Najjar, R., Fitch, M., Hellerstein, M., Gibson, B. W., and Verdin, E. (2015) SIRT5 Regulates both Cytosolic and Mitochondrial Protein Malonylation with Glycolysis as a Major Target. *Mol. Cell*. **59**, 321–332

79. Schilling, B., Christensen, D., Davis, R., Sahu, A. K., Hu, L. I., Walker-Peddakotla, A., Sorensen, D. J., Zemaitaitis, B., Gibson, B. W., and Wolfe, A. J. (2015) Protein acetylation dynamics in response to carbon overflow in Escherichia coli. *Mol. Microbiol.* **98**, 847–863

80. Basisty, N., Meyer, J. G., Wei, L., Gibson, B. W., and Schilling, B. (2018) Simultaneous Quantification of the Acetylome and Succinylome by 'One-Pot' Affinity Enrichment. *Proteomics*. **18**, 1800123

81. Cao, J., Wang, T., Wang, Q., Zheng, X., and Huang, L. (2019) Functional Insights into Protein Acetylation in the Hyperthermophilic Archaeon Sulfolobus islandicus. *Mol. Cell. Proteomics*. **18**, 1572–1587

82. Zhang, D., Tang, Z., Huang, H., Zhou, G., Cui, C., Weng, Y., Liu, W., Kim, S., Lee, S., Perez-Neut, M., Ding, J., Czyz, D., Hu, R., Ye, Z., He, M., Zheng, Y. G., Shuman, H. A., Dai, L., Ren, B., Roeder, R. G., Becker, L., and Zhao, Y. (2019) Metabolic regulation of gene expression by histone lactylation. *Nature*. **574**, 575–580

83. Rajasekaran, N. S., Connell, P., Christians, E. S., Yan, L. J., Taylor, R. P., Orosz, A., Zhang, X. Q.,

Stevenson, T. J., Peshock, R. M., Leopold, J. A., Barry, W. H., Loscalzo, J., Odelberg, S. J., and Benjamin, I. J. (2007) Human αB-Crystallin Mutation Causes Oxido-Reductive Stress and Protein Aggregation Cardiomyopathy in Mice. *Cell*. **130**, 427–439

84. Trotter, E. W., and Grant, C. M. (2002) Thioredoxins are required for protection against a reductive stress in the yeast Saccharomyces cerevisiae. *Mol. Microbiol.* **46**, 869–878

85. Sauve, A. A., Wolberger, C., Schramm, V. L., and Boeke, J. D. (2006) The Biochemistry of Sirtuins. *Annu. Rev. Biochem.* **75**, 435–465

86. Anderson, K. A., Huynh, F. K., Fisher-Wellman, K., Stuart, J. D., Peterson, B. S., Douros, J. D., Wagner, G. R., Thompson, J. W., Madsen, A. S., Green, M. F., Sivley, R. M., Ilkayeva, O. R., Stevens, R. D., Backos, D. S., Capra, J. A., Olsen, C. A., Campbell, J. E., Muoio, D. M., Grimsrud, P. A., and Hirschey, M. D. (2017) SIRT4 Is a Lysine Deacylase that Controls Leucine Metabolism and Insulin Secretion. *Cell Metab.* **25**, 838-855.e15

87. Pannek, M., Simic, Z., Fuszard, M., Meleshin, M., Rotili, D., Mai, A., Schutkowski, M., and Steegborn, C. (2017) Crystal structures of the mitochondrial deacylase Sirtuin 4 reveal isoform-specific acyl recognition and regulation features. *Nat. Commun.* **8**, 1–13

88. Gallus, C., and Schink, B. (1994) Anaerobic degradation of pimelate by newly isolated denitrifying bacteria. *Microbiology*. **140**, 409–416

89. Harrison, F. H., and Harwood, C. S. (2005) The pimFABCDE operon from Rhodopseudomonas palustris mediates dicarboxylic acid degradation and participates in anaerobic benzoate degradation. *Microbiology*. **151**, 727–736

90. Schöcke, L., and Schink, B. (1999) Energetics and biochemistry of fermentative benzoate degradation by Syntrophus gentianae. *Arch. Microbiol.* **171**, 331–337

91. Müller, N., Worm, P., Schink, B., Stams, A. J. M., and Plugge, C. M. (2010) Syntrophic butyrate and propionate oxidation processes: From genomes to reaction mechanisms. *Environ. Microbiol. Rep.*

2, 489–499

92. Boll, M., Geiger, R., Junghare, M., and Schink, B. (2020) Microbial degradation of phthalates: biochemistry and environmental implications. *Environ. Microbiol. Rep.* **12**, 3–15

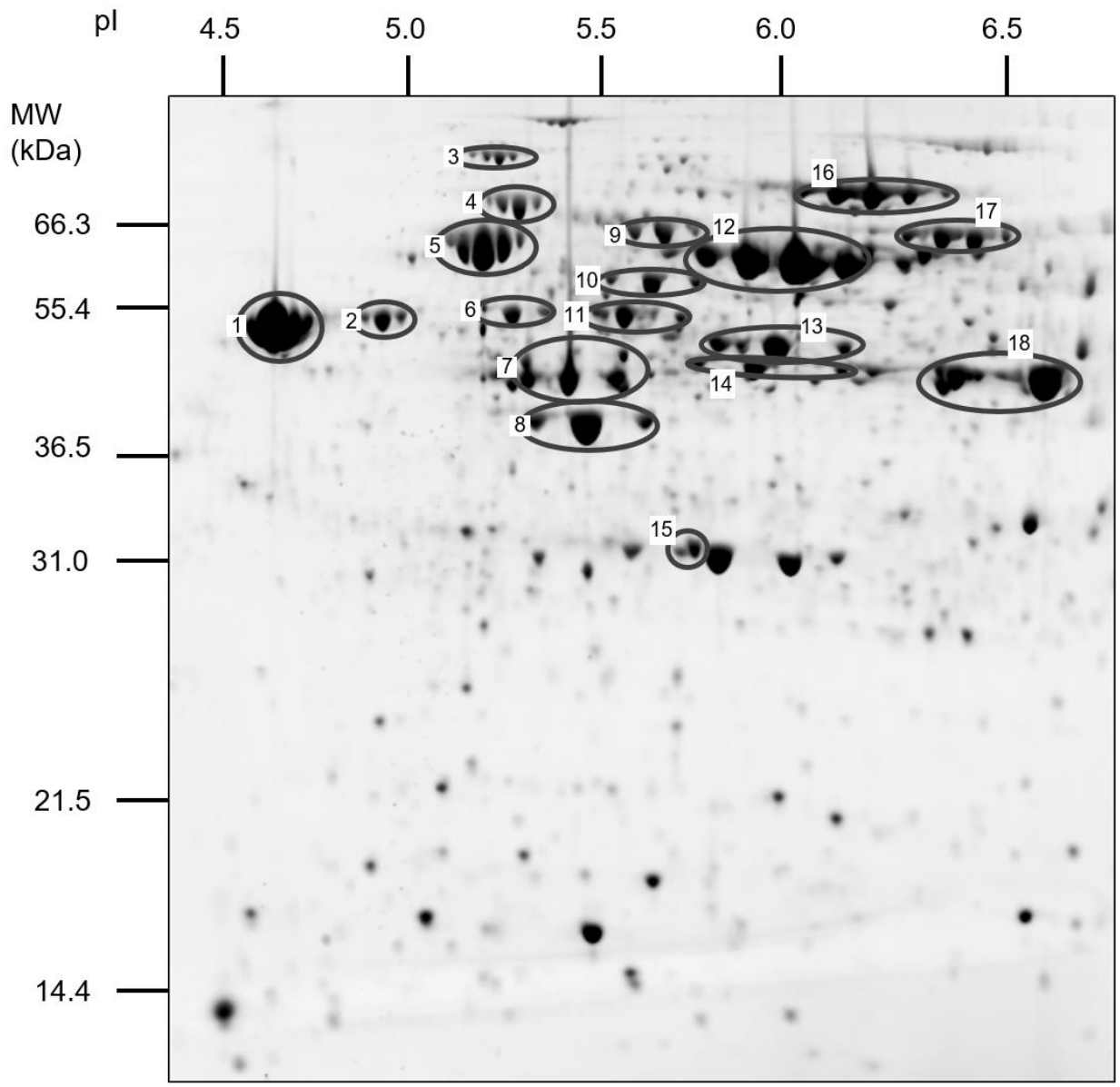


Figure 1. Proteomic patterns of *S. aciditrophicus* as detected by 2D-PAGE. Two-dimensional gel analysis of the *S. aciditrophicus* proteome for cells grown on crotonate supplemented with benzoate. The x-axis is separated by isoelectric focusing (IEF), whereas the y-axis is separated by SDS-PAGE. Circled are protein that show IEF separation for a given molecular weight. Protein spot identifications are listed in Table 2.

Acyl Modification	Chemical Formula	Monoisotopic Mass Shift	Average Mass Shift	Immonium Ion	Cyclized Immonium Ion
Benzoyl	C_7H_4O	104.0262	104.1063	205.1336	188.1070
Cyclohexa-1,5-diene-1-carboxyl	C_7H_6O	106.0419	106.1222	207.1493	190.1227
6-Oxocyclohex-1-ene-1-carboxyl	$C_7H_6O_2$	122.0368	122.1216	223.1442	206.1176
Cyclohex-1-ene-1-carboxyl	C_7H_8O	108.0575	108.1381	209.1649	192.1383
6-Hydroxycyclohex-1-ene-1-carboxyl	$C_7H_8O_2$	124.0524	124.1375	225.1598	208.1332
2-Oxocyclohexane-carboxyl	$C_7H_8O_2$	124.0524	124.1375	225.1598	208.1332
2-Heptenedioyl	$C_7H_8O_3$	140.0473	140.1369	241.1547	224.1281
3-Oxopimelyl	$C_7H_8O_4$	156.0423	156.1363	257.1497	240.1231
Cyclohexane-1-carboxyl	$C_7H_{10}O$	110.0732	110.154	211.1806	194.1540
2-Hydroxycyclohexane-carboxyl	$C_7H_{10}O_2$	126.0681	126.1534	227.1755	210.1489
Pimelyl	$C_7H_{10}O_3$	142.063	142.1528	243.1704	226.1438
3-Hydroxypimelyl	$C_7H_{10}O_4$	158.0579	158.1522	259.1653	242.1387
Glutaryl	$C_5H_6O_3$	114.0317	114.0996	215.1391	198.1125
Glutaconyl	$C_5H_4O_3$	112.016	112.0837	213.1234	196.0968
Crotonyl	C_4H_4O	68.02621	68.07413	169.1336	152.1070
Acetoacetyl	$C_4H_4O_2$	84.02113	84.07353	185.1285	168.1019
Succinyl	$C_4H_4O_3$	100.016	100.0729	201.1234	184.0968
Butyryl	C_4H_6O	70.04186	70.09001	171.1492	154.1226
3-Hydroxybutyryl	$C_4H_6O_2$	86.03678	86.08942	187.1441	170.1176
Acetyl	C_2H_2O	42.01056	42.03677	143.1179	126.0913

Table 1. List of searched acylations. All modifications searched against the *S. aciditrophicus* data set. Modifications were selected from known intermediates in anaerobic benzoate degradation species.

Spot Number	Locus Tag	Protein Function
1	SYN_03116	Hypothetical Exported Protein
2	SYN_00544	ATP Synthase Beta Chain
3	SYN_02966	Phosphoenolpyruvate Synthase
4	SYN_01983	Chaperone Protein
5	SYN_03223	60 kDa Chaperonin
6	SYN_00198	Porin
7	SYN_01909	60 kDa chaperonin 3
8	SYN_00480	Acyl-CoA Dehydrogenase
9	SYN_01709	Ketol-acid reductoisomerase
10	SYN_00546	ATP Synthase Subunit Alpha 1
11	SYN_00983	Elongation Factor Tu
12	SYN_02898	Benzoate-CoA Ligase (Bcl1)
13	SYN_01681	Acetyl-CoA Acetyltransferase
14	SYN_02586	Cyclohexane-1-carbonyl-CoA Dehydrogenase
15	SYN_01653	Enoyl-CoA Hydratase
16	SYN_02635	Acetyl-CoA Synthetase (Acs1)
17	SYN_03128	Cyclohexane-1-carboxylate-CoA Ligase (12)
18	SYN_01654	6-oxocyclohex-1-ene-1-carbonyl-CoA Hydrolase

Table 2. Protein identified in 2D gel spots. Gel spots numbered in Figure 1 were excised and their identities were determined by mass spectrometry on a QStar-XL qTOF mass spectrometer.

Modification	Proteins Identified	Sites Identified	Cyclized Immonium Ion *m/z*	Peptides with Cyclized Immonium Ion (%)
Benzoylation	1	2	188.107	100%
3-Hydroxypimelylation	16	20	224.129, 242.139	75%,55%
Glutarylation	4	6	198.112	100%
Crotonylation	2	2	152.107	100%
3-Hydroxybutyrylation	4	6	170.118	83%
Acetylation	48	104	126.091	94%
Butyrylation	3	3	154.123	33%

Table 3. Summary of acyl-lysine modifications identified in *S. aciditrophicus*.

For each type of modification, the number of unique identified proteins and the total number of sites are presented. The cyclized immonium ion, which gives confidence to the presence of an acyl modification, is stated. The percentage of peptides with spectra that contained the diagnostic ion are stated.

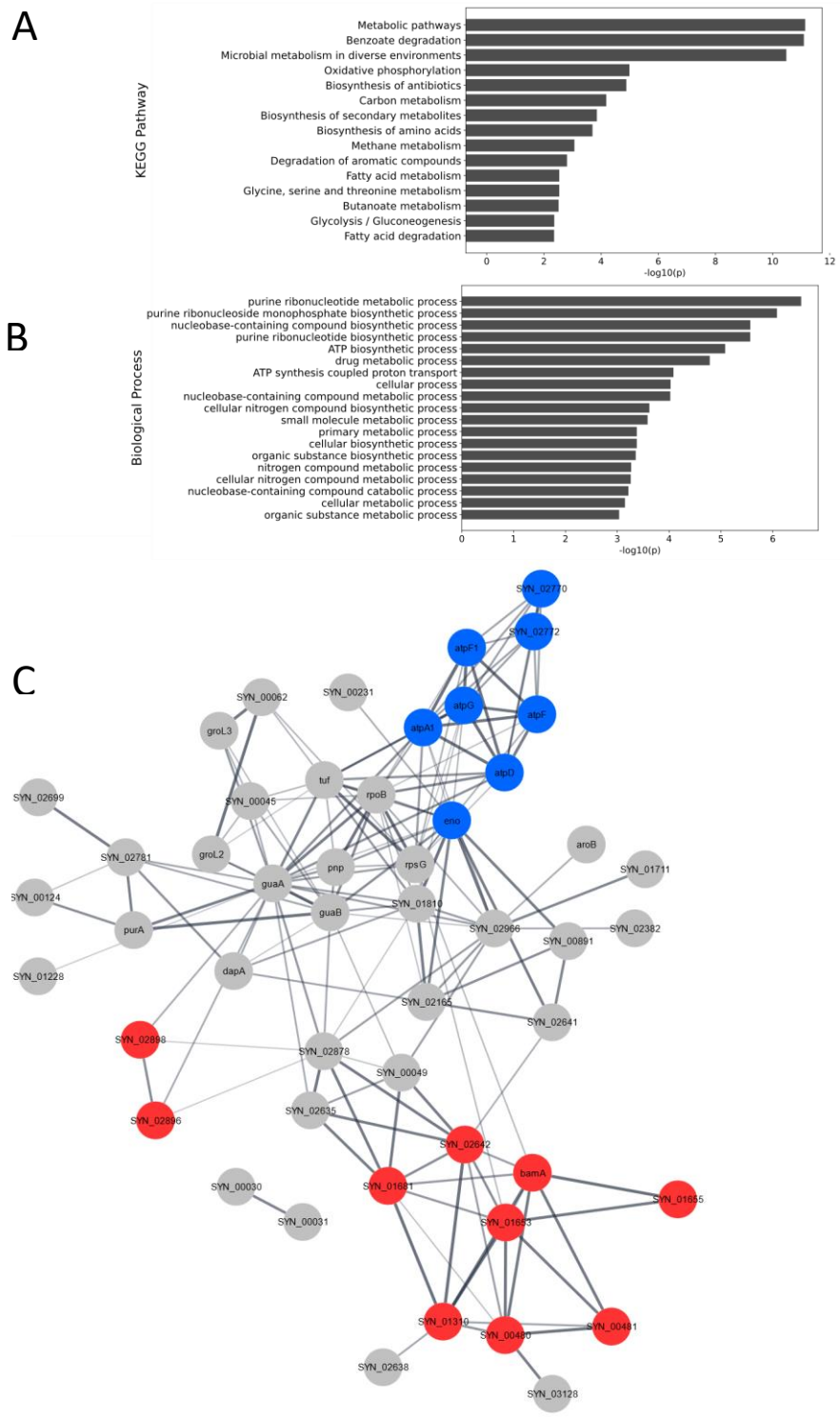


Figure 2. Functional enrichment of acylated proteins. A) Gene ontology biological processes that are enriched B) Enriched KEGG pathway processes. C) A functional protein association network generated using the STRING database. Proteins involved in oxidative phosphorylation are colored in blue. Proteins involved in degrading benzoate are red.

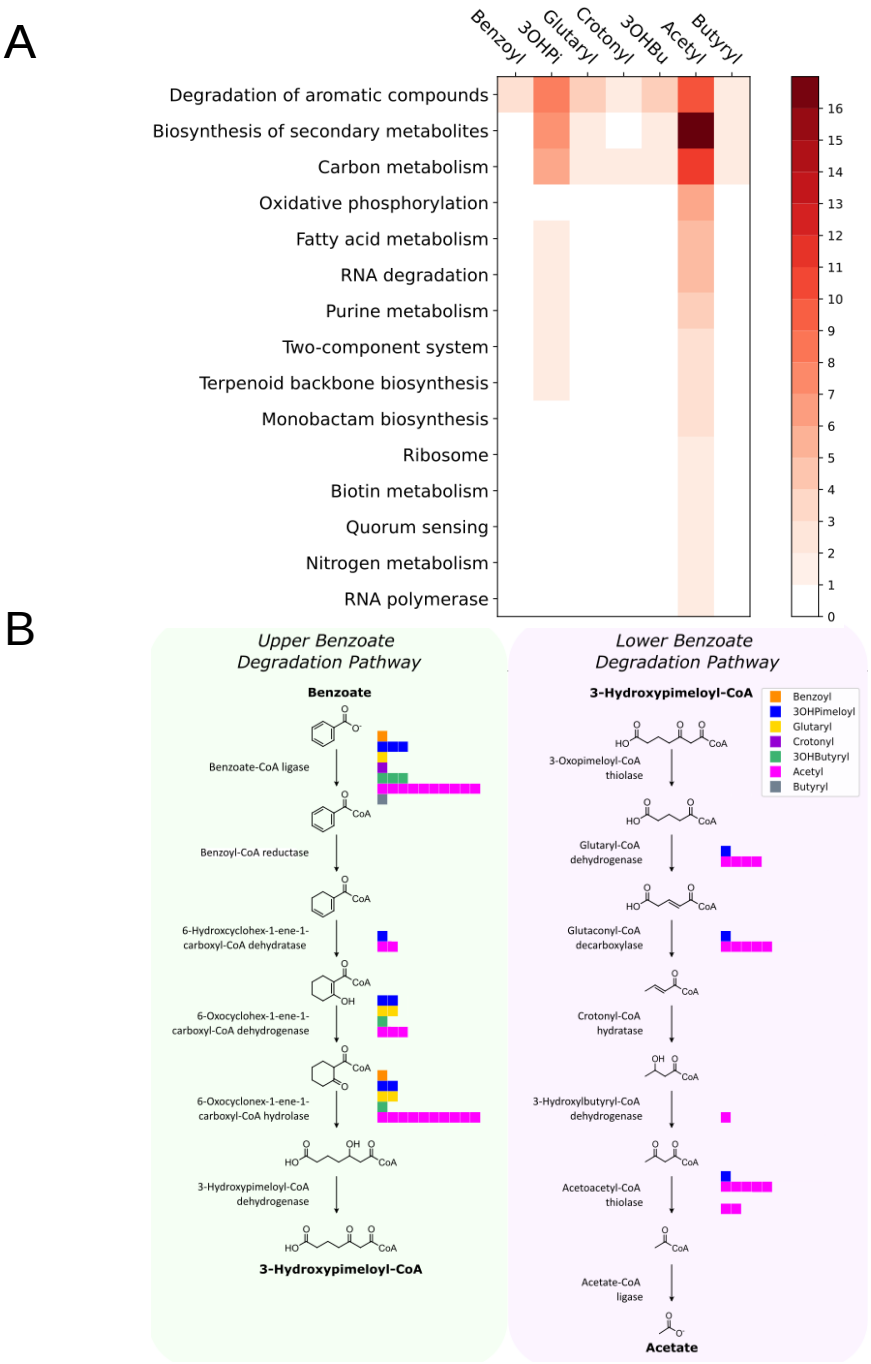


Figure 3. Mapping sites of protein lysine acylation based on COG functional pathway analysis. A) A heatmap indicating the pathways (KEGG Ontology) involved for the acyl-modified protein identified. The color indicates the number of proteins in each category that have the specified acylation. B) shows modified sites identified in the 'upper' and 'lower' benzoate pathway. Bar plots illustrate the range of modifications found on associated enzymes. One square indicates one site of modification and each color represents a different acylation. RACS modification types are indicated in the upper right corner boxed color code.

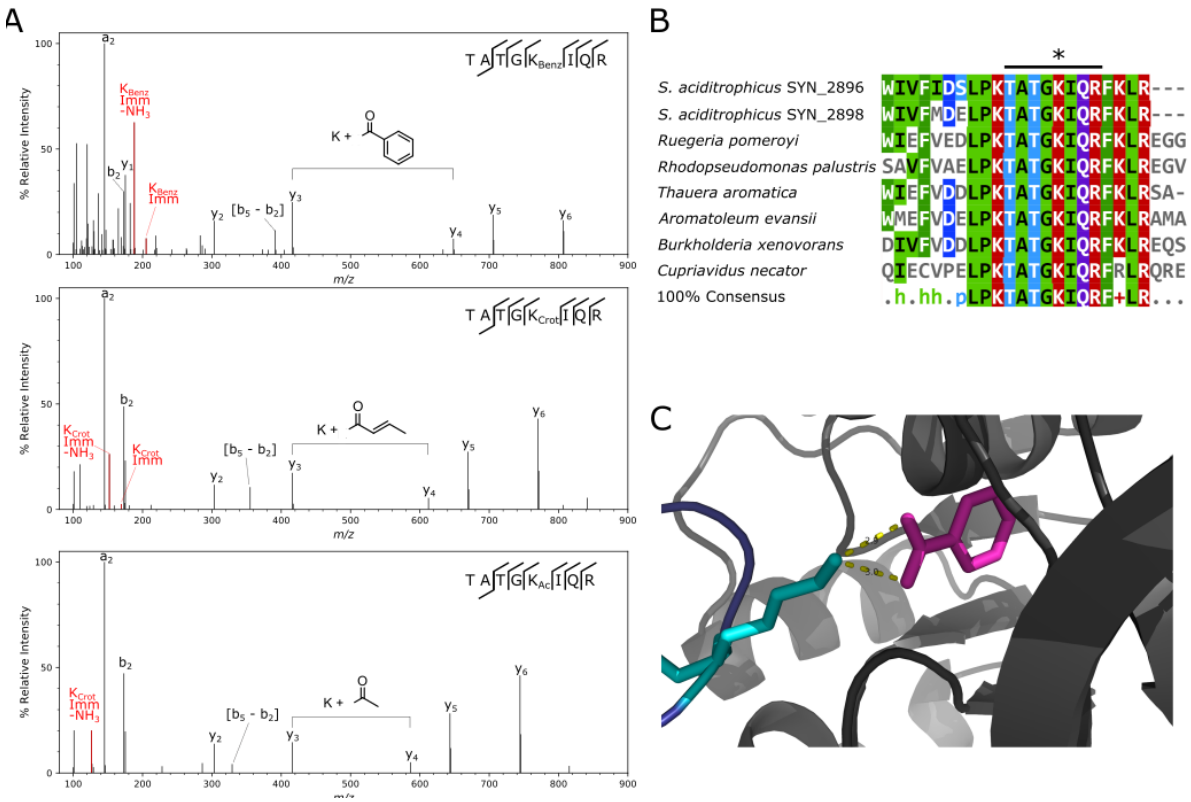


Figure 4. Insight into the heavily modified benzoate-CoA ligase enzyme. A) HCD spectra of select acylated spectra of the peptide **TATGKIQR** from the benzoate-CoA ligase (BCL) protein. The benzoylated spectrum (top) was collected with a stepped NCE method of 27 V/40 V. Other spectra were collected at 27 NCE. Prominent fragment ions that assist in the identification of the sequence and modification are annotated. Immonium ions of modifications are denoted in red. B) Sequence alignment of known BCLs across microbial systems. The line above the sequence (TATG**K**IQR) denotes the peptide identified in (A). The asterisk (*) denotes the corresponding modified lysine residue identified in *S. aciditrophicus*. C) Crystal structure of a closely related BCL from *Burkholderia xenovorans*, determined by Bains and Boulanger (48) , indicates that the lysine residue of interest is proximal to a bound benzoate molecule in the binding pocket.

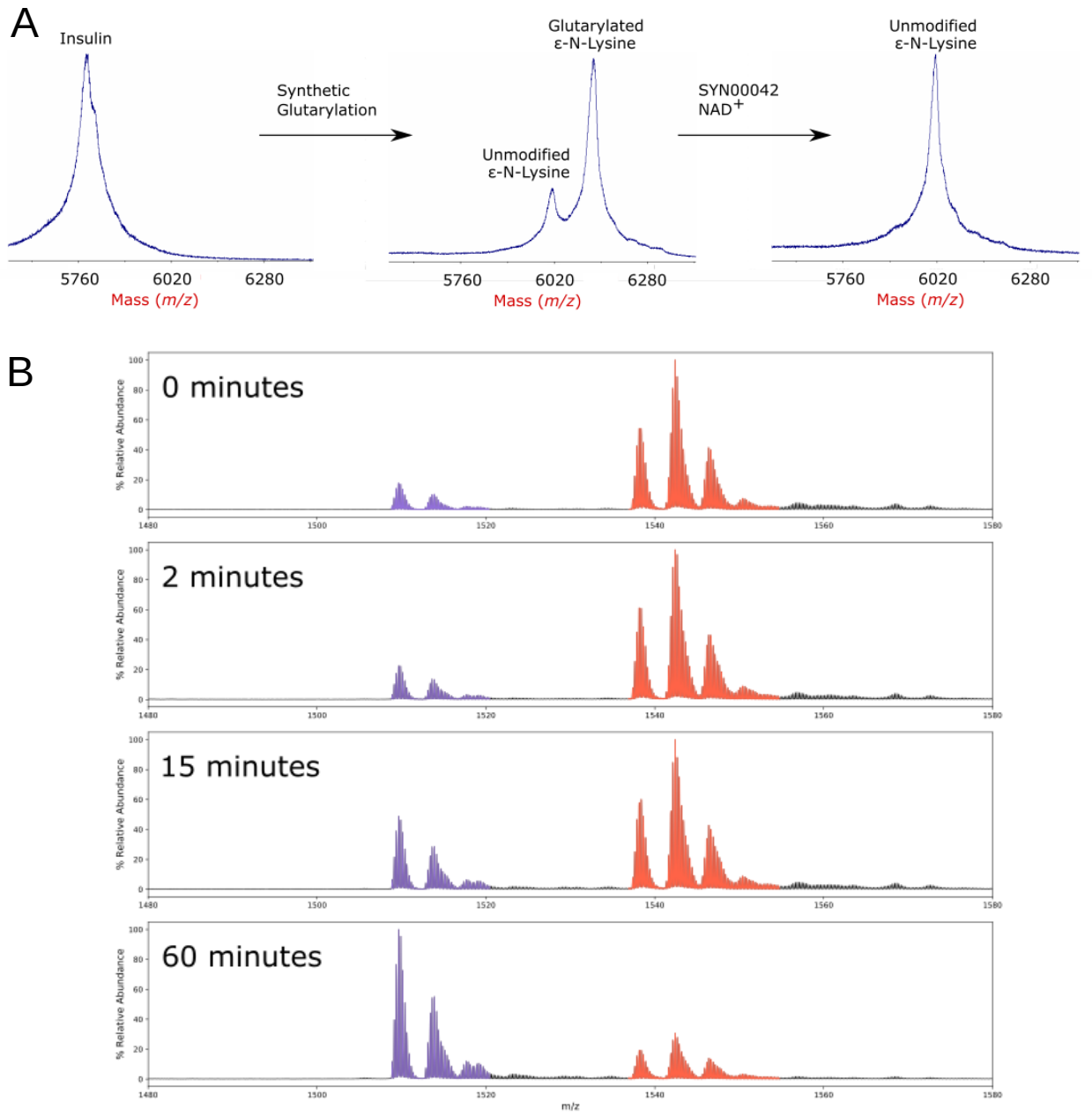


Figure 5. *S. aciditrophicus* sirtuin SYN_00042 shows *in vitro* deacylase activity. A) MALDI spectra of insulin and modified insulin protein. The peak corresponding to "Glutarylated N^{ε} -Lysine" corresponds with a mass shift of 3 glutaryl modifications (+342 Da), 2 at the N -terminal sites and 1 at lysine. B) ESI-TOF mass spectra of the 4+ charge states at 0, 2, 15 and 60 minutes after the addition of sirtuin. The peak clusters in purple on the left corresponds to insulin modified at the two N-termini, the red peak clusters on the right of insulin modified at all 3 sites. Ammonium adducts are also present in the spectra (+17 Da).

Supplemental Figure

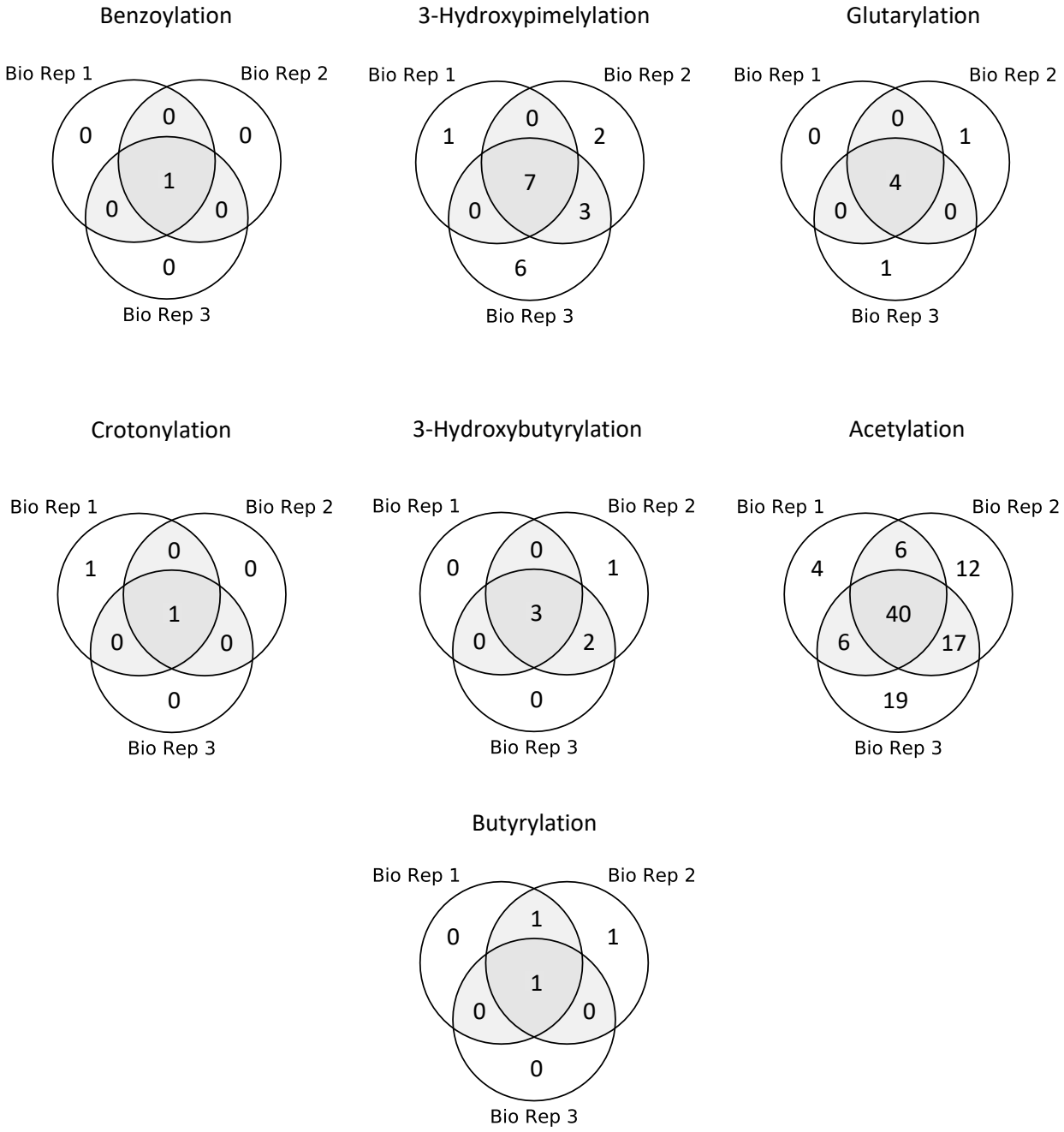


Figure S1. Acylations are reproducible across biological replicates. Venn diagrams showing the number of each acyl modification identified across biological replicates. Acylations only identified using the stepped collisional energy method are not in this figure, as only a single biological replicate was used for that analysis.

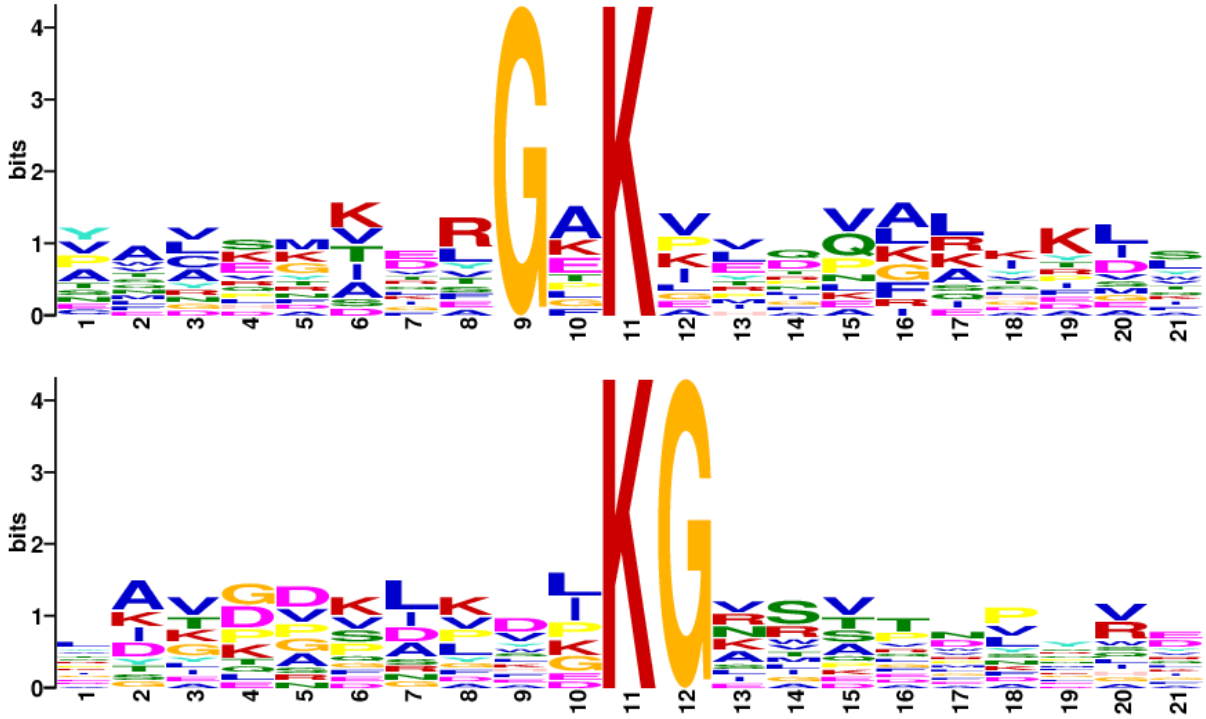


Figure S2. Sequence motifs of lysine acetylation. Motifs that were enriched from peptides containing acetyl-lysine were identified using the MoMo Modification Motifs program (Version 5.3.0) on the MEME suite platform (38).

BamA_*S. acidtrophicus* 6-oxocyclohex-1-ene-1-carbonyl-CoA hydrolase

```
1   MSLDWMPREHGLKNHSRHTEQWWGTEAPCTVYEKRPLKDPKGNVVPGLYSAWIRLNNPGQ
61  YNSYTTEMVKGVIAGFENSSTDREVVAVVFTGTGPNAFCTGGNTKEYSEYYSMRPEEYGS
121 YMELFNNMVDSILMCKKPVICRVNGMRVAGGQEIGTATDITVSSDLAIFGQAGPRHGSAP
181 VGGASDFLPWFLSIEDAMWNCVSCEMWSAYKMKAKNLISKALPVLKDDKGNWVRNPQVYT
241 DTYVKDGEIVYGEPKTGEEA**K**QARAWVNEKLKNNDYDFSLIDAEVDRIVWVFANLFPGCL
301 MKSIDGIRQKKKFWWDQIKNDHRYWLGTNMMGEAFLGFGAFNTKKITGKDTIDFIKNRQL
361 IAEGALVDEAFMEQVLGKPLAK
```

Figure S3. Sequence of BamA. The sequence of the heavily modified BamA protein, with the heavily modified K261 residue highlighted in **bold**.

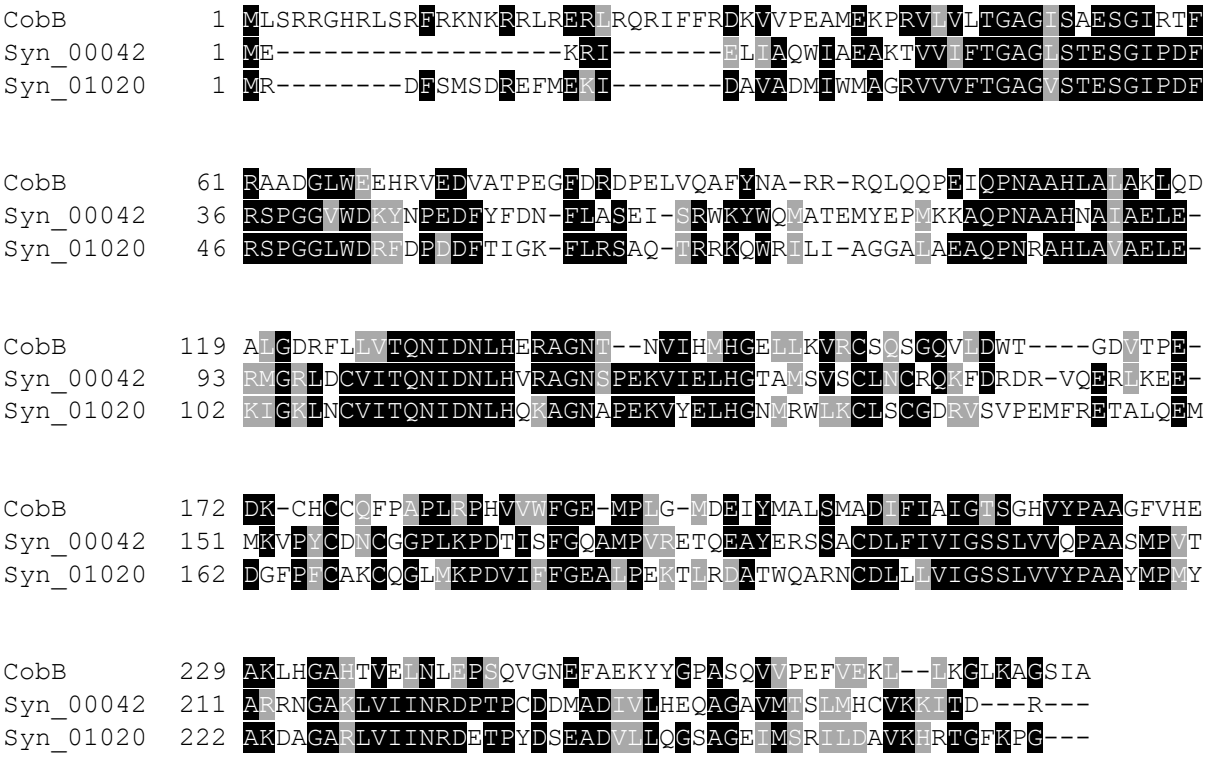


Figure S4. Sequence alignment of sirtuin homologs. The sirtuin homolog (*cobB*) of *Escherichia coli* was used to identify putative sirtuin homologs present in the *S. aciditrophicus*, Syn_00042 (31.43% identity) and Syn_01020 (29.73% identity).

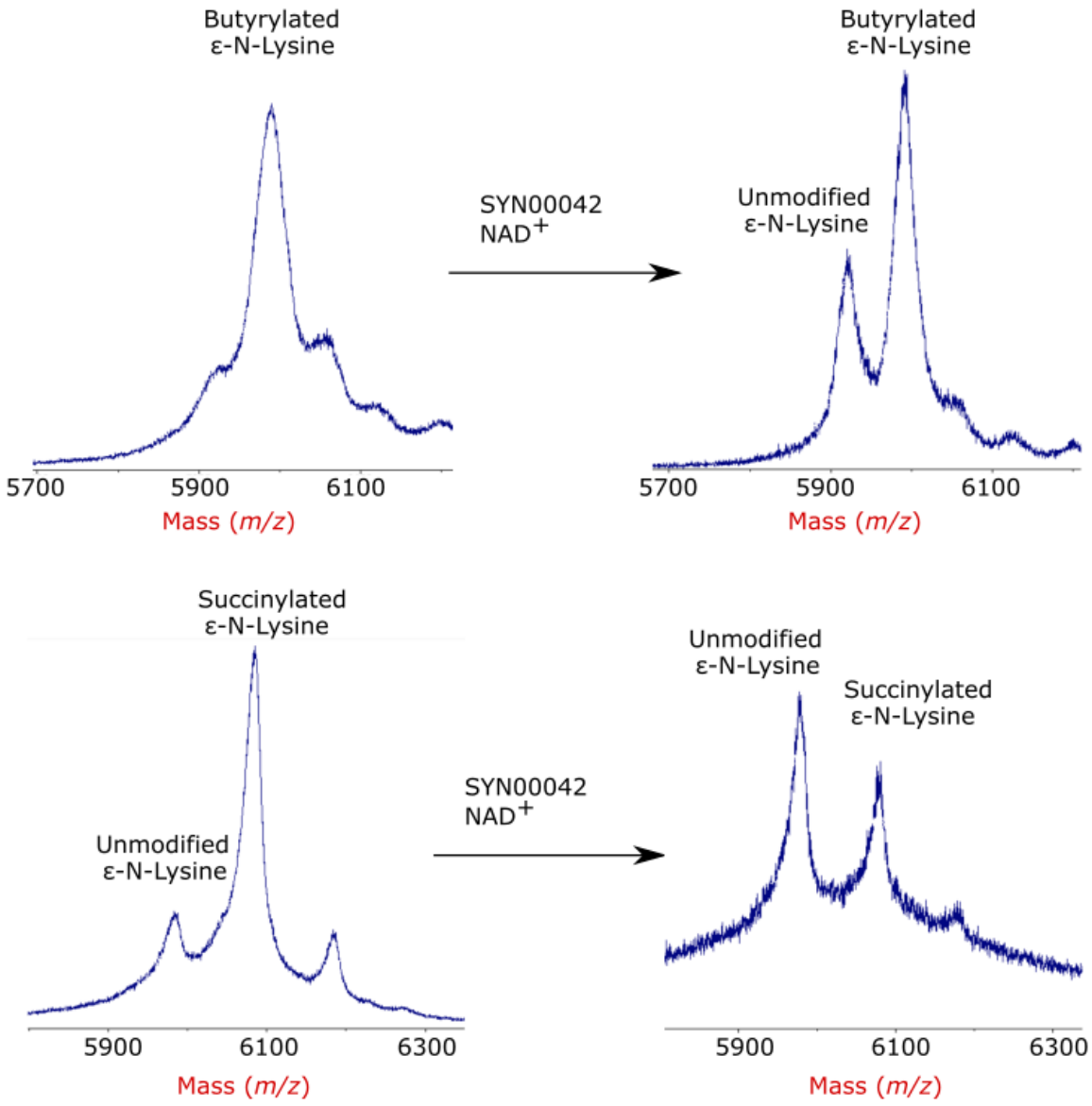


Figure S5. Syn_00042 activity on non-glutaryl acyl modifications. Sirtuin activity was assessed on acyl modifications other than glutaryl. The synthetically modified insulin with the indicated acyl-group was used as the substrate.

CHAPTER 4

Acylome of *Syntrophomonas wolfei* changes in response to altered growth substrates

4.1 Abstract

Syntrophomonas wolfei is an anaerobic microbe that degrades fatty acid chains that are of 4 to 8 carbons long. This process is described as being "thermodynamically challenging", as standard state conditions would not allow the reaction to progress forward. While the energy is limited in these systems, it is simultaneously true that the intermediates in these systems tend to include high abundances of reactive acyl-Coenzyme A species (RACS). Many systems have demonstrated a correlation between RACS as metabolites and acylations on lysine residues on metabolic proteins, though studies examining this in syntrophs remain scarce. We hypothesize that this will result in abundant acylations that correspond with the RACS present in *S wolfei's* fatty-acid degradation pathways. Here, we analyzed using liquid-chromatography tandem mass spectrometry *S. wolfei* cells grown on different carbon sources to identify the presence of lysine acylations. Modified proteins were first measured using shotgun proteomic sequencing methods. Modified enzymes within the fatty-acid degradation pathway were then targeted for analyses to allow for rigorous quantitative comparisons. It was found that acyl modifications related to RACS are present in the proteome of the bacteria. Further, we were able to demonstrate that the profile of the acyl-proteome was altered upon changing the carbon substrates of the cells.

4.2 Introduction

Microbial syntrophy is an important component of global carbon cycling that requires cooperation and crossfeeding between different species of microbes (1). Many of the intermediates generated during the degradation of larger biopolymers to smaller molecules, like hydrogen and methane, are aliphatic and aromatic acids. The degradation of these resulting intermediates to their smaller components that are consumed by methanogens is an endergonic process under steady state conditions (2). Under methanogenic conditions, the free energy

associated with degradation of these molecules is highly dependent on entropy (3). In effect, this means that the flux of compounds into or out of the system will play a large role in whether or not these reactions will proceed forward spontaneously. Given the propensity of methanogens to consume hydrogen and acetate towards the production of methane, the system exists with a great deal of efflux of these degradation products (3). The shift of equilibrium when methanogens are present alters the thermodynamics of the reaction such that it will occur spontaneously. The result is a necessary network of bacterial species that rely on one another to completely break down larger organic compounds. The bacterial class that is able to degrade the aromatic and aliphatic acids are called syntrophs and typically requires hydrogen consuming partners to grow (2).

Syntophomonas wolfei is a species of syntrophic bacteria that has been identified as a fatty acid degrading microbe (4, 5). Degradation of saturated fatty acids requires growth with a hydrogen consumer; typically the methanogen *Methanospirillum hungatei* (4, 6). However this species can also be grown on unsaturated fatty acids, such as crotonate, and can be cultivated as a pure culture (7, 8), thus allowing for comparisons on how these bacteria respond to changing metabolic conditions. This allows for the establishment of *S. wolfei* as a model organism with which to study anaerobic, syntrophic butyrate degraders. The importance of understanding degradation across a variety of short carbon moieties demands a thorough understanding of how these processes occur (1, 9–11).

S. wolfei utilizes the β-oxidation pathway to degrade short chain fatty acids, of which there are many paralogs at the genomic level (12). While the main constituents of the pathways have been identified, many questions remain. Evidence at the proteomic level appears to indicate that varying conditions do not result in significant changes in protein expression (13). Despite the

consistency of the enzymatic constituents of the cells, there has been evidence that different growth conditions may result in different enzymatic rates of catalysis (14, 15). Rather than the organism using transcriptional or translational regulation to alter enzyme activity, we sought to find if the enzymatic proteins themselves could be altered via the addition of post-translational modifications (PTMs) in a manner that correlates with changing environmental conditions.

PTMs are present in all kingdoms of life and have a plethora of physicochemical properties, structures, and by extension, functions (16). PTMs have long been attractive candidates for regulators of various cellular functions and metabolic pathways, as they can often reflect the intracellular conditions of the cell and directly alter proteins as a response. One class of such modifications is that of lysine acylation; a modification that has many diverse members in its family. Uniquely, lysine acylations can be either a result of enzyme-mediated reactions (17) or by spontaneous reactions with reactive metabolites known as reactive-acyl Coenzyme A species (RACS) (18–20). Lysine acylation has been shown to play a significant role in the physiology of bacterial species, particularly with respect to their metabolic processes (21, 22). Acylation can both impede enzymatic activity, as well as initiate enzyme activity (23, 24), presenting a theoretical mechanism of enzymatic regulation that may be conserved throughout many organisms. Both the presence of acylations throughout domains of life (21, 25), as well as conserved mechanisms of acyl-regulation (26, 27) indicate the significance of understanding the role acyl modifications in bacteria.

Here we take a systems level approach to identify acyl-modifications in the *S. wolfei* proteome. We demonstrate that there are different global lysine acylome profiles that are associated with different growth conditions of the cells, both qualitatively with respect to the modifications present, as well as quantitatively with respect to the abundance of different acyl

modifications. We identified that the enzymes involved in the β-oxidation pathway that degrade short chain acylations are heavily modified with an abundance of different acylations, and that even specific sites on these enzymes are altered under different conditions. Following global acylome analysis, we targeted specific modifications that are present in β-oxidation pathway enzymes to quantify the changes. Indeed, we demonstrate that changes in abundance of modified sites are present and reflect the expected fluctuations of the abundance of RACS under different conditions.

4.3 Experimental

4.3.1 Culturing of Cells

Syntrophomonas wolfei was grown by the McInerney group at The University of Oklahoma, either axenically on crotonate (4), or in the presence of a methanogenic partner, *Methanospirillum hungatei*, on both crotonate or butyrate as has been previously described (6).

4.3.2 Sample Preparation

Cells were harvested from cultures into 4.0% v/v ammonium lauryl sulfate, 0.1% w/v sodium deoxycholate, and 5 mM tris(2-carboxyethyl)phosphine in 100 mM ammonium bicarbonate. Peptides were prepared using enhanced filter assisted sample preparation as described by Erde et al (28, 29). Briefly, the lysate was exchanged into buffer composed of 8 M urea, 0.1% w/v sodium deoxycholate, and 0.1% w/v *n*-octyl glucoside with a 10 kDa Microcon ultrafiltration unit (Millipore). Proteins were alkylated with 17mM iodoacetamide and digested overnight with a 1:100 ratio of trypsin:protein at 37^{O}C. Resulting peptides were then extracted from the detergents in the buffer using ethyl acetate extraction. After extraction, the peptides were dried, acidified with 0.1% acetic acid, and desalted with STAGE tips using 3M Empore C18 Solid

Phase Extraction Disks and dried again. The resulting peptides were resuspended in LC-MS injection buffer consisting of 3% acetonitrile and 0.1% formic acid.

4.3.3 Shotgun Experiments on the Orbitrap Exploris 480

Resulting peptide samples were measured by Thermo Fisher (San Jose) and analyzed on the Orbitrap Exploris 480 Mass Spectrometer using reversed phase liquid chromatography-tandem mass spectrometry (LC-MS/MS). Peptides were separated with an EASY nLC-1200 (Thermo Scientific) before being loaded onto the trap mass spectrometer. Peptides were separated by a 1.6 μm C18 ultra high-performance liquid chromatography column (75 μm x 250 mm, Ionopticks^{TM} Aurora^{TM}). Buffer A (0.1% formic acid) and buffer B (0.1% formic acid in 80% acetonitrile) were used for HPLC separation at a flow rate of 300 nL/min with the following gradient: 3-4% B in 1 minute, 4-18% B in 72 minutes, 18-28% B in 28 minutes, 28-37% B in 18 minutes, and 37-98% B in 3 minutes.

Data-dependent acquisition (DDA) mode was used to select ions in the scan range of *m/z* 350-1200 for fragmentation. MS^1 scans were obtained at a resolution of 60,000 with positive polarity and a normalized automatic gain control (AGC) target of 300%. Putatively peptide ions were fragmented using higher-energy collisional dissociation (HCD) at a normalized collisional energy of 28%. Dynamic exclusion was set to 45 sec and at a mass window of ±10 ppm. MS^2 scans were collected with a first fixed mass of 120 Th, isolation window of 1.6, orbitrap resolution of 15,000, and normalized AGC target of 75%.

4.3.4 Hydrophilic Interaction Liquid Chromatography Library Building

Peptides resulting from the eFASP procedure were separated off-line via hydrophilic interaction chromatography (HILIC) to increase depth and establish a library for the parallel reaction monitoring (PRM) experiments to follow. 50 μg of peptide were loaded onto a

BioPureSPN MACRO PolyHYDROXYETHYL A (The Nest Group) column. Peptides were bound to the column using a mobile phase consisting of 90% acetonitrile and 150 mM ammonium formate, pH 3. The peptides were then sequentially eluted into six fractions that used the following organic fraction of mobile phase elution buffers: 80% acetonitrile, 78% acetonitrile, 74% acetonitrile, 71% acetonitrile, 40% acetonitrile, and 35.5% acetonitrile. Fractions 1 and 6 were combined for analysis. Peptides were desalted with STAGE tips as previously described.

Resulting peptides were measured via LC-MS/MS using a QExactive mass spectrometer (Thermo Fisher Scientific, San Jose, CA, USA). Peptides (200 ng) were separated using an EASY nLC1000 (Thermo Scientific) loaded onto an Acclaim PepMap100 C18 trap column (Thermo Scientific, Product #16-494-6, 75 μm × 2 cm, 100 Å) followed by separation on an Acclaim PepMap RSLC C18 analytical column (Thermo Scientific, Product #03-251-873, 75 μm × 25 cm, 100 Å). Buffer A (0.1% formic acid) and buffer B (0.1% formic acid in 100% acetonitrile) were implemented at 300 nL/min with the gradient of 3-20% B in 62 minutes, 20-30% B in 31 minutes, 30-50% B in 5 minutes, and 50-80% B in 2 minutes.

The QExactive quadrupole-orbitrap mass spectrometer was used in DDA mode with HCD fragmentation. MS scans in a range of *m/z* 300-1800 were acquired at 70,000 resolution, with an AGC target of 1E6 and a maximum fill time of 100 ms. The 10 most abundant precursor ions were dissociated sequentially using a normalized collisional energy of 27, and MS/MS spectra were acquired at 17,500 resolution with an AGC target of 1E5 at a maximum fill time of 80 ms.

4.3.5 Targeted Acylation Quantification

Targeted analysis by parallel reaction monitoring (PRM) (30) was utilized to quantify peptides of interest using the peptide library determined by HILIC DDA analysis. Liquid

chromatography was performed identical to the HILIC library building step. The inclusion list used for PRMs is shown in **Table 1**. The MS2 scans were obtained at a resolution of 17,500 with an AGC target of 2e5 and an isolation window of *m/z* 4.0 using HCD fragmentation with a 27 NCE.

4.3.6 Data Analysis

RAW data files from DDA experiments were analyzed using ProteomeDiscoverer (version 1.4) and the protein database was searched with the Mascot search algorithm (Matrix Science) (31). A concatenated database composed of UniProt *S. wolfei* and *M. hungatei* protein sequences (as of February 1, 2017) and the sequences of common contaminants was searched. Parameters for the Mascot database search were: enzyme name, trypsin; maximum missed cleavage sites, 2; precursor mass tolerance, 10 ppm; fragment mass tolerance, 0.02 Da; and variable methionine oxidation and cysteine carbamidomethylation. Searches also considered variable acyl modifications on lysines of acetoacetyl (+84.02113), acetyl (+42.01056), crotonyl (+68.02621), 3-hydroxybutyryl (+86.03678), and butyryl (+70.04186). For HILIC searches, the results were imported into Skyline to build libraries for the targeted analyses of PRM experiments.

PRM RAW files were imported into Skyline. Cyclized immonium ions (32) for the acyl-lysine modifications of interest were added as customized ions to the Skyline analysis workflow. The top four to six coeluting product ion extracted ion chromatograms were selected to determine the peak area abundance for acylated peptide ions. As an internal control, each sample was normalized to one to three peptides selected from a control ribosomal protein, Swol_1491 (FGPGDTVK, LQVFEGTVIK, and GTGLSQTFTVR) that were identified in a given run. To compare across different conditions that may contain different protein concentrations, peptides

from butyrate cocultures were multiplied by the ratio of protein abundance in butyrate coculture to crotonate monoculture (**Table 2**).

4.4 Results and Discussion

Cells grown under different conditions were analyzed first via shotgun mass spectrometry to obtain an overview of the proteome. Proteome differences found in these conditions are displayed in **Figure 1**. Many of the trends that were expected were identified. Acetylation is present in the largest overall numbers in the crotonate grown samples. Butyrylation was identified most often in cells grown on butyrate. 3-hydroxybutyrlation seem to be present more in cells grown on crotonate monocultures. These trends seem to indicate that the points in the pathways that have the largest positive free energy values are associated with the production of more associated acylations. Notably, several of the acyl-lysine modifications identified are present in all of the growth conditions probed. This is especially interesting for butyrate, an intermediate that is "upstream" of crotonate in butyrate oxidation. This implies that the pathway, under certain conditions, may be reversible; a phenomenon seen in other syntrophs (33).

Many of the PTMs identified were found to be present on enzymes involved in the butyrate degradation pathway (**Figure 2A**). Butyrate degradation enzymes consist of multiple paralogs for each reaction that are simultaneously expressed in *S. wolfei* (12). As such, the catalytic reactions are grouped by the reaction catalyzed to identify trends in the amount of different acyl-lysines identified under growth conditions of different substrates. The results displayed in **Figure 2B-C** indicate that the carbon substrate quantitatively impacts the modified sites within the acylome of the syntroph. Butyryl modifications are in much higher abundance in syntrophs grow on butyrate as a carbon source, whereas acetylation is typically in greater abundance in the species grown on

crotonate. This is to be expected, as acylation is often considered in microbes to be a function of RACS buildup and local concentration (18, 34). The removal of electrons from butyrate, the first step of butyrate oxidation, is an energetically challenging step to overcome, meaning this metabolite is likely to build at this point. Bypassing this step allows for more acetyl-CoA and acetyl-phosphate, resulting in the production of more of modifications associated with these reactive metabolites. Interestingly, the enzymes that tend to be more heavily butyrylated when grown on butyrate are earlier in the pathway, and those with an increase of acetylation when grown on crotonate tend to be later in the pathway, supporting the role of reactive metabolite localization in modification. The abundance changes of these modifications are of particular note in this pathway, as carbon source has been shown to impact the enzymatic rate at reactions in the pathway are catalyzed (14, 15).

Several of the proteins in the pathway also showed signs of extreme heterogeneity of proteoforms due to lysine acylation. Two proteins of interest from this pathway were Swol_2030, a 3-hydroxybutyryl-CoA dehydrogenase, and Swol_2051, an acetyl-CoA acetyltransferase. Swol_2030 has nine different lysine sites of modification, and Swol_2051 has eight. Further, both of these proteins demonstrate the presence of four different acyl-lysine modifications. Limitations innate to bottom-up proteomics make it very difficult to assess how these modifications are in combination with one another throughout the protein. Stoichiometries between modified and unmodified residues are also difficult to assess owing to differential cleavage of modified and unmodified lysine residues by trypsin, as well as differences in peptide ionization efficiency. However, it is possible to compare the profiles of modified sites across different conditions. On both of these proteins, there was a single residue that displayed four different acylations under the three conditions on which *S. wolfei* was grown. Site occupancy

was considered to demonstrate a shift in acylation in response to changing conditions the cells were grown (**Figure 3**). While the sites cannot be compared to one another due to the innate differences with protein expression and peptide ionization, modifications can be reasonably compared across different conditions upon identification. Site occupancy of two sites on the select enzymatic paralogs indicate that the modifications do shift with changing conditions of the cell. This phenomenon is particular drastic when comparing the cells grown on crotonate to those grown on butyrate, which have a greater proportion of lysine residue occupied by butyryl groups and no identified residue occupied by 3-hydroxybutyryl (**Figure 3**). Changes in site occupancy ratio still require further consideration, as it is possible that total stoichiometries differ across conditions, meaning that modifications are not being traded-off for one another, but rather that the rates at which modification are occurring differ as conditions change.

To further interrogate the quantitative aspect of these modifications, the cells were fractionated by HILIC fractionation to get better depth and breadth than what was previously identified. From this fractionation, a list of ions was generated that included identified acyl-lysine peptides that were both present in enzymes related to the β-oxidation pathway, and that had more than one type of acylation, thus allowing for a site occupancy comparison. The library of selected ions and peptides is shown in **Table 1**. One of the key challenges of comparing these microbial consortia is the difference in the number of bacteria within the population, as growth conditions may impact the species makeup. To overcome this limitation, as well as biological variability and potential bias in the sample preparation, three peptides from a large ribosomal submit were chosen as housekeeping proteins to act as internal standards by which modified peptides can be compared for relative quantitation. Peptides chosen did not always elute at consistent retention times (**Figure 4A-B**), likely caused by matrix effects of different sample

compositions (35), resulting in crotonate monoculture samples where only two, or in one case one, of the three peptides identified. In cases all three peptides were not identified, using one or two that were could generate reasonable estimates (**Figure 4C**). As the communities contain different constituents, the amount of each protein present must be controlled for to allow a quantitative comparison. To normalize the analyses across conditions, DDA-determined protein abundances were used to determine the different ratios in which proteins were present in each sample. The ratios at the protein level were used to compare quantities of peptides across the different samples for proteins in which MS^2-level data was collected (**Table 2**).

MS^2 level quantitative analysis of these peptides is demonstrated in **Figure 5** to compare peptide abundances across different conditions. **Figure 5A-B** demonstrate the utility of using MS^2 level analysis to specifically target certain ions in complex samples, something needed given the low stoichiometry of the modified peptides in the sample. Because of previously reported specificity and sensitivity of immonium ion derivatives (32, 36, 37), these ions were incorporated into quantitation (**Figure 5C**). Incorporation of these ions also allowed for high confidence of which peaks identified were associated with modified peptides; something that is not possible with MS^1 analysis alone.

Modifications as identified through the PRMs that are impactful to the β-oxidation pathway are demonstrated in **Figure 6**. The first point of note is that some of the identified acylations were significantly different with respect to acylations identified under different carbon sources in some cases and not in others, thus allowing a starting point with which sites that are of interest may be identified. To best determine which modifications are of interest, several changes are shown throughout the β-oxidation pathway (**Figure 6A-E**). Interestingly, while the data is complex, there are some trends that can be teased out. Butyrylation is identified as significantly

enriched in 5 of 8 peptides when cells are grown on the butyrate substrate. None of the 8 peptides were enriched on crotonate growth conditions. Two of the three 3-hydroxybutyrylation modifications were significantly enriched on the crotonate substrate, as were 2 of the 4 acetylation that demonstrated a significant difference between conditions. Interestingly, the single residue that seemingly goes against these trends is the K238 residue on the acetyl-CoA acetyltransferase Swol_2051 (**Figure 6E**), in which all modified residues are enriched in the butyrate coculture condition. While these sites are highly conserved and proximal to the CoA binding site in a related species, *Clostridium acetobutylicum*, there is no evidence of direct interactions between these residues and any substrates (38).

4.5 Conclusion

The breadth and scope of acyl-modifications is rapidly expanding. A family of modifications that was first identified as an acetyl group on histones in the 1960s has grown to include many types of modifications on thousands of proteins across all domains of life (25). Despite growing interest in these modifications, there still remain many questions as to both causes of these modifications as well as the effects they produce. While many labs are engaged in the search for these modifications in relatively simple systems such as histones (39–41), leveraging the unique metabolic properties and features of bacteria has been largely ignored, leaving it a fertile ground for analysis. Here we demonstrate that through informed analysis of acylations based on metabolic pathways, we can identify the acylations that will modify lysine residues on proteins. This information can be used to streamline analysis of acylations in different systems to help identify putative modifications.

Further, while the scope of acylations has been broadening, a thorough understanding of their biological impact is not yet known. By altering the conditions of substrates that *S. wolfei* can oxidize, we have demonstrated a correlation between RACS present and lysine acylations identified, strengthening the evidence for non-enzymatic acylation and supporting the "carbon stress model" as a proximate explanation for lysine acylation (18, 19, 34). Drawing a correlation to previous works has also opened a new avenue to understand the role of lysine acylations. That altering the conditions of cell growth changes both the enzymatic activities of butyrate degradation enzymes and the acylation profiles on the protein can give targets that can be further studied to identify a causative relationship. Notably, many of these enzymes have been identified as members of complexes (42), which have been demonstrated to be altered by lysine acylations in other systems (43–45). While many of these acylations are structurally similar, differing in some cases by only a methyl or hydroxyl group, these slight differences have been demonstrated as sufficient for significant alterations for enzymes using acyl-lysine as a substrate (46). Ultimately, the studies performed on the *S. wolfei* acyl-proteome provide a roadmap to understanding the impact of the acylations on enzymatic activity.

4.6 References

1. McInerney, M. J., Sieber, J. R., and Gunsalus, R. P. (2009) Syntrophy in anaerobic global carbon cycles. *Curr. Opin. Biotechnol.* **20**, 623–632
2. Schink, B. (1997) Energetics of syntrophic cooperation in methanogenic degradation. *Microbiol. Mol. Biol. Rev.* **61**, 262–80
3. McInerney, M. J., and Beaty, P. S. (1988) Anaerobic community structure from a

nonequilibrium thermodynamic perspective. *Can. J. Microbiol.* **34**, 487–493

4. McInerney, M. J., Bryant, M. P., Hespell, R. B., and Costerton, J. W. (1981) Syntrophomonas wolfei gen. nov. sp. nov., an Anaerobic, Syntrophic, Fatty Acid-Oxidizing Bacterium. *Appl. Environ. Microbiol.* **41**, 1029–39
5. Lorowitz, W. H., Zhao, H., and Bryant, M. P. (1989) Syntrophomonas wolfei subsp. saponavida subsp. nov., a Long-Chain Fatty-Acid-Degrading, Anaerobic, Syntrophic Bacterium; Syntrophomonas wolfei subsp. wolfei subsp. nov.; and Emended Descriptions of the Genus and Species. *Int. J. Syst. Bacteriol.* **39**, 122–126
6. McInerney, M. J., Bryant, M. P., and Pfennig, N. (1979) Anaerobic bacterium that degrades fatty acids in syntrophic association with methanogens. *Arch. Microbiol.* **122**, 129–135
7. Beaty, P. S., and McInerney, M. J. (1987) Growth of Syntrophomonas wolfei in pure culture on crotonate. *Arch. Microbiol.* **147**, 389–393
8. Amos, D. A., and McInerney, M. J. (1990) Growth of Synthrophomonas wolfei on unsaturated short chain fatty acids. *Arch. Microbiol.* **154**, 31–36
9. Stams, A. J. M., Sousa, D. Z., Kleerebezem, R., and Plugge, C. M. (2012) Role of syntrophic microbial communities in high-rate methanogenic bioreactors. *Water Sci. Technol.* **66**, 352–362
10. Boll, M., Geiger, R., Junghare, M., and Schink, B. (2020) Microbial degradation of phthalates: biochemistry and environmental implications. *Environ. Microbiol. Rep.* **12**, 3–15
11. Amos, D. A., and McInerney, M. J. (1989) Poly-β-hydroxyalkanoate in Syntrophomonas wolfei. *Arch. Microbiol.* **152**, 172–177

12. Sieber, J. R., Sims, D. R., Han, C., Kim, E., Lykidis, A., Lapidus, A. L., McDonnald, E., Rohlin, L., Culley, D. E., Gunsalus, R., and McInerney, M. J. (2010) The genome of Syntrophomonas wolfei: new insights into syntrophic metabolism and biohydrogen production. *Environ. Microbiol.* **12**, no-no

13. Sieber, J. R., Crable, B. R., Sheik, C. S., Hurst, G. B., Rohlin, L., Gunsalus, R. P., and McInerney, M. J. (2015) Proteomic analysis reveals metabolic and regulatory systems involved in the syntrophic and axenic lifestyle of Syntrophomonas wolfei. *Front. Microbiol.* **6**, 115

14. McInerney, M. J., and Wofford, N. Q. (1992) Enzymes involved in crotonate metabolism in Syntrophomonas wolfei. *Arch. Microbiol.* **158**, 344–349

15. Wofford, N. Q., Beaty, P. S., and McInerney, M. J. (1986) Preparation of cell-free extracts and the enzymes involved in fatty acid metabolism in Syntrophomonas wolfei. *J. Bacteriol.* **167**, 179–185

16. Walsh, G., and Jefferis, R. (2006) Post-translational modifications in the context of therapeutic proteins. *Nat. Biotechnol.* **24**, 1241–1252

17. Ali, I., Conrad, R. J., Verdin, E., and Ott, M. (2018) Lysine Acetylation Goes Global: From Epigenetics to Metabolism and Therapeutics. *Chem. Rev.* **118**, 1216–1252

18. Wagner, G. R., and Hirschey, M. D. (2014) Nonenzymatic Protein Acylation as a Carbon Stress Regulated by Sirtuin Deacylases. *Mol. Cell*. **54**, 5–16

19. Wagner, G. R., and Payne, R. M. (2013) Widespread and enzyme-independent Nε-acetylation and Nε-succinylation of proteins in the chemical conditions of the mitochondrial matrix. *J. Biol. Chem.* **288**, 29036–29045

20. Baeza, J., Smallegan, M. J., and Denu, J. M. (2015) Site-Specific Reactivity of

Nonenzymatic Lysine Acetylation. *ACS Chem. Biol.* **10**, 122–128

21. VanDrisse, C. M., and Escalante-Semerena, J. C. (2019) Protein Acetylation in Bacteria. *Annu. Rev. Microbiol.* **73**, 111–132

22. Christensen, D. G., Xie, X., Basisty, N., Byrnes, J., McSweeney, S., Schilling, B., and Wolfe, A. J. (2019) Post-translational Protein Acetylation: An elegant mechanism for bacteria to dynamically regulate metabolic functions. *Front. Microbiol.* **10**, 1604

23. Gardner, J. G., Grundy, F. J., Henkin, T. M., and Escalante-Semerena, J. C. (2006) Control of Acetyl-Coenzyme A Synthetase (AcsA) Activity by Acetylation/Deacetylation without NAD+ Involvement in Bacillus subtilis. *J. Bacteriol.* **188**, 5460–5468

24. Garrity, J., Gardner, J. G., Hawse, W., Wolberger, C., and Escalante-Semerena, J. C. (2007) N-lysine propionylation controls the activity of propionyl-CoA synthetase. *J. Biol. Chem.* **282**, 30239–30245

25. Verdin, E., and Ott, M. (2015) 50 years of protein acetylation: From gene regulation to epigenetics, metabolism and beyond. *Nat. Rev. Mol. Cell Biol.* **16**, 258–264

26. Sanders, B. D., Jackson, B., and Marmorstein, R. (2010) Structural basis for sirtuin function: what we know and what we don't. *Biochim. Biophys. Acta*. **1804**, 1604–16

27. Greiss, S., and Gartner, A. (2009) Sirtuin/Sir2 phylogeny, evolutionary considerations and structural conservation. *Mol. Cells*. **28**, 407–415

28. Erde, J., Loo, R. R. O., and Loo, J. A. (2014) Enhanced FASP (eFASP) to increase proteome coverage and sample recovery for quantitative proteomic experiments. *J. Proteome Res.* **13**, 1885–1895

29. Erde, J., Loo, R. R. O., and Loo, J. A. (2017) Improving proteome coverage and sample recovery with enhanced FASP (eFASP) for quantitative proteomic experiments. in

Methods in Molecular Biology, pp. 11–18, Humana Press Inc., **1550**, 11–18

30. Peterson, A. C., Russell, J. D., Bailey, D. J., Westphall, M. S., and Coon, J. J. (2012) Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics. *Mol. Cell. Proteomics*. **11**, 1475–88

31. Perkins, D. N., Pappin, D. J. C., Creasy, D. M., and Cottrell, J. S. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. in *Electrophoresis*, pp. 3551–3567, Wiley-VCH Verlag, **20**, 3551–3567

32. Muroski, J. M., Fu, J. Y., Nguyen, H. H., Ogorzalek Loo, R. R., and Loo, J. A. (2021) Leveraging Immonium Ions for Targeting Acyl-Lysine Modifications in Proteomic Datasets. *Proteomics*. **21**, 2000111

33. James, K. L., Kung, J. W., Crable, B. R., Mouttaki, H., Sieber, J. R., Nguyen, H. H., Yang, Y., Xie, Y., Erde, J., Wofford, N. Q., Karr, E. A., Loo, J. A., Ogorzalek Loo, R. R., Gunsalus, R. P., and McInerney, M. J. (2019) Syntrophus aciditrophicus uses the same enzymes in a reversible manner to degrade and synthesize aromatic and alicyclic acids. *Environ. Microbiol.* **21**, 1833–1846

34. Trub, A. G., and Hirschey, M. D. (2018) Reactive Acyl-CoA Species Modify Proteins and Induce Carbon Stress. *Trends Biochem. Sci.* **43**, 369–379

35. Fang, N., Yu, S., Ronis, M. J. J., and Badger, T. M. (2015) Matrix effects break the LC behavior rule for analytes in LC-MS/MS analysis of biological samples. *Exp. Biol. Med.* **240**, 488–497

36. Trelle, M. B., and Jensen, O. N. (2008) Utility of immonium ions for assignment of ε-N-acetyllysine-containing peptides by tandem mass spectrometry. *Anal. Chem.* **80**, 3422–3430

37. Zolg, D. P., Wilhelm, M., Schmidt, T., Médard, G., Zerweck, J., Knaute, T., Wenschuh, H., Reimer, U., Schnatbaum, K., and Kuster, B. (2018) ProteomeTools: Systematic Characterization of 21 Post-translational Protein Modifications by Liquid Chromatography Tandem Mass Spectrometry (LC-MS/MS) Using Synthetic Peptides. *Mol. Cell. Proteomics*. **17**, 1850–1863

38. Kim, S., Jang, Y. S., Ha, S. C., Ahn, J. W., Kim, E. J., Hong Lim, J., Cho, C., Shin Ryu, Y., Kuk Lee, S., Lee, S. Y., and Kim, K. J. (2015) Redox-switch regulatory mechanism of thiolase from Clostridium acetobutylicum. *Nat. Commun.* 10.1038/ncomms9410

39. Zhang, D., Tang, Z., Huang, H., Zhou, G., Cui, C., Weng, Y., Liu, W., Kim, S., Lee, S., Perez-Neut, M., Ding, J., Czyz, D., Hu, R., Ye, Z., He, M., Zheng, Y. G., Shuman, H. A., Dai, L., Ren, B., Roeder, R. G., Becker, L., and Zhao, Y. (2019) Metabolic regulation of gene expression by histone lactylation. *Nature*. **574**, 575–580

40. Huang, H., Zhang, D., Wang, Y., Perez-Neut, M., Han, Z., Zheng, Y. G., Hao, Q., and Zhao, Y. (2018) Lysine benzoylation is a histone mark regulated by SIRT2. *Nat. Commun.* **9**, 1–11

41. Xie, Z., Zhang, D., Chung, D., Tang, Z., Huang, H., Dai, L., Qi, S., Li, J., Colak, G., Chen, Y., Xia, C., Peng, C., Ruan, H., Kirkey, M., Wang, D., Jensen, L. M., Kwon, O. K., Lee, S., Pletcher, S. D., Tan, M., Lombard, D. B., White, K. P., Zhao, H., Li, J., Roeder, R. G., Yang, X., and Zhao, Y. (2016) Metabolic Regulation of Gene Expression by Histone Lysine β-Hydroxybutyrylation. *Mol. Cell*. **62**, 194–206

42. Crable, B. R., Sieber, J. R., Mao, X., Alvarez-Cohen, L., Gunsalus, R., Loo, R. R. O., Nguyen, H., and McInerney, M. J. (2016) Membrane complexes of Syntrophomonas wolfei involved in syntrophic butyrate degradation and hydrogen formation. *Front.*

Microbiol. **7**, 1795

43. El Kennani, S., Crespo, M., Govin, J., and Pflieger, D. (2018) Proteomic analysis of histone variants and their PTMs: Strategies and pitfalls. *Proteomes*. 10.3390/proteomes6030029

44. Lin, C., Zeng, H., Lu, J., Xie, Z., Sun, W., Luo, C., Ding, J., Yuan, S., Geng, M., and Huang, M. (2015) Acetylation at lysine 71 inactivates superoxide dismutase 1 and sensitizes cancer cells to genotoxic agents. *Oncotarget*. **6**, 20578–20591

45. Zhu, Y., Zou, X., Dean, A. E., Brien, J. O., Gao, Y., Tran, E. L., Park, S. H., Liu, G., Kieffer, M. B., Jiang, H., Stauffer, M. E., Hart, R., Quan, S., Satchell, K. J. F., Horikoshi, N., Bonini, M., and Gius, D. (2019) Lysine 68 acetylation directs MnSOD as a tetrameric detoxification complex versus a monomeric tumor promoter. *Nat. Commun.* 10.1038/s41467-019-10352-4

46. McClure, J. J., Inks, E. S., Zhang, C., Peterson, Y. K., Li, J., Chundru, K., Lee, B., Buchanan, A., Miao, S., and Chou, C. J. (2017) Comparison of the Deacylase and Deacetylase Activity of Zinc-Dependent HDACs. *ACS Chem. Biol.* **12**, 1644–1655

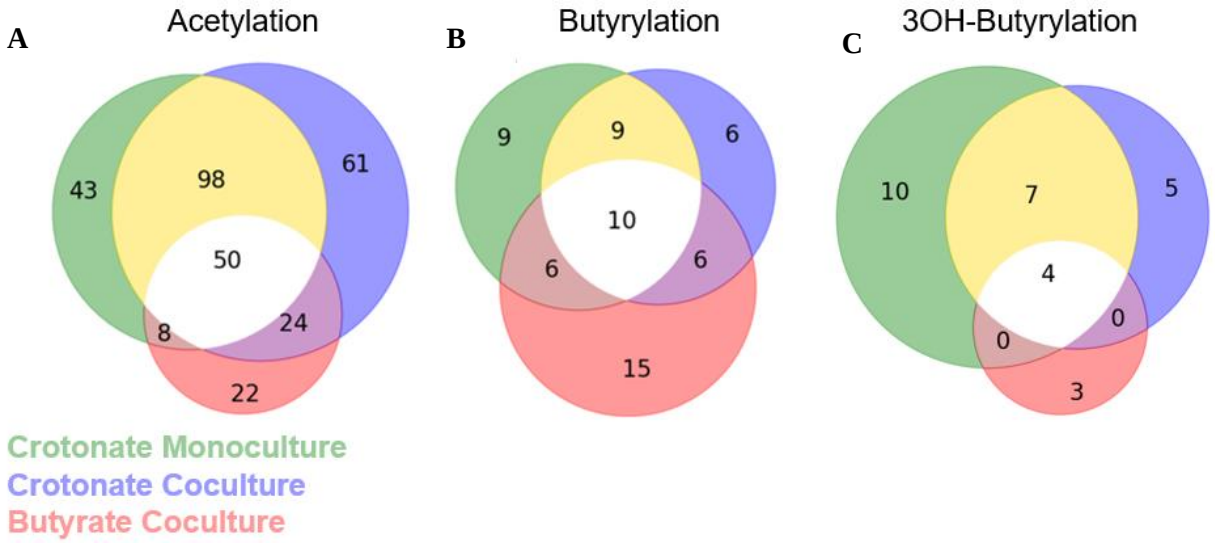


Figure 1. Profiles of lysine acylation in different growth conditions: The count of peptides identified with different types of acylations are identified under different conditions.

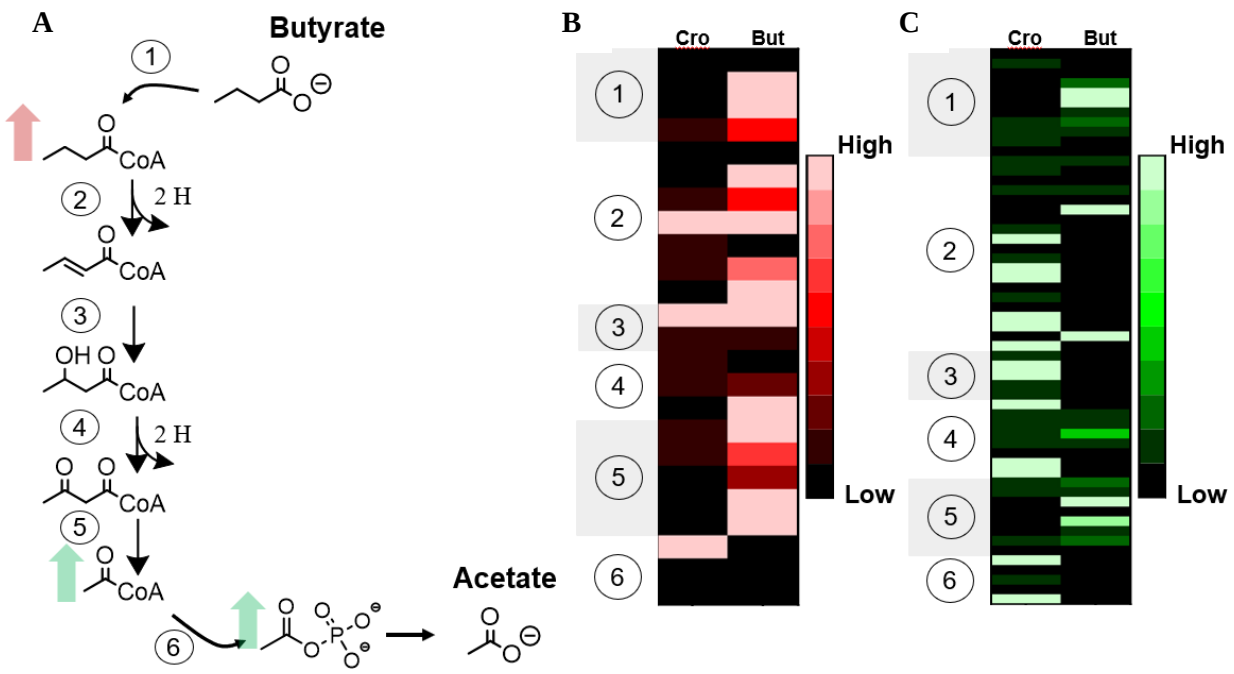


Figure 2. Quantitative shifts in lysine acylation in the β-oxidation pathway: Heatmaps demonstrate the abundance profile changes of modified peptides of peptides found in the β-oxidation pathway. **2A** identifies the pathway with key enzymatic reactions labeled numerically. **2B-C** identify modifications the abundance of modified peptides for both butyrylation (**2B**) and acetylation(**2C**).

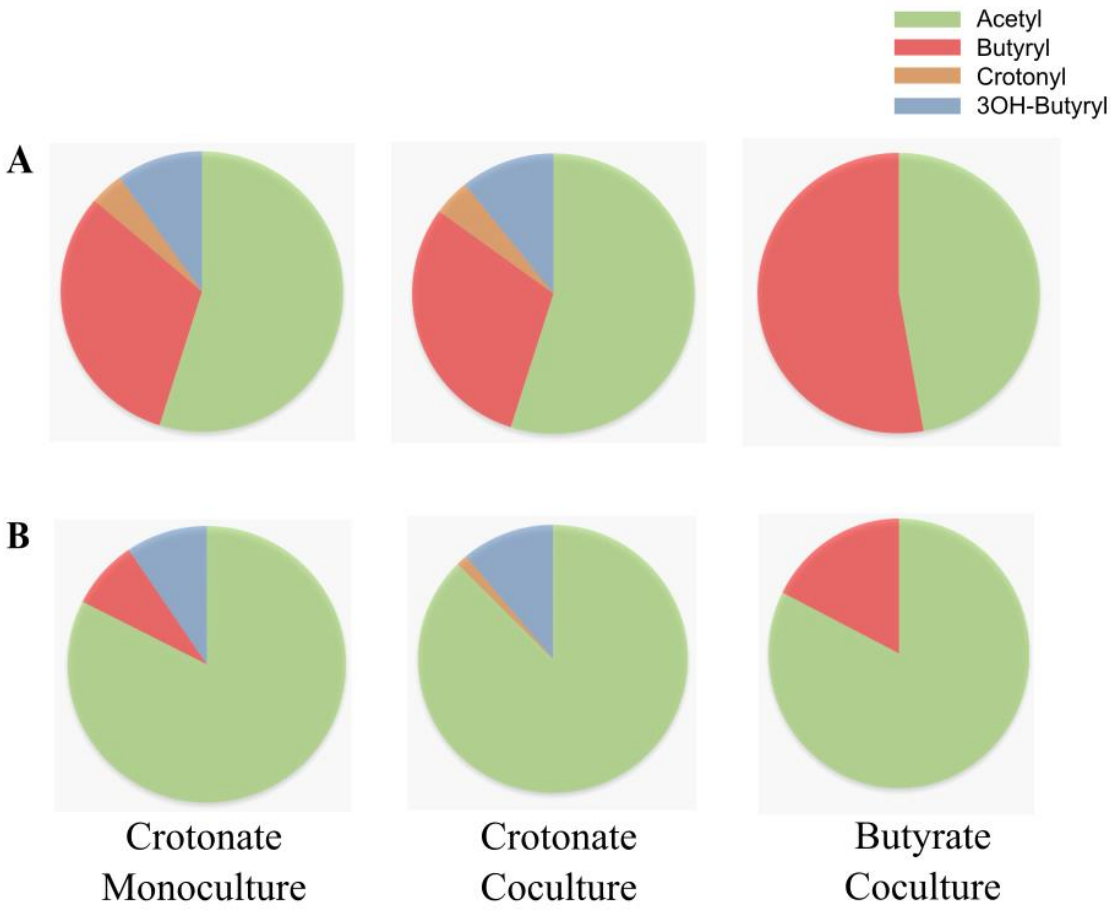


Figure 3. Quantitative shifts in select heterogeneous lysine modifications: Site occupancies on two select modified residues found within the β-oxidation pathway. K41 of the 3-hydroxybutyryl-CoA dehydrogenase Swol_2030 (**3A**) and K208 of the acetyl-CoA acetyltransferase Swol_2051 (**3B**).

Accession Number	Sequence	Modification	*m/z*	Retention Time, Start (min)	Retention Time, End (min)
1855	AVVDEIKTLSKPVESR	K7(Butyryl)	614.3517	56.90	60.90
1855	AVVDEIKTLSKPVESR	K7(Acetyl)	605.008	52.38	56.38
412	DFIGVVEKVTGK	K8(Butyryl)	681.3883	84.57	88.57
412	DFIGVVEKVTGK	K8(Acetyl)	667.3728	79.05	83.05
412	DFIGVVEKVTGK	K8(3OHBu)	689.3855	78.96	82.96
815	EELEKLQLR	K5(Acetyl)	600.3355	44.12	48.12
815	EELEKLQLR	K5(Butyryl)	614.3519	52.59	56.59
675	EIVPVVIKSK	K7(Acetyl)	577.3627	43.82	47.82
675	EIVPVVIKSK	K7(Butyryl)	591.3803	51.86	55.86
2052	ESDEIGMKHVGGNEK	M7(Oxidation); K8(Acetyl)	563.259	17.22	21.22
2052	ESDEIGMKHVGGNEK	M7(Oxidation); K8(Butyryl)	572.6043	24.38	28.38
2051	FKDEIVPVVIKGK	K11(Acetyl)	505.3045	47.92	51.92
2051	FKDEIVPVVIKGK	K11(Butyryl)	514.6491	55.17	59.17
2030	FVDKAIGAIK	K4(3OHBu)	574.3405	29.29	33.29
2030	FVDKAIGAIK	K4(Croto)	565.3351	50.84	54.84
2030	FVDKAIGAIK	K4(Acetyl)	552.3	43.05	47.05
2030	FVDKAIGAIK	K4(Butyryl)	566.3424	50.97	54.97
2030	GKAAPGTADAVVGR	K2(3OHBu)	678.3704	24.70	28.70
2030	GKAAPGTADAVVGR	K2(Butyryl)	670.3712	33.41	37.41
768	GKEVDVATADSK	K2(3OHBu)	653.3306	13.62	17.62
768	GKEVDVATADSK	K2(Acetyl)	631.3181	20.54	24.54
768	GKEVDVATADSK	K2(Butyryl)	645.3333	28.69	32.69
675	IIGMKVGLPVR	M4(Oxidation); K5(Croto)	633.8856	53.93	57.93
675	IIGMKVGLPVR	M4(Oxidation); K5(3OHBu)	642.887	33.25	37.25
675	IIGMKVGLPVR	M4(Oxidation); K5(Acetyl)	620.8765	46.45	50.45
675	IIGMKVGLPVR	M4(Oxidation); K5(Butyryl)	634.893	53.32	57.32
436	IVPTLKPGAIVTDPR	K6(Croto)	822.9889	61.92	65.92
436	IVPTLKPGAIVTDPR	K6(Butyryl)	823.9979	61.07	65.07
436	IVPTLKPGAIVTDPR	K6(Acetyl)	809.9818	52.89	56.89
815	KDLESNLGIGVK	K1(Acetyl)	657.8668	44.83	48.83
815	KDLESNLGIGVK	K1(3OHBu)	679.8802	36.98	40.98
2051	KGDTVFDTDEHPR	K1(3OHBu)	534.9162	21.30	25.30
2051	KGDTVFDTDEHPR	K1(Croto)	528.915	34.13	38.13
2051	KGDTVFDTDEHPR	K1(Acetyl)	520.2396	25.18	29.18
2051	KGDTVFDTDEHPR	K1(Butyryl)	529.5916	34.10	38.10
412	KINAESLGEAIK	K1(Butyryl)	671.8834	50.62	54.62

412	KINAESLGEAIK	K1(Acetyl)	657.8687	46.00	50.00
1855	KISSIGEVLPVLEK	K1(Acetyl)	777.4629	68.71	72.71
1855	KISSIGEVLPVLEK	K1(Butyryl)	791.4774	73.06	77.06
815	KLEELGGLIR	K1(Acetyl)	585.3485	58.66	62.66
815	KLEELGGLIR	K1(3OHBu)	607.3641	44.24	48.24
2051	KSTPEAMAK	K1(Acetyl); M7(Oxidation)	510.7556	9.26	13.26
2051	KSTPEAMAK	K1(Butyryl); M7(Oxidation)	524.7708	13.34	17.34
2051	LAPAFKK	K6(Butyryl)	422.7679	32.31	36.31
2051	LAPAFKK	K6(Acetyl)	408.7523	23.51	27.51
384	LLTNPKAGR	K6(3OHBu)	528.3137	22.24	26.24
384	LLTNPKAGR	K6(Butyryl)	520.3173	32.51	36.51
2030	NVILYDIDMKFVDK	M9(Oxidation); K10(Acetyl)	885.955	76.79	80.79
2030	NVILYDIDMKFVDK	M9(Oxidation); K10(Butyryl)	899.9665	81.16	85.16
436	SKIVPTLKPGAIVTDPR	K2(Acetyl)	917.5487	47.71	51.71
436	SKIVPTLKPGAIVTDPR	K2(3OHBu)	939.5582	41.28	45.28
2052	STDPKGPSVR	K5(Acetyl)	543.2841	14.90	18.90
2052	STDPKGPSVR	K5(Butyryl)	557.2992	20.23	24.23
2051	STPEAMAKLAPAFK	K8(Acetyl)	760.3961	42.83	46.83
2051	STPEAMAKLAPAFK	K8(Butyryl)	774.4114	49.89	53.89
791/2030	VKEGPGFVVNR	K2(Butyryl)	636.3597	42.03	46.03
791/2030	VKEGPGFVVNR	K2(Acetyl)	622.3434	32.85	36.85
1855	VTLGPKGR	K6(3OHBu)	457.2771	12.42	16.42
1855	VTLGPKGR	K6(Acetyl)	435.264	20.48	24.48
2030	YRPAPPLKQLVR	K8(3OHBu)	508.64	32.83	36.83
2030	YRPAPPLKQLVR	K8(Acetyl)	493.9652	40.34	44.34
1491	FGPGDTVK	N/A	410.713	15.87	18.87
1491	GTGLSQTFTVR	N/A	583.813	36.81	39.81
1491	LQVFEGTVIK	N/A	567.3308	44.41	47.41

Table 1. Ions targeted for quantitative analysis: Ions were selected for a parallel reaction monitoring (PRM) analysis of *S. wolfei* samples. Peptides were selected for analysis largely based on their presence in the butyrate degradation pathway of the organism, or by the presence of different modifications identified on the same site.

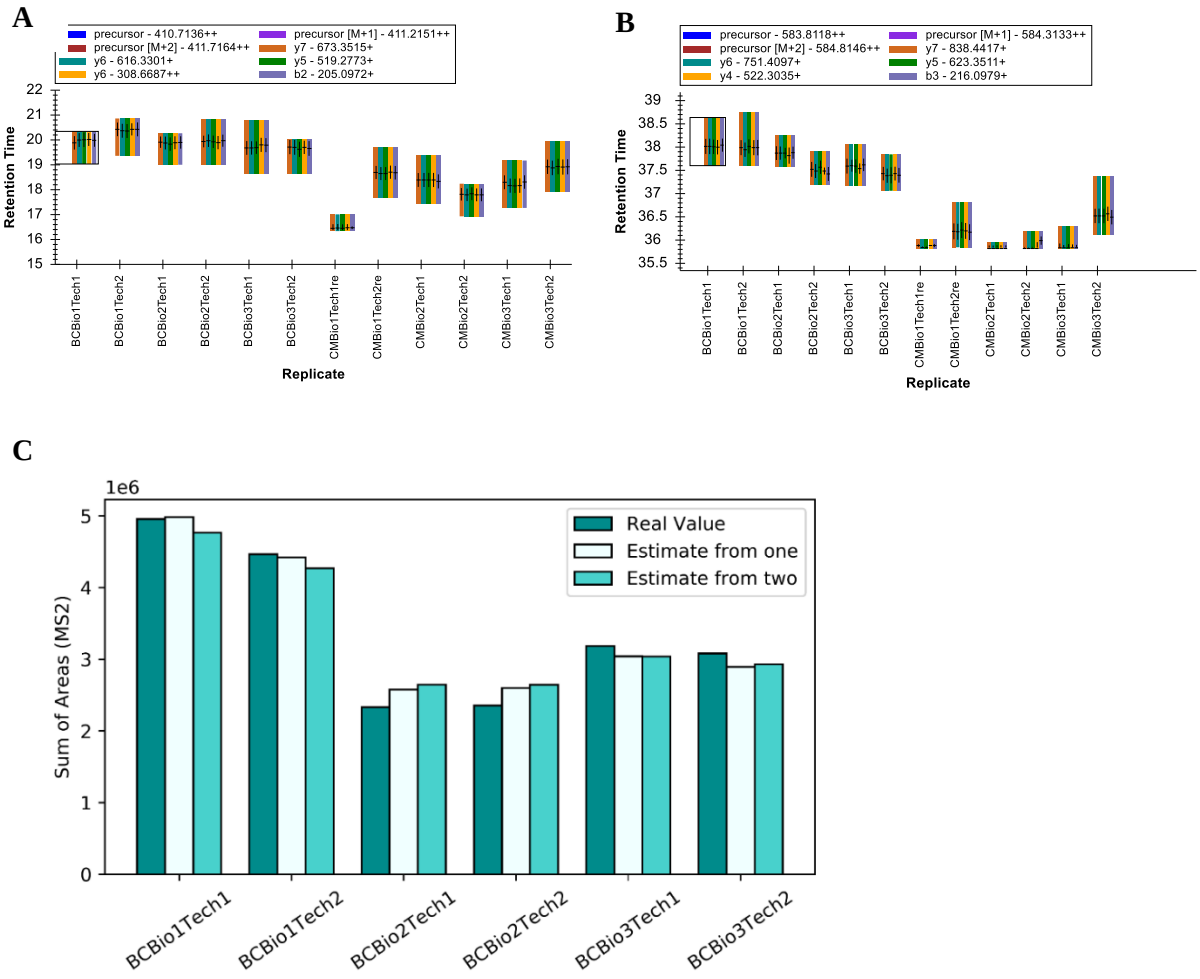


Figure 4. Validation of estimations used in incomplete datasets: Figures **4A** and **4B** indicate the retention time elution profile of two of the peptides selected as internal standards (FGPGDTVK and GTGLSQTFTVR respectively). For butyrate coculture, where all replicates successfully identified all three peptides, a comparison is shown between the true value of sums identified as well as estimates that could be determined from the presence of either one or two of the identified peptides (**4C**).

Accession Number	Protein Function	Protein Ratio (Crotonate Mono : Butyrate Co)
0384	Acyl-CoA dehydrogenase	1.18
0436	Butyrate:acetyl-CoA coenzyme A-transferase	1.41
2030	3-hydroxyacyl-CoA dehydrogenase	1.69
2051	Acetyl-CoA acetyltransferase	1.20
2052	Butyryl-CoA dehydrogenase	1.66

Table 2. Proteins abundance ratios for MS^2 quantitation: Ratios of proteins selected for quantitative analysis as identified in previous DDA analysis using the top 3 peptide method of quantitation.

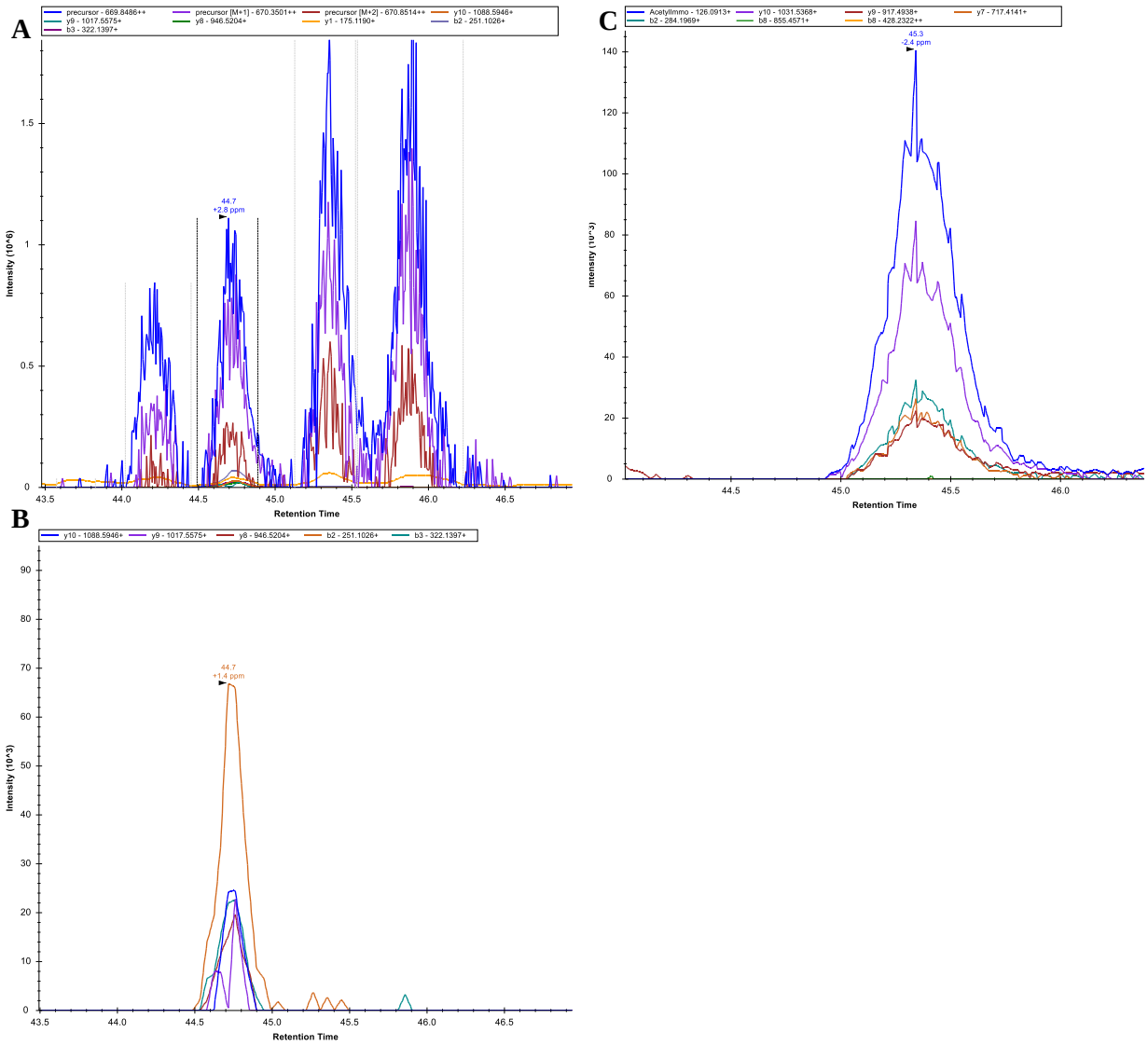


Figure 5. MS^2 level quantitation of acyl peptides: Elution profiles of peptides analyzed by MS^1 and MS^2 analysis. **5A** demonstrates the elution profile of a peptide when the MS^1 extracted ion chromatogram is selected, in comparison to the MS^2 ion chromatogram (**5B**). **5C** demonstrates the elution profile of the extracted MS2 ions for both fragmented ions and the immonium ion for the K1 acetylated form of the KINAESLGEAIK peptide.

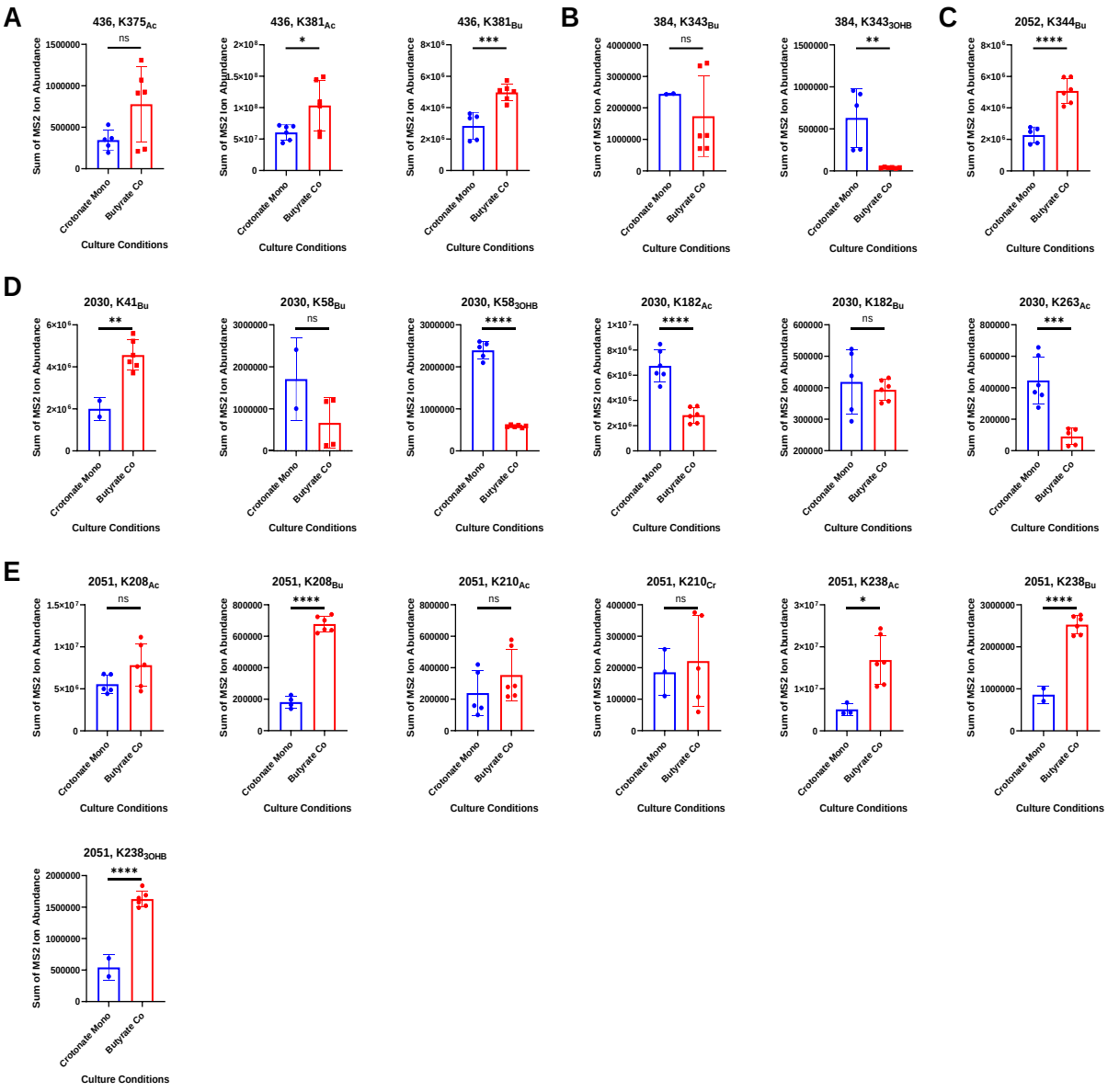


Figure 6. Quantitative change of acyl modifications in crotonate and butyrate growth conditions: Acyl-peptides were identified and quantified in a number of enzymes in the butyrate degradation pathway (**5A-E**). An unpaired t-test was used to determine significance (ns, $P > 0.05$; *, $P < 0.05$; **, $P < 0.005$; ***, $P < 0.001$; ****, $P < 0.0001$).

CHAPTER 5

Conclusion and Perspectives

5.1 Moving Forward in Proteomics

The field of proteomics has advanced significantly over the last twenty years. Tremendous technological advances, such as the development of the high resolution orbitrap mass analyzer (1), newer fragmentation methods such as electron transfer dissociation (ETD) and ultraviolet photodissociation (UVPD) (2, 3), incorporation of ion mobility based gas-phase separation techniques (4, 5), and multiplexing of mass spectra (6) have enhanced the amount of data that can be generated in a single proteomics run. However, this generation of raw data does not *necessarily* mean that all data is usable or can provide meaningful conclusions. As the field of proteomics turns toward generating larger and deeper datasets, it must be kept in mind that a rather large portion of mass spectra generated in a proteomics shotgun experiment (up to 75%) are do not lead to a protein identification (7, 8). Many attempts to overcome and assign these spectra have been previously discussed in Chapter 1, as using "open searches" that allow for unbiased search of mass shifts that can later be assigned to a chemical formula have become more common in assigning spectra that contain non-canonical cleavage sites, amino acid mutations, and post-translational modifications (9–11). However, this type of searching also comes with it the drawback that for large datasets, determining true, high confident assignments from misidentifications is not as clear as with normal false-discovery determination methods. Therefore, the notion that these larger data sets coupled with larger search spaces will generate inherently more meaningful or usable data requires further considerations. Many issues may arise from isobaric peptides and side reactions caused by sample processing and preparation (12–14). Rather than just focusing on generating *more* spectra, focus should also be placed on the validation of assignments made and determination of what data is significant.

The push to make data more interpretable and meaningful has taken several forms in recent years. Many methods and approaches have begun to focus on understanding fragment-spectra complexities and minimizing noise that may occur. For example, chimeric spectra, spectra in which co-eluting peptides also co-fragment in a single spectra, can represent as much as 50% of the spectra in a proteomic experiment and significantly impair confident peptide identification (15). Increasingly these spectra are being addressed by software to help identify peptides that would otherwise be ignored or unassignable (16). Likewise, addressing the contaminants that may factor into an analysis can help to parse through data to find the meaningful sets. To address the signals resulting from contamination from the sample preparation steps, databases such as the CRAPome (17) have allowed for the identification of proteins that are common in the background of experiments (and subsequently removed for consideration when interpreting the results). At the mass spectrometric level, software has been written to combat contaminating ions that may permeate fragment spectra throughout the course of a proteomics experiment (18). These resources are just a few of the growing considerations being made to help mitigate, or potentially even make use of, interferences that can otherwise impede data assignments or contribute to ambiguity in data analysis.

Along a similar vein, the transition of the proteomics community toward MS^2 level analysis, even for purposes of quantification, has allowed for improved accuracy in determining what is of chemical and biological significance. Permutations of amino acids can be distinguished using fragment spectra in a manner not possible from LC-MS analysis alone. The power of this methodology also increases when post-translational modifications (PTMs) are considered in the context of proteomic analysis. Indeed, Basisty *et. al* demonstrated that fragment-level quantitation can distinguish between isomeric PTM combinations that elute within a minute of

one another; a feature impossible to do with precursor mass measurement alone (19). The ability to validate this type of data analysis has also grown significantly with advances in tandem mass spectra predictive algorithms. Machine learning methods have allowed for the prediction of peptide fragment spectra with remarkable accuracy (20, 21). As the sizes of proteomic datasets expand and groups continue to integrate both identified and theoretical PTMs into training datasets, the power and accuracy of these algorithms will continue to grow. Agreement of these predictions and empirical data will continue to give strong credibility to the assignment of a peptide's identity.

Altogether, the current trends in proteomics are trending to not only collect *more* data and assign the identity of *more* ions, but to do so with greater accuracy using a multitude of methods and factors. In this way, both the number of peptides identified can increase as can the confidence of these assignments. As ambiguity is a key limitation to the proteomics field, giving confidence to the assignment of peptides is a key frontier to overcome, especially when considering the growing contribution of non-canonical proteins and PTMs to key biological questions.

The three-pronged advancements in proteomics; the ability to collect more data, assign more data, and have more confidence in data assignments, expands potential for current applications of mass spectrometry-based proteomics as a field. Chapters 2, 3, and 4 discussed data that were collected from microbial communities to better understand interactions across these communities. Indeed, one of the next frontiers of proteomic applications will be growth in the size of sample complexity that can be analyzed. Syntrophs are one class of biological systems that require partners to grow and therefore require scientific interrogation in consortia. However,

the capabilities of studying organisms in physically relevant conditions expands well beyond these microbes.

Increasingly in recent years, focus has turned toward analyzing many of these larger and more complex biological systems. Viruses, by necessity, integrate into host cells in order to grow, mature and replicate. To fully understand this process, looking at the host-virus interactions, and from a proteomics perspective, the interactions of proteins at the interface are critical. These interactions have been elucidated at the proteomic level for a number of different viruses, thereby defining the interactome of a number of different viral pathogens and their hosts (22–24). Increasing in size and complexity, bacterial pathogens have likewise been analyzed at the proteomic level with MS based techniques, thereby demonstrating significant interactors that can result in pathogenicity (25, 26). While these techniques clearly extend to human health in the host-pathogen capacity, they also have the power to act in much larger microbial communities that can have a significant impact as well. Microbes in soil proximal to plants, often in the rhizosphere, can have a large effect on plant growth. Metaproteomic studies of these communities have become critical to understanding impact that microbial communities play in plant growth and health (27). Extrapolating to even more complex systems, metaproteomics has been applied to the human gut microbiome to understand how the microbial communities within the human gut may result in a variety of medical dysfunctions (28–30). As proteomics transitions to larger communities, many more physiologically relevant conditions can be tested, and thus the emerging properties within these systems can be identified. Though a relatively simple system, the existence of syntrophs demonstrates how isolating and simplifying systems through reductionist approaches may miss out on critical features.

5.2 References

1. Hu, Q., Noll, R. J., Li, H., Makarov, A., Hardman, M., and Graham Cooks, R. (2005) The Orbitrap: a new mass spectrometer. *J. Mass Spectrom.* **40**, 430–443
2. Ly, T., and Julian, R. R. (2009) Ultraviolet photodissociation: developments towards applications for mass-spectrometry-based proteomics. *Angew. Chemie - Int. Ed.* **48**, 7130–7137
3. Syka, J. E. P., Coon, J. J., Schroeder, M. J., Shabanowitz, J., and Hunt, D. F. (2004) Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 9528–9533
4. Meier, F., Brunner, A. D., Koch, S., Koch, H., Lubeck, M., Krause, M., Goedecke, N., Decker, J., Kosinski, T., Park, M. A., Bache, N., Hoerning, O., Cox, J., Räther, O., and Mann, M. (2018) Online parallel accumulation–serial fragmentation (PASEF) with a novel trapped ion mobility mass spectrometer. *Mol. Cell. Proteomics*. **17**, 2534–2545
5. Bonneil, E., Pfammatter, S., and Thibault, P. (2015) Enhancement of mass spectrometry performance for proteomic analyses using high-field asymmetric waveform ion mobility spectrometry (FAIMS). *J. Mass Spectrom.* **50**, 1181–1195
6. Egertson, J. D., Kuehn, A., Merrihew, G. E., Bateman, N. W., MacLean, B. X., Ting, Y. S., Canterbury, J. D., Marsh, D. M., Kellmann, M., Zabrouskov, V., Wu, C. C., and MacCoss, M. J. (2013) Multiplexed MS/MS for improved data-independent acquisition. *Nat. Methods*. **10**, 744–6
7. Griss, J., Perez-Riverol, Y., Lewis, S., Tabb, D. L., Dianes, J. A., Del-Toro, N., Rurik, M., Walzer, M., Kohlbacher, O., Hermjakob, H., Wang, R., and Vizcano, J. A. (2016) Recognizing millions of consistently unidentified spectra across hundreds of shotgun

proteomics datasets. *Nat. Methods*. **13**, 651–656

8. Chick, J. M., Kolippakkam, D., Nusinow, D. P., Zhai, B., Rad, R., Huttlin, E. L., and Gygi, S. P. (2015) A mass-tolerant database search identifies a large proportion of unassigned spectra in shotgun proteomics as modified peptides. *Nat. Biotechnol.* **33**, 743–749
9. Ye, D., Fu, Y., Sun, R. X., Wang, H. P., Yuan, Z. F., Chi, H., and He, S. M. (2010) Open MS/MS spectral library search to identify unanticipated post-translational modifications and increase spectral identification rate. *Bioinformatics*. 10.1093/bioinformatics/btq185
10. Bittremieux, W., Meysman, P., Noble, W. S., and Laukens, K. (2018) Fast Open Modification Spectral Library Searching through Approximate Nearest Neighbor Indexing. *J. Proteome Res.* **17**, 3463–3474
11. Kong, A. T., Leprevost, F. V., Avtonomov, D. M., Mellacheruvu, D., and Nesvizhskii, A. I. (2017) MSFragger: Ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nat. Methods*. **14**, 513–520
12. Rodriguez, J., Gupta, N., Smith, R. D., and Pevzner, P. A. (2008) Does trypsin cut before proline? *J. Proteome Res.* **7**, 300–305
13. Kim, M. S., Zhong, J., and Pandey, A. (2016) Common errors in mass spectrometry-based analysis of post-translational modifications. *Proteomics*. **16**, 700–714
14. Lee, S., Tan, M., Dai, L., Kwon, O. K., Yang, J. S., Zhao, Y., and Chen, Y. (2013) MS/MS of synthetic peptide is not sufficient to confirm new types of protein modifications. *J. Proteome Res.* **12**, 1007–1013
15. Houel, S., Abernathy, R., Renganathan, K., Meyer-Arendt, K., Ahn, N. G., and Old, W. M. (2010) Quantifying the impact of chimera MS/MS spectra on peptide identification in

large-scale proteomics studies. *J. Proteome Res.* **9**, 4152–4160

16. Dorfer, V., Maltsev, S., Winkler, S., and Mechtler, K. (2018) CharmeRT: Boosting Peptide Identifications by Chimeric Spectra Identification and Retention Time Prediction. *J. Proteome Res.* **17**, 2581–2589

17. Mellacheruvu, D., Wright, Z., Couzens, A. L., Lambert, J. P., St-Denis, N. A., Li, T., Miteva, Y. V., Hauri, S., Sardiu, M. E., Low, T. Y., Halim, V. A., Bagshaw, R. D., Hubner, N. C., Al-Hakim, A., Bouchard, A., Faubert, D., Fermin, D., Dunham, W. H., Goudreault, M., Lin, Z. Y., Badillo, B. G., Pawson, T., Durocher, D., Coulombe, B., Aebersold, R., Superti-Furga, G., Colinge, J., Heck, A. J. R., Choi, H., Gstaiger, M., Mohammed, S., Cristea, I. M., Bennett, K. L., Washburn, M. P., Raught, B., Ewing, R. M., Gingras, A. C., and Nesvizhskii, A. I. (2013) The CRAPome: A contaminant repository for affinity purification-mass spectrometry data. *Nat. Methods*. **10**, 730–736

18. Huan, T., Xing, S., Yu, H., Liu, M., Jia, Q., Sun, Z., and Fang, M. (2021) Recognizing contamination fragment ions in liquid chromatography-tandem mass spectrometry data. *J. Am. Soc. Mass Spectrom.* **22**, 17

19. Basisty, N., Meyer, J. G., Wei, L., Gibson, B. W., and Schilling, B. (2018) Simultaneous Quantification of the Acetylome and Succinylome by 'One-Pot' Affinity Enrichment. *Proteomics*. **18**, 1800123

20. Searle, B. C., Swearingen, K. E., Barnes, C. A., Schmidt, T., Gessulat, S., Küster, B., and Wilhelm, M. (2020) Generating high quality libraries for DIA MS with empirically corrected peptide predictions. *Nat. Commun.* **11**, 1–10

21. Röst, H. L. (2019) Deep learning adds an extra dimension to peptide fragmentation. *Nat. Methods*. **16**, 469–470

22. Watanabe, T., Kawakami, E., Shoemaker, J. E., Lopes, T. J. S., Matsuoka, Y., Tomita, Y., Kozuka-Hata, H., Gorai, T., Kuwahara, T., Takeda, E., Nagata, A., Takano, R., Kiso, M., Yamashita, M., Sakai-Tagawa, Y., Katsura, H., Nonaka, N., Fujii, H., Fujii, K., Sugita, Y., Noda, T., Goto, H., Fukuyama, S., Watanabe, S., Neumann, G., Oyama, M., Kitano, H., and Kawaoka, Y. (2014) Influenza virus-host interactome screen as a platform for antiviral drug development. *Cell Host Microbe*. **16**, 795–805

23. Gordon, D. E., Jang, G. M., Bouhaddou, M., Xu, J., Obernier, K., White, K. M., O'Meara, M. J., Rezelj, V. V., Guo, J. Z., Swaney, D. L., Tummino, T. A., Hüttenhain, R., Kaake, R. M., Richards, A. L., Tutuncuoglu, B., Foussard, H., Batra, J., Haas, K., Modak, M., Kim, M., Haas, P., Polacco, B. J., Braberg, H., Fabius, J. M., Eckhardt, M., Soucheray, M., Bennett, M. J., Cakir, M., McGregor, M. J., Li, Q., Meyer, B., Roesch, F., Vallet, T., Mac Kain, A., Miorin, L., Moreno, E., Naing, Z. Z. C., Zhou, Y., Peng, S., Shi, Y., Zhang, Z., Shen, W., Kirby, I. T., Melnyk, J. E., Chorba, J. S., Lou, K., Dai, S. A., Barrio-Hernandez, I., Memon, D., Hernandez-Armenta, C., Lyu, J., Mathy, C. J. P., Perica, T., Pilla, K. B., Ganesan, S. J., Saltzberg, D. J., Rakesh, R., Liu, X., Rosenthal, S. B., Calviello, L., Venkataramanan, S., Liboy-Lugo, J., Lin, Y., Huang, X. P., Liu, Y. F., Wankowicz, S. A., Bohn, M., Safari, M., Ugur, F. S., Koh, C., Savar, N. S., Tran, Q. D., Shengjuler, D., Fletcher, S. J., O'Neal, M. C., Cai, Y., Chang, J. C. J., Broadhurst, D. J., Klippsten, S., Sharp, P. P., Wenzell, N. A., Kuzuoglu-Ozturk, D., Wang, H. Y., Trenker, R., Young, J. M., Cavero, D. A., Hiatt, J., Roth, T. L., Rathore, U., Subramanian, A., Noack, J., Hubert, M., Stroud, R. M., Frankel, A. D., Rosenberg, O. S., Verba, K. A., Agard, D. A., Ott, M., Emerman, M., Jura, N., von Zastrow, M., Verdin, E., Ashworth, A., Schwartz, O., d'Enfert, C., Mukherjee, S., Jacobson, M., Malik, H. S., Fujimori, D. G.,

Ideker, T., Craik, C. S., Floor, S. N., Fraser, J. S., Gross, J. D., Sali, A., Roth, B. L., Ruggero, D., Taunton, J., Kortemme, T., Beltrao, P., Vignuzzi, M., García-Sastre, A., Shokat, K. M., Shoichet, B. K., and Krogan, N. J. (2020) A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature*. **583**, 459–468

24. García-Dorival, I., Wu, W., Dowall, S., Armstrong, S., Touzelet, O., Wastling, J., Barr, J. N., Matthews, D., Carroll, M., Hewson, R., and Hiscox, J. A. (2014) Elucidation of the Ebola Virus VP24 Cellular Interactome and Disruption of Virus Biology through Targeted Inhibition of Host-Cell Protein Function. *J. Proteome Res.* **13**, 5120–5135
25. Díaz-Pascual, F., Ortíz-Severín, J., Varas, M. A., Allende, M. L., and Chávez, F. P. (2017) In vivo Host-Pathogen Interaction as Revealed by Global Proteomic Profiling of Zebrafish Larvae. *Front. Cell. Infect. Microbiol.* **7**, 334
26. Michalik, S., Depke, M., Murr, A., Gesell Salazar, M., Kusebauch, U., Sun, Z., Meyer, T. C., Surmann, K., Pförtner, H., Hildebrandt, P., Weiss, S., Palma Medina, L. M., Gutjahr, M., Hammer, E., Becher, D., Pribyl, T., Hammerschmidt, S., Deutsch, E. W., Bader, S. L., Hecker, M., Moritz, R. L., Mäder, U., Völker, U., and Schmidt, F. (2017) A global Staphylococcus aureus proteome resource applied to the in vivo characterization of host-pathogen interactions. *Sci. Rep.* **7**, 1–16
27. Tartaglia, M., Bastida, F., Sciarrillo, R., and Guarino, C. (2020) Soil metaproteomics for the study of the relationships between microorganisms and plants: A review of extraction protocols and ecological insights. *Int. J. Mol. Sci.* **21**, 1–20
28. Zhang, X., Deeke, S. A., Ning, Z., Starr, A. E., Butcher, J., Li, J., Mayne, J., Cheng, K., Liao, B., Li, L., Singleton, R., Mack, D., Stintzi, A., and Figeys, D. (2018) Metaproteomics reveals associations between microbiome and intestinal extracellular

vesicle proteins in pediatric inflammatory bowel disease. *Nat. Commun.* **9**, 1–14

29. Zhang, X., and Figeys, D. (2019) Perspective and Guidelines for Metaproteomics in Microbiome Studies. *J. Proteome Res.* **18**, 2370–2380

30. Long, S., Yang, Y., Shen, C., Wang, Y., Deng, A., Qin, Q., and Qiao, L. (2020) Metaproteomics characterizes human gut microbiome function in colorectal cancer. *npj Biofilms Microbiomes*. **6**, 1–10

APPENDIX TO THE DISSERTATION

Introduction

The appendix includes several projects that are related to several of the themes of the dissertation, such as mass spectrometry, post-translational modifications, and environmental microbes. However, these works either do not tie into *all* the themes of the dissertation or require future work to complete the entirety of the story.

APPENDIX I: In vitro reconstitution of sortase-catalyzed pilus polymerization reveals structural elements involved in pilin cross-linking

In vitro reconstitution of sortase-catalyzed pilus polymerization reveals structural elements involved in pilin cross-linking

Chungyu Chang[a,1], Brendan R. Amer[b,c,1], Jerzy Osipiuk[d,e], Scott A. McConnell[b,c], I-Hsiu Huang[f], Van Hsieh[b,c], Janine Fu[b,c], Hong H. Nguyen[b,c], John Muroski[b,c], Erika Flores[a], Rachel R. Ogorzalek Loo[b,c], Joseph A. Loo[b,c], John A. Putkey[g], Andrzej Joachimiak[d,e], Asis Das[h], Robert T. Clubb[b,c,2], and Hung Ton-That[a,2]

[a]Department of Microbiology and Molecular Genetics, University of Texas Health Science Center, Houston, TX 77030; [b]Department of Chemistry and Biochemistry, University of California, Los Angeles, CA 90095; [c]University of California, Los Angeles-US Department of Energy Institute of Genomics and Proteomics, University of California, Los Angeles, CA 90095; [d]Center for Structural Genomics of Infectious Diseases, Argonne National Laboratory, Argonne, IL 60439; [e]Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, IL 60637; [f]Department of Microbiology and Immunology, College of Medicine, National Cheng Kung University, Tainan 701, Taiwan; [g]Department of Biochemistry and Molecular Biology, University of Texas Health Science Center, Houston, TX 77030; and [h]Department of Molecular Biology and Biophysics, University of Connecticut Health Center, Farmington, CT 06030

Edited by Ralph R. Isberg, Howard Hughes Medical Institute and Tufts University School of Medicine, Boston, MA, and approved May 2, 2018 (received for review January 18, 2018)

Covalently cross-linked pilus polymers displayed on the cell surface of Gram-positive bacteria are assembled by class C sortase enzymes. These pilus-specific transpeptidases located on the bacterial membrane catalyze a two-step protein ligation reaction, first cleaving the LPXTG motif of one pilin protomer to form an acyl-enzyme intermediate and then joining the terminal Thr to the nucleophilic Lys residue residing within the pilin motif of another pilin protomer. To date, the determinants of class C enzymes that uniquely enable them to construct pili remain unknown. Here, informed by high-resolution crystal structures of corynebacterial pilus-specific sortase (SrtA) and utilizing a structural variant of the enzyme ($SrtA^{2M}$), whose catalytic pocket has been unmasked by activating mutations, we successfully reconstituted in vitro polymerization of the cognate major pilin (SpaA). Mass spectrometry, electron microscopy, and biochemical experiments authenticated that $SrtA^{2M}$ synthesizes pilus fibers with correct Lys–Thr isopeptide bonds linking individual pilins via a thioacyl intermediate. Structural modeling of the SpaA–SrtA–SpaA polymerization intermediate depicts $SrtA^{2M}$ sandwiched between the N- and C-terminal domains of SpaA harboring the reactive pilin and LPXTG motifs, respectively. Remarkably, the model uncovered a conserved TP(Y/L)XIN(S/T)H signature sequence following the catalytic Cys, in which the alanine substitutions abrogated cross-linking activity but not cleavage of LPXTG. These insights and our evidence that $SrtA^{2M}$ can terminate pilus polymerization by joining the terminal pilin SpaB to SpaA and catalyze ligation of isolated SpaA domains in vitro provide a facile and versatile platform for protein engineering and bio-conjugation that has major implications for biotechnology.

Corynebacterium diphtheriae | sortase | pilus polymerization | protein ligation | transpeptidation

Adhesive protein polymers, called "pili" or "fimbriae," are expressed on the cell envelope by many Gram-negative and Gram-positive bacteria, and they are critical for bacterial virulence (1). Many types of Gram-negative pili have been reported, including the well-studied retractable type IV, conjugative, and chaperone-assisted pili (2). These pili are formed by distinct pathways (2, 3); however, none of these pili are covalently linked polymers, unlike the sortase-catalyzed pili found in many Gram-positive bacteria, including *Actinomyces oris*, *Enterococcus faecalis*, *Bacillus cereus*, and numerous species of streptococci and lactobacilli (4–6).

One of the well-studied sortase-mediated pilus assembly systems involves *Corynebacterium diphtheriae* (7), the causative agent of pharyngeal diphtheria (8). *C. diphtheriae* produces three distinct pilus types (7, 9, 10), each comprised of a pilus tip adhesin, a pilus shaft made of the major pilin, and a base pilin that is covalently anchored to the cell wall (11). The archetypal SpaA-type pilus, which mediates adherence to the pharyngeal epithelium (12), consists of the tip pilin SpaC, shaft pilin SpaA, and pilus base SpaB (13). A pilus-specific sortase named "SrtA" is required for pilus polymerization (13), performing a repetitive, irreversible transpeptidation reaction that covalently links the pilin subunits via an isopeptide bond (14). Although each Spa pilin harbors a cell wall sorting signal (CWSS), which starts with a conserved LPXTG motif, followed by a stretch of hydrophobic amino acids and a positively charged tail (15), SpaA contains a pilin motif with the Lys residue K190 acting as a nucleophile for the aforementioned transpeptidation reaction (13). According to the current model (16), SrtA cleaves the LPXTG motif of Spa pilins between Thr and Gly, forming acyl-enzyme intermediates

Significance

Gram-positive sortase enzymes represent two broad functional categories—those that cross-link proteins to the cell wall and those that can catalyze this reaction and polymerize proteins to build adhesive pilus fibers. Here we report an in vitro reproduction of a robust pilus polymerization reaction using a variant of a corynebacterial pilus-specific sortase in which the catalytic center is unmasked. By molecular modeling, we uncovered a conserved structural element of pilus-specific sortases critical for protein ligation in vitro and further demonstrated that the activated sortase ligates the isolated domains of the pilin harboring the donor and acceptor motifs for ligation. Besides enabling future molecular studies and antibiotic development, our system provides a powerful platform for bioconjugation and protein engineering.

Author contributions: C.C., B.R.A., J.O., J.A.P., A.J., R.T.C., and H.T.-T. designed research; C.C., B.R.A., J.O., S.A.M., I.-H.H., V.H., J.F., H.H.N., J.M., E.F., R.R.O.L., and J.A.P. performed research; C.C., B.R.A., J.O., J.A.L., J.A.P., A.J., A.D., R.T.C., and H.T.-T. analyzed data; and C.C., B.R.A., J.O, A.D., R.T.C., and H.T.-T. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.wwpdb.org (PDB ID codes 5K9A and 6BWE).

[1]C.C. and B.R.A. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: rclubb@mbi.ucla.edu or ton-that.hung@uth.tmc.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1800954115/-/DCSupplemental.

Published online May 29, 2018.

between the Thr residue and the SrtA catalytic Cys residue. This intermediate is then nucleophilically attacked by the reactive K190 of an incoming SpaA subunit. In pilus biogenesis, the SpaC–SrtA acyl-enzyme intermediate forms first, resulting in the joining of the ε-amine group of K190 to the Thr carbonyl carbon atom in the LPXT of SpaC. Pilus polymerization ensues when additional SpaA protomers are joined progressively to the pilus base by the SrtA enzyme via the same Lys-mediated transpeptidation reaction. Polymerization is terminated with the entry of SpaB into the pilus base (11), which is then anchored to the cell wall by the housekeeping sortase SrtF (17). This cell wall anchoring of pilus polymers is likely similar to that of surface proteins catalyzed by the prototype SrtA enzyme from *Staphylococcus aureus* (18, 19). While most of this biphasic model of pilus assembly in Gram-positive bacteria (6), i.e., pilus polymerization followed by cell wall anchoring, has been validated experimentally, the molecular determinants that make up a pilus-specific sortase and enable the enzyme to join proteins together remain unknown.

The SrtA enzyme is classified as a member of the class C sortase subgroup within the sortase superfamily that has the unique ability to cross-link proteins via Lys–Thr isopeptide bonds (20, 21). Although all sortases share a canonical β-barrel sortase superfamily fold (22, 23), class C enzymes are distinguished by the presence of a conserved N-terminal region that forms a "lid" that covers the active site structurally and functionally (24–26). In *Streptococcus pneumoniae*, X-ray crystallographic evidence originally suggested that the lid region was flexible, possibly modulating substrate binding; however, subsequent studies in solution utilizing NMR showed this region to be relatively rigid in the SrtC1 enzyme (24, 27–29). Mutations of the lid region in *A. oris* SrtC2 or *Streptococcus agalactiae* SrtC1 did not alter the pilus polymerizing activities in vivo (30, 31); nonetheless, the mutations caused enzyme instability and increased hydrolytic activity in *S. agalactiae* SrtC1 (30), supporting a regulatory role for the N-terminal lid. However, the unique structural properties that enable class C sortase enzymes to cross-link proteins have not been identified.

We report here the crystal structures of the *C. diphtheriae* class C sortase SrtA lacking the signal peptide and transmembrane domain (referred to as "SrtA^{WT}") and a mutant of this protein that has substitutions in the lid interface which normally masks

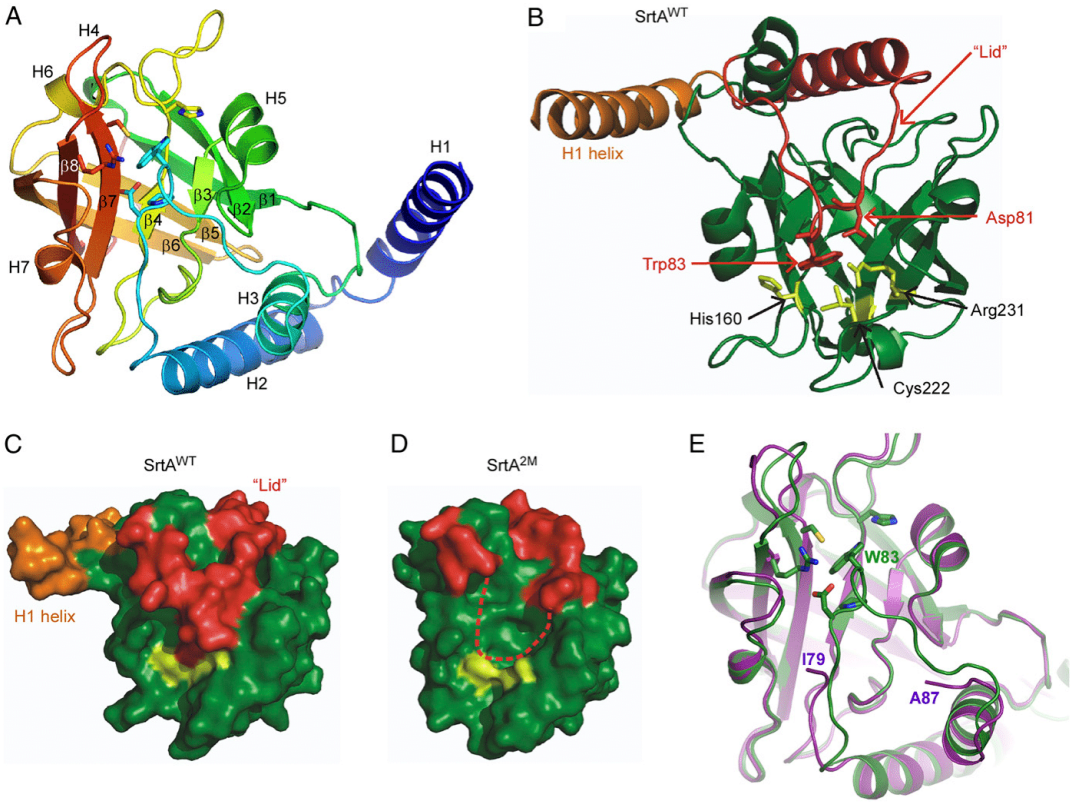


Fig. 1. Structural analysis of the *C. diphtheriae* pilus-specific sortase SrtA. (*A*) The crystal structure of SrtA was determined to 2.1-Å resolution, with the overall protein fold presented as rainbow coloring from blue to red corresponding to the N- to C-terminal positions. The helices are marked as H1–H7, and the β-strands are marked as β1–β8. (*B*) The lid region is marked in red with conserved lid residues D81 and W83 and catalytic residues C222, H160, and R231 in yellow. The C222 residue is shown only in the main conformation. (*C* and *D*) Hydrophobic surface renderings of WT SrtA (SrtA^{WT}) (*C*) and the lid mutant (SrtA^{2M}) structures (*D*) with the H1 helix and lid loop structures in red. The H1 helix of SrtA^{2M} is absent, and its lid structure is not visible, as indicated by a red dashed line. (*E*) Superposition of the SrtA^{WT} (green) and the lid mutant SrtA^{2M} (pink) structures was generated by PyMOL.

Table 1. Crystal data collection statistics

Statistics	$SrtA^{WT}$	$SrtA^{2M}$
X-ray wavelength, Å	0.9792	0.9792
Space group	P 6_122	P 2_1
Unit cell dimensions	a = b = 77.7 Å, c = 202.3 Å, $\alpha = \beta = 90°$, $\gamma = 120°$	a = 65.2 Å, b = 45.5 Å, c = 74.9 Å, $\alpha = \gamma = 90°$, $\beta = 96.4$
Resolution, Å	38.9–2.1 (2.14–2.1)	38.9–1.85 (1.88–1.85)
No. of unique reflections	21,872 (1,050)	36,705 (1,431)
Completeness, %	99.7 (100)	97.7 (77.4)
R-merge	0.142 (0.916)	0.087 (0.410)
CC1/2, $Å^2$	–(0.939)	–(0.741)
I/σ	30.7 (6.4)	16.5 (1.95)
Redundancy	15.4 (15.2)	3.5 (2.2)
Molecules per asymmetric unit	1	2
No. of protein residues	215	430

Numbers in parenthesis are shown for the highest-resolution shell.

the catalytic pocket ($SrtA^{2M}$). Using these recombinant enzymes and an SpaA substrate that lacks the signal peptide and transmembrane domain, we succeeded in reconstituting the SpaA pilus shaft polymerization reaction in vitro, demonstrating that the removal of SrtA's lid not only unmasks the catalytic center structurally but also enables the polymerizing activity in vitro. Subsequently, by structural modeling and phylogenic and mutational analyses, we identified two structural elements that enable SrtA to cross-link proteins. Importantly, we showed that the activated sortase can ligate the isolated pilin domains, thus defining the donor and acceptor motifs for the ligation reaction. The system we report provides a platform for in vitro mechanistic investigations of Gram-positive pilus assembly, antibiotic development, and biotechnological applications of protein modification and conjugation via a unique transpeptidation reaction.

Results and Discussions

Structure of the *C. diphtheriae* Pilus-Specific Sortase. The archetypal SpaA pilus polymer produced by corynebacteria is built by the dedicated pilus-specific sortase SrtA (7, 13). To gain insight into the mechanism of pilus polymerization, we determined the structure of SrtA by X-ray crystallography. We performed crystallization screens using a soluble fragment encompassing the catalytic domain of SrtA (residues 37–257, $SrtA^{WT}$), which was cloned, expressed, and purified from *Escherichia coli*. $SrtA^{WT}$ crystallized as a homodimer in the $P6_1$ 2 2 space group. Diffraction data were collected to 2.1-Å resolution and were phased by molecular replacement (Tables 1 and 2). The electron density for residues 37–248 was well defined, enabling their structure to be modeled, while density for the remaining C-terminal residues is missing, presumably due to a disordered state.

Table 2. Structure refinement statistics

Statistics	$SrtA^{WT}$	$SrtA^{2M}$
Resolution range, Å	38.9–2.1 (2.157–2.1)	38.9–1.85 (1.898–1.85)
Reflections	21,846 (1,533)	36,692 (2,195)
σ cutoff	None	None
R-value, all, %	16.22	17.26
R-value (R-work), %	16.05 (17.4)	17.12 (23.7)
Free R-value, %	19.56 (24.0)	20.08 (23.7)
Rms deviations from ideal geometry		
Bond length, Å	0.017	0.012
Angle, °	1.76	1.60
Chiral, Å	0.101	0.088
No. of atoms		
Protein	1,737	2,974
Sulfate	5	—
Water	167	258
Mean B-factor, $Å^2$		
All atoms	32.0	30.7
Protein atoms	31.2	30.3
Protein main chain	28.6	29.0
Protein side chain	33.9	31.6
Sulfate	66.0	—
Water	38.5	35.9
MolProbity summary		
Ramachandran outliers, %	0.0	0.0
Ramachandran favored, %	97.18	98.89
Rotamer outliers, %	2.02	0.89
C-beta deviations	0.0	0.0
Clash score	1.73	1.34
MolProbity score	1.31	0.86

BIOCHEMISTRY

The overall structure of $SrtA^{WT}$ conforms to the typical sortase fold described previously (22), containing an eight-stranded β-barrel core flanked by several 3_{10} and α-helices (Fig. 1 *A* and *B*). Three additional α-helices are located at the N terminus of $SrtA^{WT}$ (Fig. 1 *A* and *B*) and contain the distinguishing lid structure that occludes the enzyme's active site in a class C sortase (Fig. 1*B*). Interestingly, the H1 helix mediates homodimerization in the crystal structure and is generally removed from the body of the enzyme (Fig. 1 *A* and *B*), while helices H2 and H3 are positioned immediately adjacent to the active site and are connected by a loop that contains the highly conserved DPW lid motif that interacts with the active site (Fig. 1 *A* and *B*). W83 in the lid participates in aromatic stacking interactions with the active site C222 and nearby H160 residues. In addition, D81 within the motif interacts with the active site R231 residue, suggesting its regulatory role in lid positioning and pilin polymerization. Importantly, residues within the catalytic triad His–Cys–Arg are well resolved, and C222 can be modeled in two distinct positions with 50% occupancy, pointing both toward and away from the active site (Fig. 1 *A* and *B*).

To investigate the functional importance of the lid in polymerization, we next generated a recombinant SrtA mutant protein in which the DPW lid motif (residues 81–83) was mutated to GPG, hereafter referred to as "$SrtA^{2M}$." We succeeded in determining the crystal structure of $SrtA^{2M}$ at 1.85-Å resolution using crystallization conditions that differed from those used for the WT protein (*Materials and Methods*). In the electron density map for $SrtA^{2M}$, residues 80–86 that represent the lid were invisible. Presumably, the lid residue substitutions prevented contacts with the active site, causing the mutant lid to adopt a range of conformations. Remarkably, a second major difference between the two structures is the absence of interpretable electron density for the H1 helix in the $SrtA^{2M}$ lid mutant, which might be caused by flexibility around the hinge between helices H1 and H2 and by the absence of stabilizing interactions with neighboring molecules in the crystals of $SrtA^{2M}$.

To evaluate the involvement of the predicted catalytic residues and the lid in pilus assembly, corynebacterial cells harboring WT and its isogenic mutants were subjected to cell fractionation, and protein samples were immunoblotted with specific antibodies against SpaA (α-SpaA), the cognate substrate of SrtA that forms the pilus shaft (7, 13). As shown in Fig. 2*A*, SpaA polymers (P) were observed in cell wall fractions of the WT strain, but they were absent in the *srtA* deletion mutant, as previously reported (11). Ectopic expression of SrtA rescued the pilus assembly defect of the Δ*srtA* mutant (Fig. 2*A*, third lane), and Ala substitution of the catalytic residues C222, H160, and R231 abrogated pilus assembly (Fig. 2*A*, last three lanes). In control experiments, we demonstrated that none of these mutations affected the assembly of the SpaH-type pili, as expected (Fig. 2*B*) (9). Strikingly, the lid mutants are catalytically active in pilus polymerization. Like the WT and the complementing strains, strains expressing mutations in the DPW motif still produced pilus polymers (Fig. 2*C*), and immunoblotting analysis of the membrane fractions revealed no changes in the SrtA protein level when the lid was mutated (Fig. 2*D*).

To visualize these SpaA polymers, corynebacterial cells were immobilized on carbon-coated nickel grids, washed with water, and stained with 0.75% uranyl formate before viewing with an electron microscope. Because the parental strain NCTC 13129 produced short pili that were hardly detected (*SI Appendix*, Fig. S1*A*, WT), we constructed a multicopy vector expressing both SpaA and SrtA. Using this vector as a template, $SrtA^{2M}$ was generated by site-directed mutagenesis (*Materials and Methods*). The generated vectors were introduced into a corynebacterial double mutant lacking both *spaA* and *srtA* (Δ*spaA*/Δ*srtA*). Compared with the WT strain, overexpression of SpaA and SrtA resulted in increased production of long pili, as expected (*SI Appendix*, Fig. S1*A*, Δ*spaA*/Δ*srtA*/pSpaA-SrtA). Consistent with the above results, mutations in the DPW motif did not affect pilus assembly (*SI Appendix*, Fig. S1*A*, Δ*spaA*/Δ*srtA*/pSpaA-$SrtA^{2M}$). To confirm that these long pilus fibers are SpaA pili, the same set of strains was subjected to immunoelectron microscopy (IEM) (32), whereby immobilized cells were stained with α-SpaA, followed by staining with gold particles conjugated with IgG, before washing and staining with uranyl acetate. As shown in *SI Appendix*, Fig. S1*B*, SpaA-stained pili were detected in both strains producing WT SrtA and $SrtA^{2M}$, whereas short pili were observed in the WT strain, and no pili were observed in the Δ*spaA*/Δ*srtA* and Δ*spaA*/Δ*srtA*/$pSpaA^{K190A}$-$SrtA^{2M}$ strains.

Thus, the overall structure of the *C. diphtheriae* $SrtA^{WT}$ enzyme resembles class C sortases, or pilus-specific sortases, which possess a distinguishing feature of this class of enzymes, the lid region (16, 21). In agreement with previous studies (30, 31), the elimination of the lid's interaction at the catalytic pocket does not dramatically affect pilus assembly in vivo.

In Vitro Reconstitution of Archetypal *C. diphtheriae* SpaA Pilus Polymerization. Previous structural and biochemical studies of pilus-specific sortase enzymes in several streptococcal species indicate that the lid may modulate substrate entry into the active site (24, 25, 28, 30). We envisioned that a loss of lid closure might increase the accessibility of the active site. To test this hypothesis, we used the thiol-reactive reagent 4,4′-dithiodipyridine (DTDP) (33, 34) to probe the solvent accessibility of the catalytic Cys residue (C222). Disulfide exchange between thiol side chains of Cys residues and DTDP gives rise to 4-thiopyridone, which shows strong absorption at 324 nm (34). The recombinant proteins $SrtA^{WT}$ or $SrtA^{2M}$ (0.6 mg/mL) were rapidly mixed with 0.32 mM DTDP, and the rate of reaction between DTDP and C222 was monitored as an increase in absorbance at 324 nm. Time-dependent changes in absorbance were fit to single or double exponential equations to derive rates, as described in *Materials and Methods*. As shown in Fig. 3*A*, data for the $SrtA^{WT}$ enzyme best fit an equation with a single exponential rate of 2.17 ± 0.02/min. In contrast, data for the lid anchor mutant best fit an equation with two much faster exponential rates, 228 ± 7/min and 16 ± 3/min (Fig. 3*A*, *Inset*), which indicates that the catalytic C222 was readily accessible in this mutant. The two different rates may be due to slow exchange between two conformations in the mutant protein. If so, the conformation with the faster rate is the dominant form, since it represents 80% of the total change in absorbance.

The increased DTDP reactivity of the active site Cys residue in $SrtA^{2M}$ described above raises the possibility that the mutant enzyme may be able to assemble pili in vitro, which has been difficult to reconstitute so far for pilus-assembling sortases. We therefore sought to reconstitute pilus polymerization in vitro using various recombinant sortase enzymes and a soluble form of SpaA (residues 30–500), which is devoid of the N-terminal signal peptide and C-terminal membrane anchor domain (see diagram in Fig. 3*E*). Sortases were mixed with SpaA at a 1:3 molar ratio, and aliquots were removed for SDS/PAGE analysis and Coomassie staining at 0, 24, and 48 h. In the $SrtA^{WT}$ samples, a few new high molecular mass (HMM) bands were weakly observed after 24 and 48 h of incubation, one migrating between the 50 and 100 kDa markers and the others around 100 kDa (Fig. 3*B*, lanes $SrtA^{WT}$). Remarkably, with the $SrtA^{2M}$ enzyme, HMM SpaA polymers ($SpaA_P$) were abundantly formed within 24 h and increased further after 48 and 72 h (Fig. 3*B*, lanes $SrtA^{2M}$). Consistent with the results in Fig. 2*A*, the catalytically inactive enzyme in which C222 was replaced by Ala, $SrtA^{C222A}$, failed to produce any SpaA polymers (Fig. 3*B*, lanes $SrtA^{C222A}$). Intriguingly, removal of the H1 helix in the $SrtA^{2M}$ enzyme also abrogated pilus polymerization (Fig. 3*B*, lanes $^{\Delta}SrtA^{2M}$). The significance of this helix in the transpeptidation activity of sortase is discussed below.

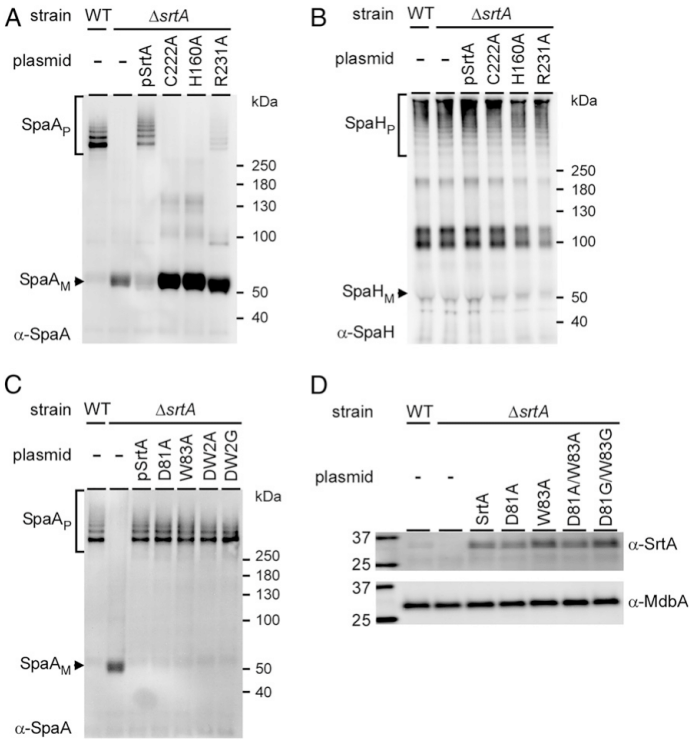


Fig. 2. Catalytic residues are required for pilus assembly in vivo. Cells of *C. diphtheriae* strains in equivalent numbers were subjected to cell fractionation. Protein samples from cell wall fractions (*A*–*C*) and protoplasts (*D*) were analyzed by immunoblotting with specific antibodies against SpaA (α-SpaA) (*A* and *C*), SpaH (α-SpaH) (*B*), and SrtA (α-SrtA) (*D*), with α-MdbA as a membrane-loading control. Pilus monomers (subscript M), polymers (subscript P), and molecular mass markers are indicated.

To visualize HMM SpaA polymers formed by $SrtA^{2M}$, the reaction mixtures after 72 h of incubation were subjected to electron microscopy, whereby aliquots were applied to nickel grids; bound proteins were washed and stained with 0.75% uranyl formate before viewing with an electron microscope. As shown in Fig. 3*C*, strands of SpaA polymers were observed in the reaction with the $SrtA^{2M}$ enzyme but not with an $SpaA^{K190A}$ mutant substrate defective in the nucleophilic attack, thus authenticating the visualization of Gram-positive pilus polymers synthesized in vitro. The synthesized pilus polymers had a width of ∼10 nm and a length ranging from ∼200–500 nm, equivalent to 25–62 subunits, with each protomer measuring about 8 nm (35). For comparison, pili produced by $SrtA^{2M}$ in vivo had widths ranging from 7.6–9.3 nm and lengths up to 2 μm (*SI Appendix*, Fig. S1*A*).

To determine if the recombinant $SrtA^{2M}$ enzyme faithfully catalyzes the pilus transpeptidation reaction, we determined whether the SpaA subunits in the HMM SpaA polymers were linked together via covalent Lys isopeptide bonds in which the Thr residue of the LPLT sorting signal was joined to the Lys residue within the pilin motif (13). Indeed, MS analysis of excised HMM SpaA polymer SDS/PAGE bands revealed the presence of an isopeptide bond between the carbonyl carbon of T494 and the sidechain amine of K190 (Fig. 3*D* and Table 3), as was observed in the native SpaA pili assembled in vivo (36). In line with the role of Lys190 in pilus polymerization, the SpaA mutant substrate in which K190 was replaced by Ala was unable to form polymers with the active $SrtA^{2M}$ enzyme (Fig. 3*B*, lanes $SpaA^{K190A}$). Remarkably, the MS data revealed the presence of $SrtA^{2M}$ and SpaA in the marked band migrating between the 50 and 100 kDa markers in both $SrtA^{WT}$ and $SrtA^{2M}$ samples (Fig. 3*B*, asterisks), suggesting that the enzyme is joined to the SpaA substrate via a labile thioacyl bond forming an acyl-enzyme heterodimer intermediate. To our astonishment, the MS analysis also revealed the presence of $SrtA^{2M}$ in the HMM SpaA polymer bands migrating at and above the 100-kDa marker (Fig. 3*B*, lanes $SrtA^{2M}$, bracket). The results are in agreement with our previous identification of the native acyl-enzyme intermediates formed between SrtA and SpaA polymers in vivo in *C. diphtheriae* as demonstrated by immunoblotting (37).

To further probe the mechanism of SpaA pilus assembly, we dissected the SpaA molecule into two components: the N-terminal domain (^{N}SpaA, residues 30–194, encompassing the pilin motif with the K190 nucleophile) and the C-terminal domain (^{C}SpaA, residues 350–500, containing the CWSS with the LPLTG motif) (Fig. 3*E*). The recombinant proteins were expressed in *E. coli* and purified (*SI Appendix*). When the two isolated domains were mixed at equal concentrations (300 μM) in a reaction with the lid-substituted $SrtA^{2M}$ enzyme (100 μM), a di-polypeptide conjugate was readily formed (Fig. 3*E*). Significantly, the presence of the expected Thr-Lys isopeptide in this conjugate was confirmed by MS (Table 3). Furthermore, control reactions demonstrated that SrtA containing the WT lid is markedly inactive in catalyzing cross-linking of the isolated domains (*SI Appendix*, Fig. S2). Together, our results

BIOCHEMISTRY

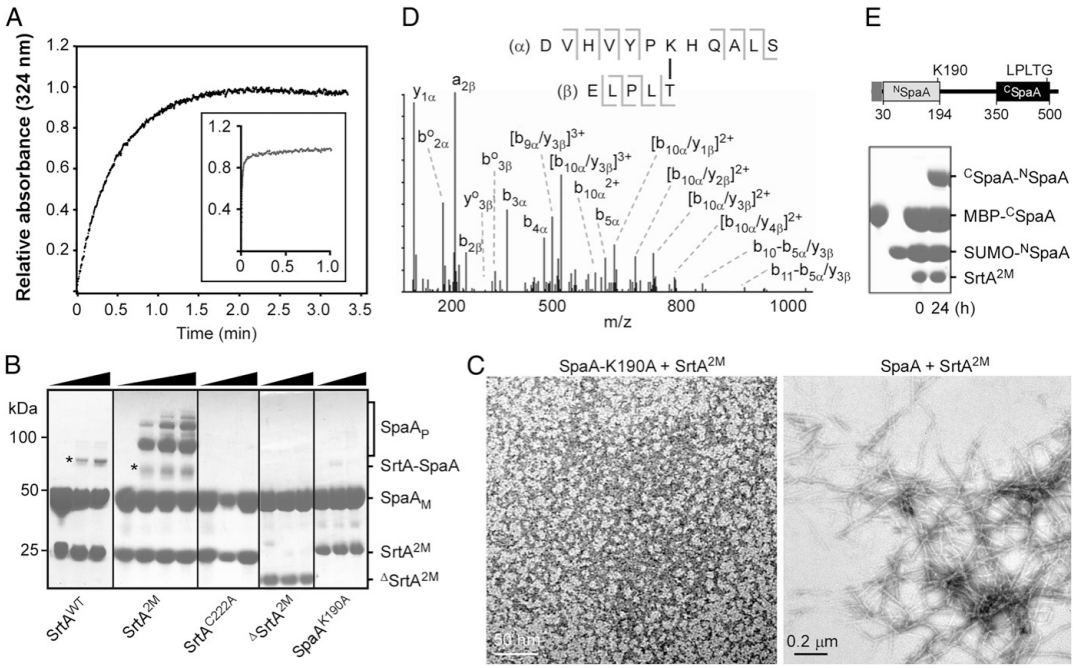


Fig. 3. Involvement of the lid in pilus polymerization in vitro. (*A*) Accessibility of the thiol group of the active site C222 residue in $SrtA^{WT}$ (main panel) or $SrtA^{2M}$ (*Inset*) enzymes was determined by stopped-flow experiments, whereby the reaction between the thiol-reactive reagent DTDP and C222 was monitored by absorbance at 324 nm. The experiments were performed in triplicate. (*B*) In vitro reconstitution of SpaA pilus polymerization was carried out at room temperature using various forms of recombinant SrtA and SpaA proteins at the molar ratio of 1:3. The reaction samples were analyzed by SDS/PAGE and Coomassie staining after 0, 24, and 48 h of incubation. Additional samples after 72 h of incubation were taken for the $SrtA^{2M}$ reactions (black triangles). SpaA monomers (M), polymers (P), and molecular mass markers are indicated. An SrtA–SpaA intermediate is marked with asterisks. (*C*) Protein samples from pilus polymerization reactions were analyzed by electron microscopy with negative staining using 0.75% uranyl formate. For comparison, recombinant $SrtA^{2M}$ and SpaA K190A proteins were included. (*D*) The isopeptide bond between residue T494 of the LPLTG motif and Lys residue K190 of the pilin motif in SpaA polymers in *B* was examined by MS/MS. Shown is the *m/z* tandem mass spectrum of the linked peptide (sequence shown in the *Inset*). (*E*) Fusion proteins between SUMO and the N-terminal SpaA domain (^{N}SpaA; residues 30–194) and between maltose-binding protein (MBP) and the C-terminal SpaA domain (^{C}SpaA; residues 350–500) were used with the $SrtA^{2M}$ enzyme in the in vitro pilus polymerization assay as described in *B*. The reaction samples were analyzed by SDS/PAGE and Coomassie staining after 24 h. The reactive Lys residue K190 and the LPXTG motif in the two domains are indicated.

support the concept that the lid in class C sortase functions in the molecular gating of substrate entry to the enzyme active site. Thus, we have demonstrated pilus polymerization in Gram-positive Actinobacteria in an in vitro reaction.

Structural Elements in a Sortase Required for Protein Polymerization. To gain insight into how SrtA joins the SpaA proteins together during polymerization, we performed molecular modeling of the ^{N}SpaA–SrtA–^{C}SpaA ternary complex in which the isopeptide bond is modeled using our previously determined crystal structures of SpaA [Protein Data Bank (PDB) ID code 3HR6] (36) and the isolated SrtA (PDB ID code 5K9A) proteins. We first generated a model of the SrtA–SpaA acyl-intermediate, juxtaposing the C terminus of the C-terminal SpaA domain with the active site C222 residue in SrtA. Because the crystal structure of SpaA lacks the CWSS that forms the acyl-intermediate with SrtA, we modeled the acyl-intermediate by placing the C-terminal domain of SpaA ~25 Å away from the active site Cys to accommodate the nine missing C-terminal residues that contain the CWSS. To construct the ternary complex, we then positioned the coordinates of the SpaA N-terminal domain near the acyl-intermediate to juxtapose the reactive Lys K190 of the pilin motif with the active site C222 residue (Fig. 4*A*). The resulting model of the ternary complex makes it readily obvious that the β7/β8 loop near the active site Cys residue and the N-terminal H1 helix in SrtA are in contact with the SpaA N-terminal domain, raising the possibility that these elements might play a role in recognizing the region of SpaA that houses the reactive Lys nucleophile. Strikingly, a primary sequence alignment of SrtA and other class C sortases indicates that they all contain a conserved TP(Y/L)XIN(S/T)H motif within the β7/β8 loop (*SI Appendix*, Fig. S3). This motif is clearly absent in other types of sortases that are known to attach proteins to the cell wall (class A, B, D, and E enzymes) but are unable to polymerize proteins. We thus postulated that the β7/β8 loop may play a role in conferring the polymerization activity in the class C enzymes.

In our model of the ternary reaction intermediate, the side chains of Y225, N228, and S229 within the TP(**Y**/L)XI**N**(**S**/T)H motif extend from the enzyme's surface in a position to contact ^{N}SpaA. To explore their possible roles in catalysis, we constructed a series of mutants of the lid-opened $SrtA^{2M}$ mutant enzyme in which each of these residues was individually replaced by Ala. The purified S229A and N228A mutant $SrtA^{2M}$ proteins were each defective in transpeptidation in vitro, as no isopeptide-linked SpaA–SpaA product was produced even after 48 h; the Y225A mutant protein had impaired transpeptidation activity as well, but

Table 3. MS analysis of synthetic SpaA pilus polymers

HPLC retention time, min	Calculated MH^+	Predicted peptides	Mass accuracy, ppm	Pilus assembly reaction
24.4	1947.0334	DVHVYPKHQALS :: ELPLT	4.0	Reaction with $SpaA_{30–500}$
26.5	1410.7627	DVHVYPK* :: ELPLT	2.8	Reaction with $SpaA_{30–500}$
22.9	1675.8802	DVHVYPKHQ* :: ELPLT	3.6	Reaction with $SpaA_{30–500}$
29.0	2336.2034	DVHVYPKHQALS :: NAGFELPLT	4.7	Reaction with $SpaA_{30–500}$
23.3	2400.2922	DVHVYPKHQALSEPVK :: ELPLT	2.9	Reaction with $SpaA_{30–500}$
ND	2485.2756	DQITLITCTPYAVNSHR :: ELPLT	ND	Reaction with $SpaA_{30–500}$
ND	2874.4455	DQITLITCTPYAVNSHR :: NAGFELPLT	ND	Reaction with $SpaA_{30–500}$
ND	1657.8717	DQITLITCTP* :: ELPLT	ND	Reaction with $SpaA_{30–500}$
ND	2047.0416	DQITLITCTP* :: NAGFELPLT	ND	Reaction with $SpaA_{30–500}$
24.5	1947.0334	DVHVYPKHQALS :: ELPLT	4.5	Reaction with ^{N}SpaA and ^{C}SpaA
26.5	1410.7627	DVHVYPK :: ELPLT	3.3	Reaction with ^{N}SpaA and ^{C}SpaA
22.9	1675.8802	DVHVYPKHQ :: ELPLT	3.5	Reaction with ^{N}SpaA and ^{C}SpaA
ND	2336.2034	DVHVYPKHQALS :: NAGFELPLT	ND	Reaction with ^{N}SpaA and ^{C}SpaA
ND	2400.2922	DVHVYPKHQALSEPVK :: ELPLT	ND	Reaction with ^{N}SpaA and ^{C}SpaA
ND	2485.2756	DQITLITCTPYAVNSHR :: ELPLT	ND	Reaction with ^{N}SpaA and ^{C}SpaA
ND	2874.4455	DQITLITCTPYAVNSHR :: NAGFELPLT	ND	Reaction with ^{N}SpaA and ^{C}SpaA
ND	1657.8717	DQITLITCTP* :: ELPLT	ND	Reaction with ^{N}SpaA and ^{C}SpaA
ND	2047.0416	DQITLITCTP* :: NAGFELPLT	ND	Reaction with ^{N}SpaA and ^{C}SpaA
36.5	1023.6448	PKLI :: ELPLT	11.9	Reaction with ^{C}SpaA and SpaB

MH^+, the mass of the singly protonated species; ND, not determined.
*Not expected cleavage sites.

to a lesser extent than the S229A and N228A mutants (Fig. 4*C*). Recall that the removal of the H1 helix in the $SrtA^{2M}$ enzyme also abrogates pilus polymerization (Fig. 3*B*, lanes $^{\Delta}SrtA^{2M}$). We have determined that the absence of the H1 helix does not cause the protein to unfold, since the ^{1}H-^{15}N heteronuclear single-quantum correlation (HSQC) spectra of $SrtA^{2M}$ and $^{\Delta}SrtA^{2M}$ are generally similar (*SI Appendix*, Fig. S4). We conclude that specific residues within the β7/β8 loop and the presence of the H1 helix form a functionally important contact surface with ^{N}SpaA. This is supported by experiments with an $SrtA^{2M}$ mutant harboring the N165A substitution in the proximal β4/β5 loop, which showed that this mutant retained nearly WT activity (Fig. 4*C*; lanes $SrtA^{N165A}$).

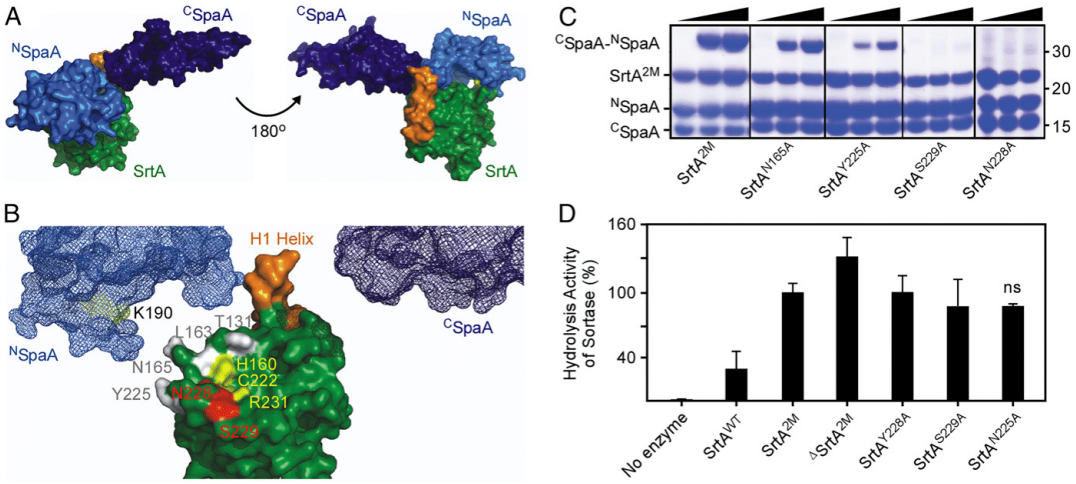


Fig. 4. Structural modeling reveals SrtA residues critical for transpeptidation activities. (*A*) An SrtA–SpaA pilin polymerase attack complex was visualized and assembled using PyMOL. Shown is an SpaA molecule splitting into its two domains, ^{N}SpaA (light blue) and ^{C}SpaA (dark blue). The SrtA enzyme is shown in green, and the H1 helix potentially bridging interactions between the two SpaA domains is seen in orange. (*B*) The detailed locations of the SrtA catalytic triad (marked in yellow) and the surrounding residues at the active site of the pilin polymerase attack complex are shown. The N-terminal H1 helix bridges the two reactive domains of SpaA and potentially facilitates interactions for the formation of the SrtA–SpaA polymerization attack complex. (*C*) Transpeptidation activity of $SrtA^{2M}$ and its variants (N165A, Y225A, N228A, and S229A) was determined in the pilus polymerization assay described in Fig. 3*E*, using domain substrates ^{N}SpaA and ^{C}SpaA. Protein samples were analyzed by SDS/PAGE and Coomassie staining after 24 h. The ligated product ^{C}SpaA–^{N}SpaA, sortase enzymes, substrates, and molecular markers are indicated. (*D*) Hydrolysis activity of SrtA enzymes was determined by an HPLC-based assay. WT or mutant SrtA (50 μM) was incubated with 500 μM KNAGFELPLTGGSGRI ($SpaA^{pep}$) in a 100-μL assay at 37 °C for 48 h. Reaction products were monitored and separated using HPLC at an absorbance of 215 nm. The peak fractions were collected and identified by MALDI-TOF-MS. The hydrolysis activity by $SrtA^{2M}$ is set as 100%. The results are presented as the average of three independent experiments; error bars indicate SDs; ns, not significant.

The model of the ternary complex raises the possibility that critical residues in the β7/β8 loop and the H1 helix may be required only for nucleophile recognition during the transpeptidation reaction but not for the other step of catalysis in which the LPXTG sorting motif is cleaved to form the thioacyl enzyme–substrate intermediate (4, 38). To test this hypothesis, we determined the importance of these structural elements in thioacyl-intermediate formation, using an established HPLC-based assay (39, 40) and an SpaA-derived peptide KNAGFELPLTGGSGRI ($SpaA^{pep}$) as the substrate. The enzymes and the $SpaA^{pep}$ substrate were mixed at a 1:10 molar ratio, and the loss of the intact peptide was monitored by HPLC with the hydrolysis activity of $SrtA^{2M}$ set as 100%. Consistent with a selective role in nucleophile recognition, none of the mutants exhibited any significant defect in cleaving the LPXTG motif (Fig. 4*D*). Importantly, the $^{\Delta}SrtA^{2M}$ enzyme, which was inactive in the pilus polymerization assay (Fig. 3*B*), cleaved the $SpaA^{pep}$ substrate with an efficiency comparable to that of the activated $SrtA^{2M}$ mutant enzyme (Fig. 4*D*). Under these conditions, the hydrolysis kinetics of the $SrtA^{2M}$ and $^{\Delta}SrtA^{2M}$ enzymes displayed a comparable V_{max} of 2.3 ± 0.2 and 3.3 ± 0.8 μM/h, respectively, unlike that of $SrtA^{WT}$, which was significantly reduced (0.6 ± 0.1 μM/h). These results prompt us to propose that the conserved TP(Y/L)XIN(S/T)H motif within the β7/β8 loop is a hallmark feature of the class C sortases that enables molecular recognition of the pilin motif Lys nucleophile in their cognate substrates. The H1 helix appears to play a similar role; however, it is not well conserved in class C sortases.

SrtA-Catalyzed Pilus Polymerization Is Terminated by SpaB. Our previous studies suggest that SpaB acts as a molecular switch that terminates pilus polymerization by incorporating into the pilus polymer as the terminal subunit, and this reaction requires the Lys residue K139 present on SpaB (11), which is then anchored to the cell wall by the housekeeping sortase SrtF (17). The SpaA polymer is presumed to be linked to the terminal SpaB via an isopeptide bond formed between the Thr residue of the SpaA LPXTG motif and K139 (Fig. 5*A*). To examine if this is the case, we produced a recombinant SpaB protein (residues 25–180), which lacked the N-terminal signal peptide and the hydrophobic domain and the C-terminal charged tail of the CWSS but contained the LAFTG motif. As a control, an SpaB mutant protein with the K139A substitution mutation was also generated. These recombinant proteins, expressed in and purified from *E. coli*, were then used in the pilus polymerization assay and analyzed by SDS/PAGE and Coomassie staining as described in Fig. 3*B*. In the presence of the $SrtA^{2M}$ enzyme, recombinant SpaA protein was polymerized into HMM species as expected (Fig. 5*B*, first two lanes). Remarkably, when SpaB was added into this reaction, the formation of SpaA polymers was significantly reduced, and the SpaA–SpaB dimer accumulated (Fig. 5*B*, third and fourth lanes, asterisk); it is important to note that while some trimeric forms of SpaA were observed, $SrtA^{2M}$ was unable to further polymerize SpaA pilins in the presence of WT SpaB, although SpaA substrates were abundantly available (Fig. 5*B*, lane 4). This suggests that the SpaB K139 may be more nucleophilic than the SpaA K190 or that SpaB K139 may have a higher affinity and the ability to outcompete SpaA for pilin cross-linking reactions. The fact that SpaA pilus polymerization catalyzed by $SrtA^{2M}$ was not affected in the presence of the SpaB K139A mutant protein is consistent with K139 as the nucleophile in the cross-linking reaction (Fig. 5*B*, last two lanes).

The observation that the SpaA–SpaB dimer is the predominant form of pilin conjugates produced by $SrtA^{2M}$ in the presence of SpaB (Fig. 5*B*, lane 4) prompted us to test whether the pilin motif is dispensable for SrtA-catalyzed SpaA–SpaB conjugation. To examine if this is the case, the recombinant protein ^{C}SpaA (see Fig. 4*C*), lacking the pilin motif, was used in place of recombinant SpaA. The reaction was performed as described in Fig. 4*C*. Indeed, after 24 h a band corresponding to the size of a truncated SpaA–SpaB conjugate was observed (Fig. 5*C*). To corroborate this and analyze the linkage between SpaA and SpaB, this band was excised for MS analysis (see Fig. 3*D* for methods). MS/MS data confirmed the isopeptide bond between the SpaA T494 residue and the SpaB K139 residue (Fig. 5*D* and Table 3). Clearly, our results demonstrate that SpaB is a termination factor for pilus polymerization.

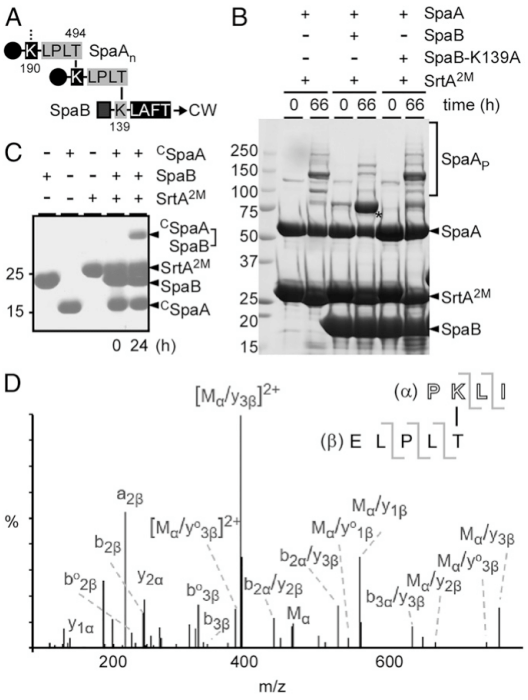


Fig. 5. SrtA-catalyzed pilus polymerization is terminated by the pilus base SpaB pilin. (*A*) Depicted is an SpaA polymer ($SpaA_n$) with individual subunits cross-linked by an isopeptide bond between K190 and T494. The SpaA polymer is linked to the terminal SpaB via an isopeptide bond between the T494 residue of SpaA and the K139 residue of SpaB, which in turn is covalently attached to the bacterial peptidoglycan (CW). (*B*) A pilus polymerization termination assay was performed with $SrtA^{2M}$ enzyme and SpaA substrate (2:1 molar ratio) in the presence (same concentration as SpaA) or absence of SpaB or SpaB with the K139A mutation. The reactions were stopped after 66 h by the addition of SDS-containing sample buffer, and protein samples were analyzed by SDS/PAGE and Coomassie staining. The asterisk indicates an SpaA–SpaB dimer. Molecular mass markers (kDa) are shown. (*C*) ^{C}SpaA (residues 350–500), SpaB, and $SrtA^{2M}$ were used in the pilus polymerization termination assay, with substrates and enzyme at a 3:1 molar ratio. Protein samples were analyzed by SDS/PAGE and Coomassie staining after 24-h incubation. (*D*) A gel band corresponding to a ^{C}SpaA–SpaC dimer was excised for tryptic digestion and MS/MS. Shown is the MS/MS spectrum, which revealed the presence of the isopeptide bond formed between T494 and K139.

In conclusion, we report here the high-resolution crystal structures of the *C. diphtheriae* pilus-specific sortase SrtA enzyme ($SrtA^{WT}$) and a mutant form of the enzyme with mutations in the lid region ($SrtA^{2M}$) and through these illuminate some of the basic features of the sortase that functions to polymerize pilus proteins in Gram-positive bacteria. The structure of the WT enzyme displayed a characteristic "closed" configuration of a class C sortase with its catalytic site occluded by a molecular lid (Fig. 1). By introducing specific amino acid substitutions within

the lid, we were able to generate an enzyme whose catalytic pocket displayed an open conformation with no other major perturbations detected in the atomic structure. The functional importance of these two states of the enzyme was demonstrated by our ability to reconstitute a robust pilus polymerization reaction in vitro using the cognate shaft pilin. While the WT form could not polymerize the shaft pilin in a reaction, the opened-lid version of the enzyme is highly active. We showed that the activated enzyme was able to recognize the sorting signal, form the relevant acyl-enzyme intermediate, recognize the pilin motif Lys residue, and catalyze isopeptide bond formation conjugating Lys of one pilin protomer with Thr of another protomer (Fig. 3). We then utilized structural modeling to identify specific structural elements conserved in a pilus-specific sortase which are important for catalyzing the transpeptidation reaction in vitro (Fig. 4). We also provide additional in vitro evidence in support of SpaB being a terminator of pilus polymerization (Fig. 5).

Given that the $SrtA^{2M}$ enzyme is able to catalyze pilus polymerization and SpaB incorporation in vitro, although pilus assembly in vivo is not apparently altered when the mutant enzyme is present, we surmise that the SrtA lid might play some form of a modulatory role in pilus polymerization and termination in vivo. For instance, as SpaB is a preferred substrate of the housekeeping sortase SrtF (17), charging SrtF with SpaB triggers an SrtF–SrtA interaction. This interaction could potentially alter the configuration of the SrtA lid, allowing SrtA-mediated entry of SpaB to the pilus base and subsequently transferring the pilus polymer to SrtF to complete the cell wall-anchoring step. Future experiments may be designed to examine the lid dynamics in the presence or absence of SpaB or SrtF.

Importantly, we have shown here that the separated domains of the pilin, one containing the pilin motif and the other containing the sorting motif, could be ligated efficiently to produce a di-polypeptide conjugate containing the Lys–Thr isopeptide bond. This provides a powerful protein ligation platform for engineering designer proteins that is mechanistically different from the "sortagging" technology developed with the archetypal *S. aureus* sortase which normally functions to cross-link surface proteins to the bacterial cell wall but does not polymerize proteins (41, 42). We envision that the surface display of protein polymers, protein labeling of living cells, and protein ligation are a few examples of many potential biotechnological and biological applications of this enzyme.

Materials and Methods

Bacterial Strains, Plasmids, and Media. Bacterial strains and plasmids used in this study are listed in *SI Appendix*, Table S1. *SI Appendix, SI Materials and Methods* contains information regarding recombinant plasmids, protein purification, protein crystallization, and structure determination. *C. diphtheriae* strains were grown in Heart Infusion (HI) broth (Becton Dickinson) or on HI agar plates at 37 °C. When needed, kanamycin was added at a concentration of 25 μg/mL. *E. coli* DH5α and BL21 (DE3), used for cloning and protein expression and purification, respectively, were grown in either Luria–Bertani or 2× YT broth (Sigma-Aldrich) at 37 °C in the presence of ampicillin at 100 μg/mL.

In Vitro Reconstitution of Pilus Polymerization. In vitro reactions were carried out at room temperature, and proteins were dissolved in assay buffer [50 mM Tris·HCl (pH 8.0), 300 mM NaCl, 1 mM DTT]. All reactions used a fixed 100-μM concertation of SrtA enzyme and 300-μM SpaA substrate (either full length or each individual domain). Reactions were stirred gently by continuous rotation. Aliquots were taken at 0 h, 24 h, 48 h, and 72 h, and reactions were quenched by the addition of two volumes of SDS loading dye. For a pilus termination assay, SpaB was used at the same concentration as SpaA, whereas the molar concentrations of $SrtA^{2M}$ enzyme and SpaA substrate followed 3:1 or 2:1 ratios. The reactions were performed in 24 or 66 h, respectively.

Probing Accessibility of the SrtA Active Site. Reaction rates of DTDP and SrtA proteins via the Cys C222 residue were determined by stopped-flow experiments, which were performed at 23 °C using an Applied Photophysics Ltd. Model SX.18 MV sequential stopped-flow spectrofluorimeter with a 150 W Xe/Hg lamp and a dead time of 1.7 ms. All triplicate reactions were carried out in reaction buffer [50 mM 3-(*N*-morpholino)propanesulfonic acid (Mops), 200 mM KCl, 1 mM EDTA, pH 7.5]. Absorbance was monitored at 324 nm after solutions were rapidly mixed in syringe A, containing 0.6 mg/mL protein, and syringe B, containing 0.32 mM DTDP. Reaction rates (k) were derived by fitting data to the following equations with one (Eq. **1**) or two (Eq. **2**) rates:

$$A = A_{max} * \left(1 - e^{-kt}\right) \quad [1]$$

or

$$A = A_{max1} * \left(1 - e^{-k1t}\right) + A_{max2} * \left(1 - e^{-k2t}\right), \quad [2]$$

where A is absorbance at 324 nm at time t, and A_{max} is the maximum absorbance.

Cell Fractionation and Western Blotting. Cell fractionation and Western blotting were performed according to published procedures with some modifications (43, 44). Briefly, midlog-phase cultures of *C. diphtheriae* strains grown at 37 °C were normalized to an OD_{600} of 1.0 and were subjected to cell wall protein extraction using mutanolysin (300 U/mL). Protein samples obtained from culture medium (S) and cell wall (W) were trichloroacetic acid precipitated and acetone washed. The protoplasts after the cell wall extraction were used for analysis of cell membrane-bound proteins. Protein samples were resuspended in SDS sample buffer containing 3% urea and were heated at 100 °C for 10 min before SDS/PAGE analysis using 3–12% or 3–20% Tris-Gly gradient gels. Detection of proteins was performed by immunoblotting with specific antibodies (1:20,000 for α-SpaA; 1:4,000 for α-SpaH; 1:5,000 for α-MdbA; and 1:4,000 for α-SrtA).

MS of Pilus Polymers. Protein digestion and isopeptide bond identification were performed according to previous protocols (36, 45). Specifically, proteins entrapped in gel bands were reduced with 10 mM DTT (Sigma) at 60 °C for 1 h and then were alkylated with 50 mM iodoacetamide (Sigma) at 45 °C for a few minutes in the dark. These reduction and alkylation steps were skipped for the acyl-intermediate samples. Samples were digested with 200 ng trypsin (Thermo Scientific) at 37 °C overnight. At the end of trypsin digestion, 200 ng of Asp-N endoproteinase (Thermo Scientific) was added for another overnight incubation. Digested peptides were extracted from the gel bands in 50% acetonitrile/49.9% water/0.1% TFA and were cleaned with C18 StageTip (46) before MS analysis.

Digested peptides were separated on an EASY-Spray column (25 cm × 75 μm i.d., PepMap RSLC C18, 2 μm; Thermo Scientific) connected to an EASY-nLC 1000 nUPLC (Thermo Scientific) using a gradient of 5–35% acetonitrile in 0.1% formic acid and a flow rate of 300 nL/min for 30 min. Tandem mass spectra were acquired in a data-dependent manner with an Orbitrap Q Exactive mass spectrometer (Thermo Fisher Scientific) interfaced to a nanoelectrospray ionization source.

The raw MS/MS data were converted into MGF format by Thermo Proteome Discoverer 1.4 (Thermo Scientific). In-house programs to search for the isopeptides were used for two different approaches. The first approach was performed as previously described (36) and was used to calculate the masses of predicted peptides containing the isopeptide linkage to guide the search. The second approach was based on published (36) and our own observations of the presence of ions specific for the fragments of ELPLT (*m/z* 215.138, 225.122, 243.132, 294.180, 312.190, 322.174, and 340.186). The in-house programs sifted through tens of thousands of mass spectra searching specifically for this information and extracted MS/MS spectra for further analyses.

Determination of SrtA Hydrolysis by an HPLC-Based Assay. In vitro hydrolysis reactions were performed based on the method developed by Kruger et al. (47). WT or mutant SrtA (50 μM) was incubated with 500 μM KNAGFELPLTGGSGRI ($SpaA^{pep}$) in 100-μL reactions at 37 °C for 24 h. The reactions were quenched by adding 50 μL of 1 M HCl and were injected onto a Waters XBridge Peptide BEH C18 reversed-phase HPLC column. Peptides were eluted by applying a gradient from 5–51% acetonitrile (in 0.1% TFA) over 25 min at a flow rate of 1 mL/min. Elution of the peptides was monitored by absorbance at 215 nm. Peak fractions were collected, and their identities were confirmed by MALDI-TOF-MS.

Electron Microscopy. For visualization of in vitro pilus polymers, pilus polymerization reactions were diluted in half with water and 7-μL aliquots were applied onto carbon-coated nickel grids, washed five times with distilled water, and stained with 0.75% uranyl formate for 2 min before viewing by a JEOL JEM-1400 electron microscope.

For visualization of pili produced by corynebacterial cells, electron microscopy and IEM were performed according to a published protocol (48).

Briefly, corynebacterial cells grown on HI agar plates were washed and suspended in PBS. Seven-microliter aliquots of the cell suspension were applied onto nickel grids, washed, and stained with 0.75% uranyl formate before viewing by an electron microscope. For IEM, cells were stained with α-SpaA (1:100 dilution), followed by staining with 12-nm gold particles conjugated to IgG, before staining with 1% uranyl acetate.

To estimate the dimension of pili, ImageJ (https://imagej.nih.gov/ij/) was employed. Twenty-five measurements were performed at different locations of pili for each strain. Statistical analysis was performed by GraphPad Prism.

ACKNOWLEDGMENTS. We thank members of the R.T.C. and H.T.-T. laboratories for critical review and discussion of the manuscript and members of the Structural Biology Center at Argonne National Laboratory for help in conducting X-ray diffraction data collection. This work was supported by Department of Health and Human Services, NIH National Institute of Allergy and Infectious Diseases Award AI52217 (to R.T.C.) and in part by Center of Structural Genomics of Infectious Diseases Contracts HHSN272201200026C and HHSN272201700060C (to A.J.), National Institute of General Medical Sciences Grant GM103479 (to J.A.L.), and National Institute of Dental and Craniofacial Research Awards DE017382 and DE025015 (to H.T.-T.).

1. Kline KA, Dodson KW, Caparon MG, Hultgren SJ (2010) A tale of two pili: Assembly and function of pili in bacteria. *Trends Microbiol* 18:224–232.
2. Hospenthal MK, Costa TRD, Waksman G (2017) A comprehensive guide to pilus biogenesis in Gram-negative bacteria. *Nat Rev Microbiol* 15:365–379.
3. Thanassi DG, Bliska JB, Christie PJ (2012) Surface organelles assembled by secretion systems of Gram-negative bacteria: Diversity in structure and function. *FEMS Microbiol Rev* 36:1046–1082.
4. Ton-That H, Schneewind O (2004) Assembly of pili in Gram-positive bacteria. *Trends Microbiol* 12:228–234.
5. Telford JL, Barocchi MA, Margarit I, Rappuoli R, Grandi G (2006) Pili in Gram-positive pathogens. *Nat Rev Microbiol* 4:509–519.
6. Mandlik A, Swierczynski A, Das A, Ton-That H (2008) Pili in Gram-positive bacteria: Assembly, involvement in colonization and biofilm development. *Trends Microbiol* 16: 33–40.
7. Ton-That H, Schneewind O (2003) Assembly of pili on the surface of *Corynebacterium diphtheriae*. *Mol Microbiol* 50:1429–1438.
8. Rogers EA, Das A, Ton-That H (2011) Adhesion by pathogenic corynebacteria. *Adv Exp Med Biol* 715:91–103.
9. Swierczynski A, Ton-That H (2006) Type III pilus of corynebacteria: Pilus length is determined by the level of its major pilin subunit. *J Bacteriol* 188:6318–6325.
10. Gaspar AH, Ton-That H (2006) Assembly of distinct pilus structures on the surface of *Corynebacterium diphtheriae*. *J Bacteriol* 188:1526–1533.
11. Mandlik A, Das A, Ton-That H (2008) The molecular switch that activates the cell wall anchoring step of pilus assembly in Gram-positive bacteria. *Proc Natl Acad Sci USA* 105:14147–14152.
12. Mandlik A, Swierczynski A, Das A, Ton-That H (2007) *Corynebacterium diphtheriae* employs specific minor pilins to target human pharyngeal epithelial cells. *Mol Microbiol* 64:111–124.
13. Ton-That H, Marraffini LA, Schneewind O (2004) Sortases and pilin elements involved in pilus assembly of *Corynebacterium diphtheriae*. *Mol Microbiol* 53:251–261.
14. Kang HJ, Coulibaly F, Clow F, Proft T, Baker EN (2007) Stabilizing isopeptide bonds revealed in Gram-positive bacterial pilus structure. *Science* 318:1625–1628.
15. Navarre WW, Schneewind O (1999) Surface proteins of Gram-positive bacteria and mechanisms of their targeting to the cell wall envelope. *Microbiol Mol Biol Rev* 63: 174–229.
16. Siegel SD, Liu J, Ton-That H (2016) Biogenesis of the Gram-positive bacterial cell envelope. *Curr Opin Microbiol* 34:31–37.
17. Swaminathan A, et al. (2007) Housekeeping sortase facilitates the cell wall anchoring of pilus polymers in *Corynebacterium diphtheriae*. *Mol Microbiol* 66:961–974.
18. Mazmanian SK, Liu G, Ton-That H, Schneewind O (1999) *Staphylococcus aureus* sortase, an enzyme that anchors surface proteins to the cell wall. *Science* 285:760–763.
19. Ton-That H, Liu G, Mazmanian SK, Faull KF, Schneewind O (1999) Purification and characterization of sortase, the transpeptidase that cleaves surface proteins of *Staphylococcus aureus* at the LPXTG motif. *Proc Natl Acad Sci USA* 96:12424–12429.
20. Dramsi S, Trieu-Cuot P, Bierne H (2005) Sorting sortases: A nomenclature proposal for the various sortases of Gram-positive bacteria. *Res Microbiol* 156:289–297.
21. Spirig T, Weiner EM, Clubb RT (2011) Sortase enzymes in Gram-positive bacteria. *Mol Microbiol* 82:1044–1059.
22. Jacobitz AW, Kattke MD, Wereszczynski J, Clubb RT (2017) Sortase transpeptidases: Structural biology and catalytic mechanism. *Adv Protein Chem Struct Biol* 109: 223–264.
23. Khare B, V L Narayana S (2017) Pilus biogenesis of Gram-positive bacteria: Roles of sortases and implications for assembly. *Protein Sci* 26:1458–1473.
24. Manzano C, Izoré T, Job V, Di Guilmi AM, Dessen A (2009) Sortase activity is controlled by a flexible lid in the pilus biogenesis mechanism of Gram-positive pathogens. *Biochemistry* 48:10549–10557.
25. Khare B, Fu ZQ, Huang IH, Ton-That H, Narayana SV (2011) The crystal structure analysis of group B *Streptococcus* sortase C1: A model for the "lid" movement upon substrate binding. *J Mol Biol* 414:563–577.
26. Persson K (2011) Structure of the sortase AcSrtC-1 from *Actinomyces oris*. *Acta Crystallogr D Biol Crystallogr* 67:212–217.
27. Neiers F, et al. (2009) Two crystal structures of pneumococcal pilus sortase C provide novel insights into catalysis and substrate specificity. *J Mol Biol* 393:704–716.
28. Jacobitz AW, et al. (2016) The "lid" in the *Streptococcus pneumoniae* SrtC1 sortase adopts a rigid structure that regulates substrate access to the active site. *J Phys Chem B* 120:8302–8312.
29. Manzano C, et al. (2008) Sortase-mediated pilus fiber biogenesis in *Streptococcus pneumoniae*. *Structure* 16:1838–1848.
30. Cozzi R, et al. (2011) Structure analysis and site-directed mutagenesis of defined key residues and motives for pilus-related sortase C1 in group B *Streptococcus*. *FASEB J* 25: 1874–1886.
31. Wu C, et al. (2012) Structural determinants of *Actinomyces* sortase SrtC2 required for membrane localization and assembly of type 2 fimbriae for interbacterial coaggregation and oral biofilm formation. *J Bacteriol* 194:2531–2539.
32. Sanchez BC, Chang C, Wu C, Tran B, Ton-That H (2017) Electron transport chain is biochemically linked to pilus assembly required for polymicrobial interactions and biofilm formation in the Gram-positive actinobacterium *Actinomyces oris*. *MBio* 8: e00399-17.
33. Putkey JA, et al. (1997) Fluorescent probes attached to Cys 35 or Cys 84 in cardiac troponin C are differentially sensitive to Ca(2+)-dependent events in vitro and in situ. *Biochemistry* 36:970–978.
34. Epps DE, Vosters AF (2002) The essential role of a free sulfhydryl group in blocking the cholesteryl site of cholesteryl ester transfer protein (CETP). *Chem Phys Lipids* 114: 113–122.
35. Echelman DJ, et al. (2016) CnaA domains in bacterial pili are efficient dissipaters of large mechanical shocks. *Proc Natl Acad Sci USA* 113:2490–2495.
36. Kang HJ, Paterson NG, Gaspar AH, Ton-That H, Baker EN (2009) The *Corynebacterium diphtheriae* shaft pilin SpaA is built of tandem Ig-like modules with stabilizing isopeptide and disulfide bonds. *Proc Natl Acad Sci USA* 106:16967–16971.
37. Guttilla IK, et al. (2009) Acyl enzyme intermediates in sortase-catalyzed pilus morphogenesis in Gram-positive bacteria. *J Bacteriol* 191:5603–5612.
38. Ton-That H, Marraffini LA, Schneewind O (2004) Protein sorting to the cell wall envelope of Gram-positive bacteria. *Biochim Biophys Acta* 1694:269–278.
39. Aulabaugh A, et al. (2007) Development of an HPLC assay for *Staphylococcus aureus* sortase: Evidence for the formation of the kinetically competent acyl enzyme intermediate. *Anal Biochem* 360:14–22.
40. Ton-That H, Mazmanian SK, Faull KF, Schneewind O (2000) Anchoring of surface proteins to the cell wall of *Staphylococcus aureus*. Sortase catalyzed in vitro transpeptidation reaction using LPXTG peptide and NH(2)-Gly(3) substrates. *J Biol Chem* 275:9876–9881.
41. Antos JM, et al. (2017) Site-specific protein labeling via sortase-mediated transpeptidation. *Curr Protoc Protein Sci* 89:15.13.11–15.13.19.
42. Proft T (2010) Sortase-mediated protein ligation: An emerging biotechnology tool for protein modification and immobilisation. *Biotechnol Lett* 32:1–10.
43. Chang C, Mandlik A, Das A, Ton-That H (2011) Cell surface display of minor pilin adhesins in the form of a simple heterodimeric assembly in *Corynebacterium diphtheriae*. *Mol Microbiol* 79:1236–1247.
44. Reardon-Robinson ME, et al. (2015) A thiol-disulfide oxidoreductase of the Gram-positive pathogen *Corynebacterium diphtheriae* is essential for viability, pilus assembly, toxin production and virulence. *Mol Microbiol* 98:1037–1050.
45. Thevis M, Ogorzalek Loo RR, Loo JA (2003) In-gel derivatization of proteins for cysteine-specific cleavages and their analysis by mass spectrometry. *J Proteome Res* 2: 163–172.
46. Rappsilber J, Mann M, Ishihama Y (2007) Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat Protoc* 2:1896–1906.
47. Kruger RG, Dostal P, McCafferty DG (2004) Development of a high-performance liquid chromatography assay and revision of kinetic parameters for the *Staphylococcus aureus* sortase transpeptidase SrtA. *Anal Biochem* 326:42–48.
48. Chang C, Huang IH, Hendrickx AP, Ton-That H (2013) Visualization of Gram-positive bacterial pili. *Methods Mol Biol* 966:77–95.

***In vitro* reconstitution of sortase-catalyzed pilus polymerization reveals structural elements involved in pilin crosslinking**

Chungyu Chang[a,1], Brendan R. Amer[b,c1], Jerzy Osipiuk[d,e], Scott A. McConnell[b,c], I-Hsiu Huang[f], Van Hsieh[b,c], Janine Fu[b,c], Hong H. Nguyen[b,c], John Muroski[b,c], Erika Flores[a], Rachel R. Ogorzalek Loo[b,c], Joseph A. Loo[b,c], John A. Putkey[g], Andrzej Joachimiak[d,e], Asis Das[h], Robert T. Clubb [b,c,2], and Hung Ton-That[a,2]

[a]*Department of Microbiology & Molecular Genetics, University of Texas Health Science Center, Houston, TX 77030;* [b]*Department of Chemistry and Biochemistry*, *University of California, Los Angeles, CA 90095;* [c]*University of California, Los Angeles-US Department of Energy Institute of Genomics and Proteomics, University of California, Los Angeles, CA 90095;* [d]*Center for Structural Genomics of Infectious Diseases, Argonne National Laboratory, Argonne, IL 60439;* [e]*Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, IL 60637;* [f]*Department of Microbiology and Immunology, College of Medicine, National Cheng Kung University, Tainan 701, Taiwan;* [g]*Department of Biochemistry and Molecular Biology, University of Texas Health Science Center, Houston, TX 77030; and* [h]*Department of Molecular Biology and Biophysics, University of Connecticut Health Center, Farmington, CT 06030*

[1]Contributed equally to this work

[2]To whom correspondence should be addressed:

Robert T. Clubb
Department of Chemistry and Biochemistry, University of California, Los Angeles, 611 Charles Young Drive East, Los Angeles, CA 90095, USA
Tel. (+1) 310 206 2334; Fax (+1) 310 206 4779; Email: rclubb@mbi.ucla.edu

Hung Ton-That
Department of Microbiology and Molecular Genetics, University of Texas Health Science Center, 6431 Fannin Street, R224/MSE, Houston, TX 77030, USA.
Tel. (+1) 713 500 5468; Fax (+1) 713 500 5499; E-mail: ton-that.hung@uth.tmc.edu

Running title: *In vitro assembly of Gram-positive pilus polymers*

Keywords: *Corynebacterium diphtheriae*; sortase; pilus polymerization; protein ligation; transpeptidation

www.pnas.org/cgi/doi/10.1073/pnas.1800954115

SUPPORTING INFORMATION

SI Materials and Methods

Recombinant Plasmids

pSUMO-SrtA^{WT} and its derivatives – To generate a recombinant plasmid expressing His-tagged SrtA^{WT} (residues 37 to 257), the *srtA* gene sequence without N-terminal signal peptide and C-terminal membrane spanning domains was PCR-amplified from the genomic DNA of *C. diphtheriae* NCTC 13129 with appropriate primers (Table S2) and inserted into the pE-SUMO (LifeSensors) expression vector using the Gibson assembly method (New England BioLabs). pSUMO-SrtA^{WT} was used as a template to generate D81A and W83A mutations (pSUMO-SrtA^{2M}), as well as Y225A, S229A or N228A mutation, using site-directed mutagenesis carried out by QuickChange method (Agilent) (see Table S2 for primers). Resulting plasmids were then transformed into XL10 for amplification prior to DNA sequence confirmation. Similarly, pSUMO-$^{\Delta}$SrtA^{2M} were generated using pSUMO-SrtA2M as a template, in which H1 helix (residues 37-54) was removed. The resulting plasmid was introduced into *E. coli* BL21 (DE3) after verification by DNA sequencing.

pMCSG-SrtA^{WT} and its derivatives – For protein crystallization, the same *srtA* fragment as the above was cloned into the pMCSG7 expression vector by ligation-independent cloning (LIC) as previously reported (1). The resulting plasmid was introduced into *E. coli* DH5α for selection and DNA sequencing, and then *E. coli* BL21 (DE3). To generate pMCSG-SrtA^{2M}, pMCSG-SrtA^{WT} was used as a template for inverse PCR amplification with a pair of phosphorylated primers carrying the intended mutation (Table S1) as previously described (1). The resulting linear PCR product was ligated before introduced into *E. coli* DH5α. The generated plasmid was verified by DNA sequencing prior to introduce into *E. coli* BL21 (DE3).

Recombinant SpaA plasmids – A plasmid expressing recombinant His-tagged SpaA protein of *C. diphtheriae* lacking the N-terminal signal peptide and C-terminal transmembrane domain (residues 30 to 500) was generated using the LIC method described above (see Table S2 for primers). The resulting plasmid (pMCSG-SpaA) was introduced into *E. coli* BL21 (DE3) after verification by DNA sequencing. Using the above site-directed mutagenesis method, pMCSG-SpaA was then used as a template to generate pMCSG-SpaA^{K190A} that expresses the same SpaA molecule with lysine 190 replaced by alanine. To generate plasmids pSUMO-^{N}SpaA and pMBP-^{C}SpaA, which express the N-terminal (residues 30 to 194) and C-terminal (residues 350 to 500) domains of SpaA, pE-SUMO and pE-MAPLE were used, respectively, in the same cloning protocol as described for pSUMO-SrtA^{WT} above.

Recombinant SpaB plasmids – Similar to the construction of the recombinant SpaA plasmids, pMCSG was used to clone a recombinant SpaB protein (residues 25 to 180), lacking the N-terminal signal peptide and the C-terminal hydrophobic domain and the charged tail. This generated plasmid was used as a template for site-directed mutagenesis to produce the SpaB-K139A mutant protein.

pSrtA and its derivatives – For *srtA* expression in *C. diphtheriae*, the *E. coli*/*Corynebacterium* shuttle vector pCGL0243 was used (2). pSrtA (3), a pCGL0243 derivative that expresses *C. diphtheriae srtA* under control of the *spaA* promoter, served as a template for site-directed mutagenesis, as described above, to generate various SrtA variants used in this study (Table S1 and see Table S2 for primers). The resulting plasmids were introduced into *E. coli* DH5α for DNA sequencing prior to electroporation into *C. diphtheriae* strains.

Protein Expression and Purification

For in vitro pilus polymerization, His-tagged proteins were purified according to a published procedure (4). Briefly, *E. coli* BL21 (DE3) cells harboring pSUMO-SrtA^{WT}, pSUMO-SrtA2M, pMCSG-SpaA, pMCSG-SpaA^{K190A}, pSUMO-^{N}SpaA, or pMBP-^{C}SpaA were grown in LB supplemented with ampicillin at 100 μg ml^{-1} at 37^{0}C until OD_{600} of ~ 0.6. Cells were equilibrated to 17°C and treated with 1 mM IPTG before they were allowed to grow overnight at 17°C to induce protein expression. Cells were then harvested by centrifugation (8,000 RPM for 20 min) and stored at -80°C for further processing. SrtA and SpaA-derived proteins were purified as a His6x-SUMO- fusion using HisPure Co^{2+} IMAC resin (Thermo) per the manufacturer's instructions. Briefly, cell pellets were resuspended in 50 mM Tris-HCl pH 8.0, 300 mM NaCl, and 5 mM $CaCl_2$ (lysis buffer) and lysed by sonication. Subsequent cell lysate was then fractionated by centrifugation (15,000 RPM for 1 hr) and the supernatant was loaded onto HisPure Co^{2+} IMAC resin. Proteins were then eluted from the resin using lysis buffer supplemented with 200 mM Imidazole. The His6x-SUMO tag was removed by the addition of His6x-Ulp1 protease, and subsequent HisPure Co^{2+} purification. Protein purity was determined by SDS-PAGE analysis.

For crystallization, recombinant proteins were purified according to a published procedure (5). Briefly, *E. coli* BL21 (DE3) cells harboring pMCSG-SrtA^{WT} or pMCSG-SrtA^{2M} was cultured in 2x YT broth containing ampicillin at 100 μg ml^{-1} at 37°C with shaking until OD_{600} of ~ 1.0. The culture was later induced with 0.4 mM IPTG and allowed to grow overnight at 18°C with shaking. Cells were harvested and disrupted by sonication. The lysate containing SrtA was purified by using Ni-NTA (Qiagen) affinity chromatography with the addition of 5 mM β-mercaptoethanol in all buffers. The N-terminal His-tag and TEV restriction sequence of the protein was removed by the TEV protease (0.15 mg for 20 mg purified protein) incubated for 16 h at 4°C, and then passed through a Ni-NTA column to remove both the TEV protease and cleaved N-terminal tags. The final step of purification was gel-filtration on HiLoad 16/60

Superdex 200pg column (GE Healthcare) in crystallization buffer 10 mM HEPES buffer pH 7.5, 200 mM NaCl and 1 mM DTT. The protein was concentrated on Amicon Ultracel 10K centrifugal filters (Millipore) up to 24 mg/ml concentration.

Protein crystallization, Data Collection, Structure Determination and Refinement

The initial crystallization condition was determined with a sparse crystallization matrix at 4°C and 16°C temperatures using the sitting-drop vapor-diffusion technique as reported (5). The best crystallization condition for the $SrtA^{WT}$ protein was found in 0.1 M MES:NaOH buffer pH 6.5, 1.6 M ammonium sulfate and 10% dioxane at 4°C. The SrtA2M protein was crystallized in 0.2 M sodium chloride and 20% PEG 3350 at 4°C. The $SrtA^{WT}$ and $SrtA^{2M}$ protein crystals selected for data collection were soaked in the crystallization buffer supplemented with either 28% sucrose or 25% glycerol, respectively, and flash-cooled in liquid nitrogen.

Single-wavelength X-ray diffraction data were collected at 100 K temperature at the 19-ID beamline using the program SBCcollect. The intensities were integrated and scaled with the HKL3000 suite (6). The SrtA structures were determined by molecular replacement using the HKL3000 suite incorporating following programs: MOLREP (7), SOLVE/RESOLVE (8), and ARP/wARP (9). The coordinates for the *A. oris* sortase SrtC-1 (10) (PDB:2XWG) were used as the starting model for the $SrtA^{WT}$ structure. Several rounds of manual adjustments of structure models using COOT (11) and refinements with Refmac program (12) from CCP4 suite (13) were performed. The stereochemistry of the structure was validated with PHENIX suite (6) incorporating MOLPROBITY(14) tools. Data collection and refinement statistics are summarized in Tables 1 and 2. Atomic coordinates and structure factors of $SrtA^{WT}$ and $SrtA^{2M}$ were deposited into the Protein Data Bank as 5K9A and 6BWE, respectively.

Nuclear magnetic resonance (NMR) analysis of $SrtA^{2M}$ and $^{\Delta}SrtA^{2M}$ sortases

^{1}H-^{15}N HSQC NMR spectra of $SrtA^{2M}$ and $^{\Delta}SrtA^{2M}$ were acquired using the following samples: 150 μM [^{15}N] $^{\Delta}SrtA^{2M}$ and 500 μM [^{15}N] $SrtA^{2M}$ proteins dissolved in 50 mM sodium phosphate, 100 mM sodium chloride, 0.01% sodium azide, 5% D_2O, pH 6.5. NMR spectra were acquired at 298 K using a Bruker Avance 600 MHz spectrometer equipped with a triple resonance cryogenic probe. A total of 16 and 32 scans were used to acquire the $SrtA^{2M}$ and $^{\Delta}SrtA^{2M}$ spectra, respectively.

SI Figure Legends

Figure S1: Visualization of *C. diphtheriae* pili assembled by the SrtA^{2M} enzyme. Corynebacterial cells of indicated strains were immobilized on nickel grids and subjected to negative staining with 0.75% uranyl formate (A) or immune-gold labeling with antibodies against SpaA prior to negative staining (B). Scale bars in (A) indicate 100 nm; (B), 200 nm.

Figure S2: SrtA containing the wild type lid is inactive in catalyzing crosslinking of the isolated domains. Fusion proteins between SUMO and the N-terminal SpaA domain (^{N}SpaA; residues 30 to 194) and between maltose-binding protein MBP and the C-terminal SpaA domain (^{C}SpaA; residues 350 to 500) were reacted with either the SrtA^{WT} or SrtA^{2M} enzyme at a 3:1 ratio, respectively. The reaction samples were analyzed by SDS-PAGE and Coomassie staining after 0, 24 h, 48 h, and 72 h of incubation.

Figure S3: Sequence alignment of Gram-positive sortases showing the class C signature sequence. Shown is an alignment of amino acids encompassing strands β7 to β8 of class C sortases. All aligned class C enzymes have been experimentally demonstrated to assemble pili by either cellular or biochemical methods. The conserved TP(Y/L)XIN(S/T)H signature sequence in class C enzymes is shaded in light blue, while the catalytic cysteine and arginine residues are colored red. The bottom of the figure shows representative class A, B, D and E enzymes sortases that do not assemble pili, but instead attach proteins to the cell wall. The GenBank accession codes are as follows: WP_010935503 (*Corynebacterium diphtheriae* SrtA); WP_010934130 (*Corynebacterium diphtheriae* SrtB); WP_010934133 (*Corynebacterium diphtheriae* SrtC); WP_010935679 (*Corynebacterium diphtheriae*, SrtD); WP_010935678 (*Corynebacterium diphtheriae*, SrtE); WP_002307920 (*Enterococcus faecium*, SrtC); WP_014569086 (*Lactobacillus rhamnosus*, SrtC1); WP_060958109 (*Actinomyces oris*, SrtC1); WP_060956887 (*Actinomyces oris*, SrtC2); WP_000047114 (*Streptococcus pneumoniae*, SrtC1); WP_050148456

(*Streptococcus agalactiae*, SrtC1); WP_000746885 (*Streptococcus agalactiae*, SrtC2); WP_000828081 (*Bacillus cereus*, SrtD); WP_037276992 (*Ruminococcus albus*, SrtC).

Figure S4: Comparison of the ^{1}H-^{15}N HSQC spectra of SrtA^{2M} and $^{\Delta}$SrtA^{2M}. The ^{1}H-^{15}N HSQC spectrum of SrtA^{2M} is shown (red). The expanded images are selected regions from the HSQC spectrum overlaid with the corresponding spectrum of $^{\Delta}$SrtA^{2M} (blue). The cross-peaks originate from backbone amides located within structured portions of each protein based on their down-field ^{1}H chemical shifts. Similar chemical shifts are observed indicating that the proteins adopt similar atomic structures.

SI Tables

Table S1: Bacterial strains and plasmids used

Strain & Plasmid	Description	Reference
Strain		
NCTC 13129	A *C. diphtheriae* type strain	(15)
HT2	Δ*srtA*; an isogenic derivative of NCTC 13129	(15)
HT52	Δ*spaA*/Δ*srtA*; an isogenic derivative of NCTC 13129	(3)
Plasmid		
pCGL0243	*C. diphtheriae*/*E. coli* shuttle vector, Kan^R	(16)
pSrtA	A pCGL0243 derivative expressing *C. diphtheriae* SrtA under control of the *spaA* promoter	(3)
$pSrtA^{2M}$	A pSrtA derivative, expressing SrtA with D81G and W83G mutations	This study
$pSrtA^{D81A}$	A pSrtA derivative, expressing SrtA with D81G mutation	This study
$pSrtA^{W83A}$	A pSrtA derivative, expressing SrtA with W83A mutation	This study
$pSrtA^{DW2A}$	A pSrtA derivative expressing SrtA with D81A and W83A mutations	This study
$pSrtA^{C222A}$	A pSrtA derivative expressing SrtA with C222A mutation	This study
$pSrtA^{H160A}$	A pSrtA derivative expressing SrtA with H160A mutation	This study
$pSrtA^{R231A}$	A pSrtA derivative expressing SrtA with R231A mutation	This study
pSpaA-SrtA	A pCGL0243 derivative expressing *C. diphtheriae* SpaA and wild type SrtA under control of the *spaA* promoter	This study
pSpaA-$SrtA^{2M}$	A pSpaA-SrtA derivative expressing *C. diphtheriae* SpaA and $SrtA^{2M}$ under control of the *spaA* promote	This study
$pSpaA^{K190A}$-$SrtA^{2M}$	A pSpaA-$SrtA^{2M}$ derivative with K190A mutation in SpaA	This study
pMCSG7	Ligation-independent cloning vector; Amp^R	(1)
pMCSG-SrtA	*srtA* gene encoding a.a. 37 to 257cloned into pMCSG7	This study
pMCSG-$SrtA^{2M}$	A pMCSG7 derivative expressing His-tagged SrtA with D81G and W83G mutations	This study
pMCSG-SpaA	A pMCSG7 derivative expressing His-tagged SpaA (residues 30 to 500)	This study
pMCSG-$SpaA^{K190A}$	A pMCSG-SpaA derivative expressing SpaA with K190A mutation	This study
pMCSG-SpaB	A pMCSG7 derivative expressing His-tagged SpaB (residues 25 to 180)	This study

pMCSG-SpaB^{K139A}	A pMCSG-SpaB derivative expressing SpaB with K139A mutation	This study
pE-SUMO	Expression vector	LifeSensors
pSUMO-SrtA^{WT}	*srtA* gene encoding a.a. 37 to 257 cloned into pE-SUMO	This study
pSUMO-SrtA^{2M}	Same as pSUMO-SrtA^{WT}, but carrying D81G and W83G mutations	This study
pSUMO-$^{\Delta}$SrtA^{2M}	The gene sequence encoding H1 helix (a.a. 37-54) was deleted from pSUMO-SrtA^{2M}	This study
pSUMO-SrtA^{Y225A}	Same as pSUMO-SrtA^{2M}, but carrying Y225A mutation	This study
pSUMO-SrtA^{N228A}	Same as pSUMO-SrtA^{2M}, but carrying N228A mutation	This study
pSUMO-SrtA^{S229A}	Same as pSUMO-SrtA^{2M}, but carrying S229A mutation	This study
pSUMO-^{N}SpaA	N-terminal sequence of SpaA (residues 30-194) cloned into pE-SUMO	This study
pE-MAPLE	Expression vector	
pMBP-^{C}SpaA	C-terminal sequence of SpaA (residues 350-500) cloned into pE-MAPLE	This study

Table S2: Primers used for cloning and site-directed mutagenesis (SDM)

Primer	Sequence[(a)]	Used for
SrtA-SDM-DW2G-5	cgGggctgaacagggcgtcgaaaaacag	pCGL-SrtA^{2M}; pMCSG-SrtA^{2M}; pSpaA-SrtA^{2M}
SrtA-SDM-DW2G-3	gaCcgaggatcgggaagcctactactaccg	pCGL-SrtA^{2M}; pMCSG-SrtA^{2M}; pSpaA-SrtA^{2M}
SrtA-SDM-D81A-5	gCtccgtggctgaacagggcgtc	pCGL-SrtA^{D81A}
SrtA-SDM-D81A-3	gaggatcgggaagcctactaccgac	pCGL-SrtA^{D81A}
SrtA-SDM-W83A-5	GCgctgaacagggcgtcgaaaaacag	pCGL-SrtA^{W83A}
SrtA-SDM-W83A-3	cggatcgaggatcgggaagcc	pCGL-SrtA^{W83A}
SrtA-SDM-DW2A-5	cgGggctgaacagggcgtcgaaaaacag	pCGL-SrtA^{DW2A}
SrtA-SDM-DW2A-3	gaCcgaggatcgggaagcctactaccg	pCGL-SrtA^{DW2A}
SrtA-SDM-C222A-5	GCcacccccctacgccgtcaac	pCGL-SrtA^{C222A}
SrtA-SDM-C222A-3	ggtgatgagtgttatttggtc	pCGL-SrtA^{C222A}
SrtA-SDM-H160A-5	GCcagcggccttgccaacgc	pCGL-SrtA^{H160A}
SrtA-SDM-H160A-3	tccggtgatcacggggtg	pCGL-SrtA^{H160A}
SrtA-SDM-R231A-5	ctcctcgtacgagcccaccgc	pCGL-SrtA^{R231A}
SrtA-SDM-R231A-3	TGCgtgggagttgacggcgtagg	pCGL-SrtA^{R231A}
SrtA-LIC-N37-5	tacttccaatccaatgcaaacaacgcgcgccaagcacg	pMCSG-SrtA
SrtA-LIC-Q257-3	ttatccacttccaatgttactattgccagattttggtgccgg	pMCSG-SrtA
SpaB-LIC-Q25-5	tacttccaatccaatgcacaagaagcaaacacattggtcattgacctc	pMCSG-SpaB
SpaB-LIC-G180-3	ttatccacttccaatgttactatccgaggatgcttgcgcc	pMCSG-SpaB
SpaB-SDM-K139A-5	cttatcgacgccacccccgg	pMCSG-SpaB^{K139A}
SpaB-SDM-K139A-3	cGCggggcgcagggtcatgg	pMCSG-SpaB^{K139A}
SpaA-SDM-K190A-5	GCgcaccaggctttgtctgagcc	pMCSG-SpaA^{K190A}; pSpaA^{K190A}-SrtA^{2M}
SpaA-SDM-K190A-3	gggatacacgtgcacgtcttggag	pMCSG-SpaA^{K190A}; pSpaA^{K190A}-SrtA^{2M}
SpaA-LIC-E30-5	tacttccaatccaatgcagaagagtcacagtatcatgcagtccaac	pMCSG-SpaA
SpaA-LIC-I500-3'	ttatccacttccaatgttagatgcgccccgaaccac	pMCSG-SpaA
SpaA-SDM-K190A-5	gcgcaccaggctttgtctgagcc	pMCSG-SpaA^{K190A}
SpaA-SDM-K190A-3	gggatacacgtgcacgtcttggag	pMCSG-SpaA^{K190A}
SrtA-SDM-Y225A-5	catcacctgcacccccGCGgccgtcaactcccaccg	pSUMO-SrtA^{Y225A}
SrtA-SDM-Y225A-3	cggtgggagttgacggcCGCgggggtgcaggtgatg	pSUMO-SrtA^{Y225A}
SrtA-SDM-N228A-5	cccctacgccgtcGCctcccaccgactc	pSUMO-SrtA^{N228A}
SrtA-SDM-N228A-3	gagtcggtgggagGCgacggcgtagggg	pSUMO-SrtA^{N228A}
SrtA-SDM-S229A-5	cccctacgccgtcaacGcGcaccgactcctcgtacg	pSUMO-SrtA^{S229A}

SrtA-SDM-S229A-3	cgtacgaggagtcggtgCgCgttgacggcgtagggg	pSUMO-SrtA^{S229A}

[(a)] Upper case letters are the mutated nucleotides for site-directed mutagenesis.

References

1. Reardon-Robinson ME*, et al.* (2015) A thiol-disulfide oxidoreductase of the Gram-positive pathogen *Corynebacterium diphtheriae* is essential for viability, pilus assembly, toxin production and virulence. *Mol Microbiol* 98(6):1037-1050.
2. Chang C, Mandlik A, Das A, & Ton-That H (2011) Cell surface display of minor pilin adhesins in the form of a simple heterodimeric assembly in *Corynebacterium diphtheriae*. *Mol Microbiol* 79(5):1236-1247.
3. Guttilla IK*, et al.* (2009) Acyl enzyme intermediates in sortase-catalyzed pilus morphogenesis in gram-positive bacteria. *J Bacteriol* 191(18):5603-5612.
4. Bornhorst JA & Falke JJ (2000) Purification of proteins using polyhistidine affinity tags. *Methods Enzymol* 326:245-254.
5. Reardon-Robinson ME*, et al.* (2015) A Disulfide Bond-forming Machine Is Linked to the Sortase-mediated Pilus Assembly Pathway in the Gram-positive Bacterium *Actinomyces oris*. *J Biol Chem* 290(35):21393-21405.
6. Adams PD*, et al.* (2002) PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr* 58(Pt 11):1948-1954.
7. Vagin A & Teplyakov A (1997) MOLREP: an automated program for molecular replacement. *Journal of Applied Crystallography* 30:1022-1025.
8. Terwilliger TC (2003) SOLVE and RESOLVE: automated structure solution and density modification. *Methods Enzymol* 374:22-37.
9. Morris RJ, Perrakis A, & Lamzin VS (2003) ARP/wARP and automatic interpretation of protein electron density maps. *Methods Enzymol* 374:229-244.
10. Persson K (Structure of the sortase AcSrtC-1 from *Actinomyces oris*. *Acta Crystallogr D Biol Crystallogr* 67(Pt 3):212-217.

11. Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60(Pt 12 Pt 1):2126-2132.

12. Murshudov GN, Vagin AA, & Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53(Pt 3):240-255.

13. Collaborative Computational Project N (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* 50(Pt 5):760-763.

14. Davis IW, Murray LW, Richardson JS, & Richardson DC (2004) MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Res* 32(Web Server issue):W615-619.

15. Ton-That H & Schneewind O (2003) Assembly of pili on the surface of *Corynebacterium diphtheriae*. *Mol Microbiol* 50(4):1429-1438.

16. Ankri S, Reyes O, & Leblon G (1996) Electrotransformation of highly DNA-restrictive corynebacteria with synthetic DNA. *Plasmid* 35(1):62-66.

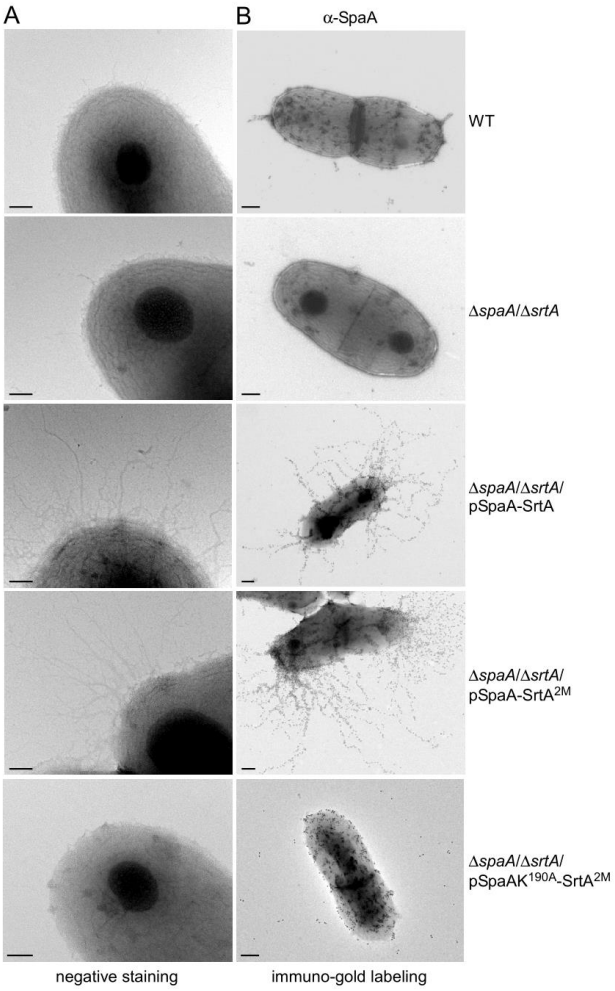


Figure S1: Chang et al.

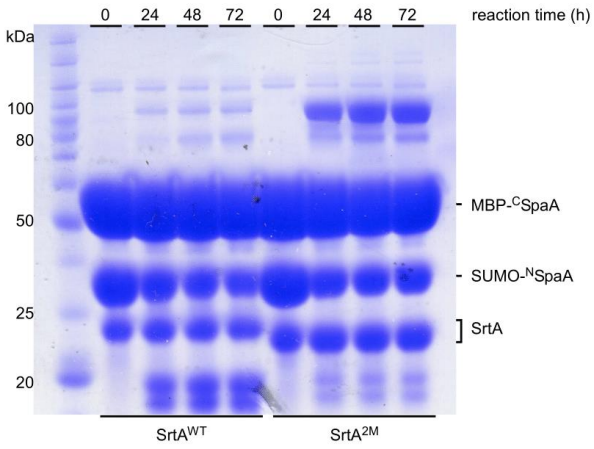


Figure S2: Chang et al.

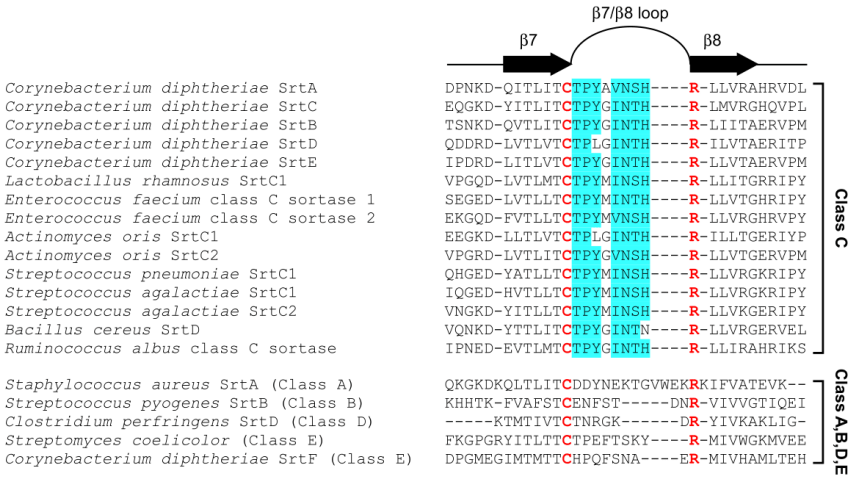


Figure S3: Chang et al.

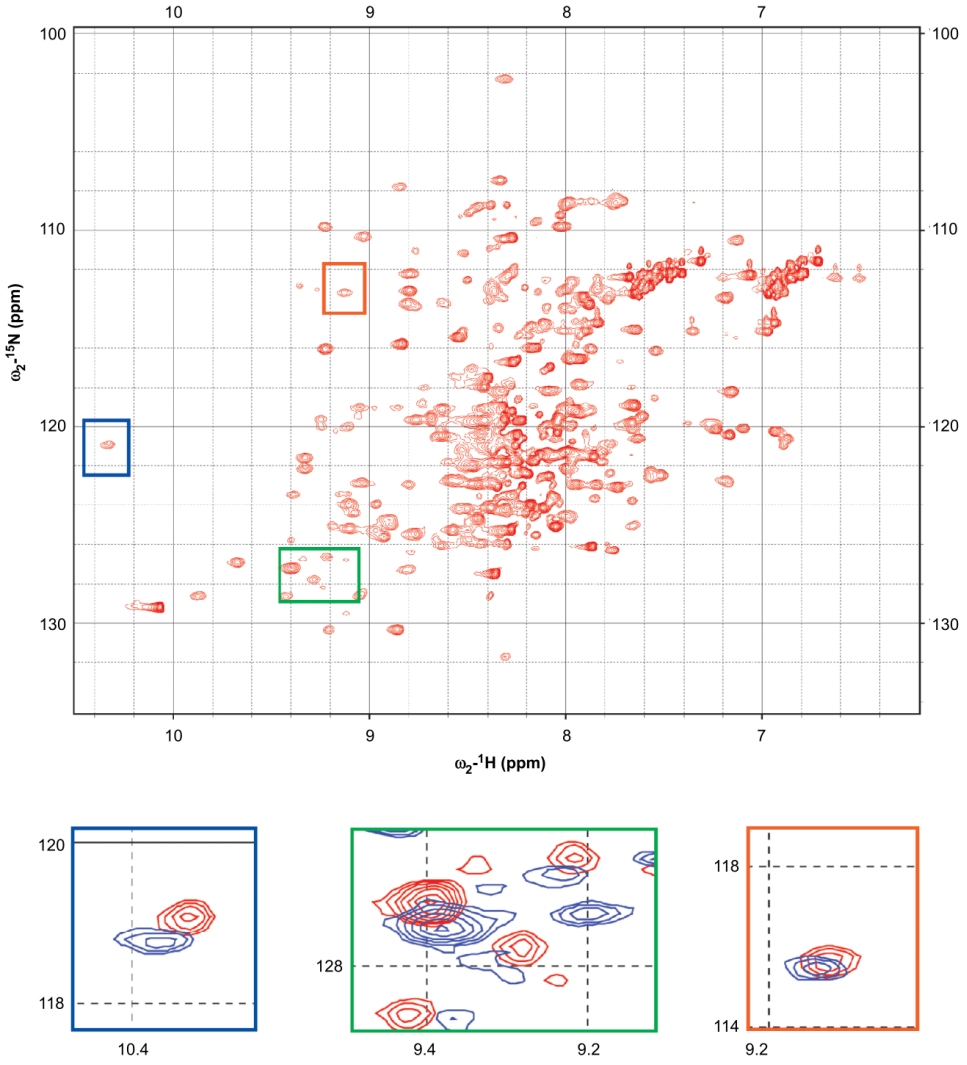


Figure S4: Chang et al.

APPENDIX II: Protein Labeling via a Specific Lysine-Isopeptide Bond Using the Pilin Polymerizing Sortase from *Corynebacterium diphtheriae*

JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Cite This: J. Am. Chem. Soc. 2018, 140, 8420−8423
Communication
pubs.acs.org/JACS

Protein Labeling via a Specific Lysine-Isopeptide Bond Using the Pilin Polymerizing Sortase from *Corynebacterium diphtheriae*

Scott A. McConnell,[†] Brendan R. Amer,[†] John Muroski,[†] Janine Fu,[†] Chungyu Chang,[§] Rachel R. Ogorzalek Loo,[†] Joseph A. Loo,[†] Jerzy Osipiuk,[‡] Hung Ton-That,[§] and Robert T. Clubb*[,†]

[†]Department of Chemistry and Biochemistry, UCLA-DOE Institute for Genomics and Proteomics and the Molecular Biology Institute, University of California, Los Angeles, Los Angeles, California 90095, United States

[‡]Structural Biology Center, Argonne National Laboratory, Argonne, Illinois 60439, United States

[§]Department of Microbiology & Molecular Genetics, University of Texas Health Science Center, Houston, Texas 77030, United States

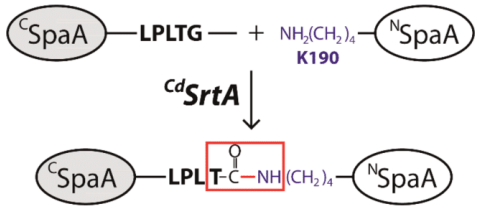
Supporting Information

ABSTRACT: Proteins that are site-specifically modified with peptides and chemicals can be used as novel therapeutics, imaging tools, diagnostic reagents and materials. However, there are few enzyme-catalyzed methods currently available to selectively conjugate peptides to internal sites within proteins. Here we show that a pilus-specific sortase enzyme from *Corynebacterium diphtheriae* (^{Cd}SrtA) can be used to attach a peptide to a protein via a specific lysine-isopeptide bond. Using rational mutagenesis we created ^{Cd}SrtA^{3M}, a highly activated cysteine transpeptidase that catalyzes *in vitro* isopeptide bond formation. ^{Cd}SrtA^{3M} mediates bioconjugation to a specific lysine residue within a fused domain derived from the corynebacterial SpaA protein. Peptide modification yields greater than >95% can be achieved. We demonstrate that ^{Cd}SrtA^{3M} can be used in concert with the *Staphylococcus aureus* SrtA enzyme, enabling dual, orthogonal protein labeling via lysine-isopeptide and backbone-peptide bonds.

Enzymatic methods that site-specifically functionalize proteins are of significant interest, as they can enable the creation of novel protein-conjugates for medical and research applications.[1−5] The *Staphylococcus aureus* sortase (^{Sa}SrtA) has been developed into a powerful protein engineering tool.[6−10] It catalyzes a transpeptidation reaction that covalently modifies the target protein via a backbone peptide bond, by joining peptide segments that contain a LPXTG "sort-tag" and an N-terminal oligoglycine amine group.[11,12] Several groups have now optimized this reaction to modify proteins with a range of molecules, including drugs, lipids, sugars, fluorophores, and peptides.[13−21] While ^{Sa}SrtA is potent tool, it is almost exclusively used to modify target proteins at their N- or C-termini, while it labels internal lysine side chains as a side reaction with low sequence specificity.[15,22,23] Here we show that a mutationally activated sortase enzyme from *Corynebacterium diphtheriae* (^{Cd}SrtA) can site-specifically install a peptide on a protein via a lysine-isopeptide bond. ^{Cd}SrtA and ^{Sa}SrtA have orthogonal activities, enabling dual peptide-fluorophore labeling of a protein via lysine isopeptide- and backbone peptide-bonds, respectively.

Gram-positive bacteria use specialized sortase enzymes to construct pili: long, thin fibers (0.2−3.0 μm × 2−10 nm) that project from the cell surface to mediate bacterial adherence to host tissues, biofilm formation and host immunity modulation.[24−26] These structures are distinct from pili produced by Gram-negative bacteria because their protein subunits (called pilins) are cross-linked by lysine-isopeptide bonds that confer enormous tensile strength.[27,28] Recently, we reconstituted *in vitro* the assembly reaction that builds the archetypal SpaA-pilus in *C. diphtheriae*, the causative agent of pharyngeal diphtheria.[29] ^{Cd}SrtA functions as a pilin polymerase, performing a repetitive transpeptidation reaction that covalently links adjacent SpaA pilin subunits together via lysine-isopeptide bonds. As shown in Scheme 1, ^{Cd}SrtA cross-links adjacent

Scheme 1. ^{Cd}SrtA-Catalyzed Isopeptide Bond Formation



SpaA proteins by connecting their N- (^{N}SpaA, residues 30−194) and C-terminal (^{C}SpaA, residues 350−500) domains, which contain a reactive WxxxVxVYPK pilin motif and LPLTG sorting signal sequences, respectively. In the reaction, ^{Cd}SrtA first cleaves the LPLTG sequence in ^{C}SpaA between the threonine and glycine, forming an acyl-enzyme intermediate in which the catalytic C222 residue in ^{Cd}SrtA is joined to ^{C}SpaA's threonine carbonyl atom. This transient intermediate is then nucleophilically attacked by the reactive K190 within ^{N}SpaA's pilin motif resulting in a T494-K190 isopeptide bond between

Received: May 17, 2018
Published: June 21, 2018

DOI: 10.1021/jacs.8b05200
J. Am. Chem. Soc. 2018, 140, 8420−8423

^{C}SpaA and ^{N}SpaA domains within adjacent pilin subunits. Previously, we demonstrated that wild-type ^{Cd}SrtA is catalytically inactive *in vitro* due to the presence of an N-terminal polypeptide segment, called a lid, that masks the enzyme's active site (Figure 1A).[30−34] Moreover, we showed that it was

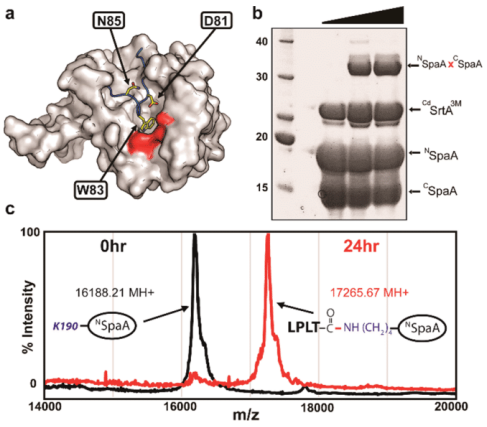


Figure 1. Mutationally activated ^{Cd}SrtA catalyzes lysine-isopeptide bond formation. (A) The structure of ^{Cd}SrtA^{WT} harbors an inhibitory "lid" structure (blue). Side chains that were mutated to activate the enzyme are shown as yellow sticks. The surface of the catalytic site is colored red. (B) Protein−protein ligation using the activated ^{Cd}SrtA^{3M} enzyme. SDS-PAGE analysis of the reaction demonstrating formation of the lysine-isopeptide linked ^{N}SpaAx^{C}SpaA product. The reaction (100 μM enzyme, 300 μM ^{C}SpaA and ^{N}SpaA) was sampled at 0, 24, and 48 h. (C) High yield protein-peptide labeling with ^{Cd}SrtA^{3M}. MALDI-MS data showing that >95% ^{N}SpaA is labeled with peptide containing the sort-tag, LPLTG$_{peptide}$. MALDI-MS spectra recorded at 0 (black) and 24 h (red) are overlaid.

possible to activate the enzyme by introducing D81G and W83G lid mutations and we demonstrated that a soluble catalytic domain harboring these mutations (^{Cd}SrtA^{2M}, residues 37−257 of ^{Cd}SrtA with D81G/W83G mutations) site-specifically ligates the isolated ^{N}SpaA and ^{C}SpaA domains *in vitro*.[29]

Toward the goal of creating a lysine modifying bioconjugation reagent we improved the ligation activity of ^{Cd}SrtA^{2M}, we defined substrate determinants that are required for catalysis. In addition to the aforementioned D81 and W83 mutations in ^{Cd}SrtA^{2M}, inspection of the crystal structure reveals three lid residues that may stabilize its positioning over the active site (I79, N85, K89). The ligation activities of triple mutants of ^{Cd}SrtA containing the D81G and W83G alterations, as well I79R, N85A or K89A substitutions were determined. A D81G/W83G/N85A triple mutant, hereafter called ^{Cd}SrtA^{3M}, has the highest level of ligation activity (Figures 1B and S1). After a 24 h incubation with the isolated ^{N}SpaA and ^{C}SpaA domains, ^{Cd}SrtA^{3M} produces 10.6-fold more cross-linked ^{N}SpaAx^{C}SpaA product than ^{Cd}SrtA WT and 35% more product than ^{Cd}SrtA^{2M} (Figure S1). The mutations in ^{Cd}SrtA^{3M} presumably further displace its lid, thereby facilitating enhanced binding of ^{C}SpaA's LPLTG sorting signal and subsequent acylation by C222. This is substantiated by our finding that the ^{Cd}SrtA^{3M} triple mutant exhibits the highest level of activity in a HPLC-based sorting signal cleavage assay that reports on formation of the acyl-enzyme intermediate (Figure S1) and previous studies that have shown that alterations in the lid increase C222 reactivity with 4,4′-dithiodipyridine.[29]

^{N}SpaA and ^{C}SpaA are joined by ^{Cd}SrtA^{3M} via their respective pilin motif and LPXTG sorting signal elements. To elucidate determinants required for recognition of the K190 nucleophile, ^{Cd}SrtA^{3M} was incubated with a peptide containing the pilin motif (DGWLQDVHVYPKHQALS) and either ^{C}SpaA or a peptide containing its C-terminal sorting signal (KNAGFELPLTGGSGRI) (Figure S2). In both instances, no detectable product was observed, indicating that ^{Cd}SrtA^{3M} requires additional tertiary elements within ^{N}SpaA to recognize K190. In contrast, when ^{Cd}SrtA^{3M} is incubated with ^{N}SpaA and the peptide containing the C-terminal sorting signal, > 95% of ^{N}SpaA is labeled with the peptide (Figure 1C). Moreover, LC-MS/MS analysis of the cross-linked species reveals that the components are joined via a site-specific isopeptide bond between the threonine within the sorting signal peptide and the Nε amine of K190 in ^{N}SpaA (Figure S3A).

We next demonstrated that ^{Cd}SrtA^{3M} can be used to label a target protein via an isopeptide bond with either a peptide fluorophore or another protein. In the labeling reaction a target protein is first expressed as a fusion with the ^{N}SpaA domain containing the pilin motif (hereafter called PM), and then reacted with a LPLTG-containing biomolecule and ^{Cd}SrtA^{3M} (Figure 2A). To demonstrate peptide fluorophore attachment using ^{Cd}SrtA^{3M}, we incubated the enzyme with ^{N}SpaA and a

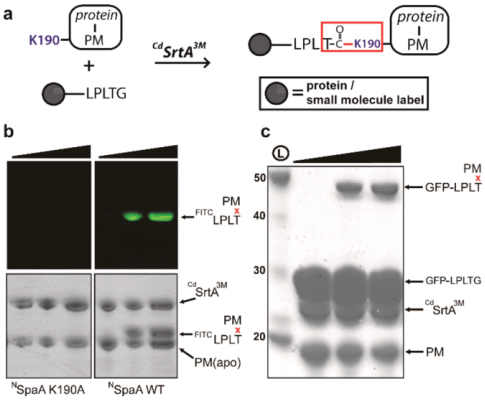


Figure 2. Labeling proteins via a lysine-isopeptide bond with ^{Cd}SrtA^{3M}. (A) Schematic showing ^{Cd}SrtA^{3M} catalyzed labeling of pilin motif (PM) fusion protein with a protein containing the LPLTG sorting signal or a LPLTG peptide with a functional label. (B) SDS-PAGE analysis of a fluoropeptide modification reaction containing ^{Cd}SrtA^{3M} (100 μM) and ^{FITC}LPLTG (1 mM) and either ^{N}SpaA (K190A) (lanes 1−3) or ^{N}SpaA WT (lanes 4−6) (both 100 μM). Top and bottom panels are the same gels visualized by fluorescence or by Coomassie staining, respectively. Reaction progress was measured at 0 (lanes 1,4), 24 (lanes 2,5) and 48 h (lanes 3,6). (C) Protein−protein ligation with ^{Cd}SrtA^{3M}. As in panel (B), except reactions contained GFP-LPLTG (300 μM) instead of the fluoropeptide. Reactants were visualized with Coomassie staining at 0, 24, and 48 h (lanes 1−3, respectively).

DOI: 10.1021/jacs.8b05200
J. Am. Chem. Soc. 2018, 140, 8420−8423

fluorescent FITC-KNAGFELPLTGGSGRI peptide (^{FITC}LPLTG). After incubating for either 24 or 48 h the reaction components were separated by SDS-PAGE and visualized by either Coomassie staining or FITC fluorescence at 530 nm. ^{Cd}SrtA^{3M} labels ^{N}SpaA with the fluorescent peptide, yielding a ^{FITC}LPLTx^{N}SpaA cross-linked product (Figure 2B, right). Fluorophore labeling is specific, as ^{N}SpaA harboring a K190A mutation is unreactive in control experiments (Figure 2B, left).

To demonstrate that ^{Cd}SrtA^{3M} can also be used to join proteins together via an isopeptide bond, the isolated PM was reacted with green fluorescent protein engineered to contain a C-terminal LPLTGGSGRI sorting signal sequence (GFP-LPLTG). Incubation of these proteins with ^{Cd}SrtA^{3M} resulted in the appearance a higher molecular weight GFP-LPLTx^{N}SpaA cross-linked product (Figure 2C, S3B). Notably, the ^{Cd}SrtA^{3M} protein–protein ligation reaction is versatile, as labeling can be achieved with the PM fused to either the N- or C-terminus of the target protein.

The ^{Cd}SrtA and ^{Sa}SrtA enzymes recognize distinct nucleophiles, suggesting that they can be used orthogonally to selectively label a single target protein at different sites. To demonstrate orthogonal labeling we created a fusion protein that contained the Small Ubiquitin-like Modifier (SUMO) protein harboring a pentaglycine peptide and PM at its N- and C-termini, respectively (G_5-SUMO$_{PM}$). Our dual modification approach involves sequential reaction of the G_5-SUMO$_{PM}$ substrate with each sortase and peptide fluorophores containing the cognate sorting signal, as outlined in Figure 3A. To selectively modify G_5-SUMO$_{PM}$ (species 1), it was first incubated with ^{Cd}SrtA^{3M} and FITC-LPLTG$_{pep}$ to create at high yield G_5-SUMO$_{PM}$-FITC (species 2) (Figure 3B). After removal of excess FITC-LPLTG peptide using a desalting column, the target protein was then labeled at its N-terminus with AlexaFluor$_{546}$-LPATG using ^{Sa}SrtA. This was achieved by incubating species 2 with ^{Sa}SrtA and AlexaFluor$_{546}$-LPATG to produce the doubly labeled protein (species 3). Separation of the reaction products by SDS-PAGE confirms dual labeling, as the expected fluorescence for each probe is detected at ~33 kDa during the procedure (Figure 3B). In particular, FITC-labeled Gly$_5$-SUMO$_{PM}$ is produced after treatment with ^{Cd}SrtA^{3M} (488/530 nm excitation/emission), and persists after treatment with ^{Sa}SrtA that catalyzes the second conjugation step with AlexaFluor$_{546}$ (532/605 nm excitation/emission). We note that a similar labeling strategy can presumably be used for fusion proteins that contain the SpaB basal pilin instead of ^{N}SpaA, as we have recently shown that ^{Cd}SrtA^{3M} can also use SpaB as a nucleophile *in vitro*.[29] A strength of our approach is the distinct nucleophile and sorting signal substrate specificities of each sortase, which limits cross reactivity. In addition to recognizing distinct nucleophiles, our findings indicate that the sortases have unique sorting signal substrate specificities; ^{Cd}SrtA^{3M} is unable to hydrolyze or use as a transpeptidation substrate sorting signals containing the sequence LPATG that is readily used by ^{Sa}SrtA, but instead it is selective for peptides containing LPLTG (Figures S4, S5). Moreover, the isopeptide bond created by ^{Cd}SrtA^{3M} is not significantly hydrolyzed by ^{Sa}SrtA or ^{Cd}SrtA after 24 h (Figure S6). Thus, ^{Cd}SrtA acts preferentially on its LPLTG sorting signal substrate, preventing potential reversal of LPATG peptides installed by ^{Sa}SrtA. Similarly, the isopeptide linkages installed by ^{Cd}SrtA are not a substrate for reversal by ^{Sa}SrtA or ^{Cd}SrtA.

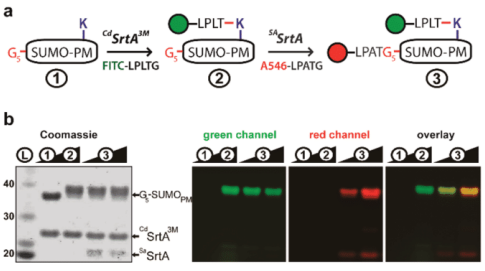


Figure 3. Orthogonal protein labeling using ^{Cd}SrtA^{3M} and ^{Sa}SrtA. (A) Sequential reaction scheme used to install fluorogenic peptides on a target protein via peptide- and isopeptide bonds. G_5-SUMO$_{PM}$ is a SUMO target protein that is fused to N- and C-terminal nucleophiles, pentaglycine (G_5) and the pilin motif (PM), respectively. (B) SDS-PAGE analysis of reaction mixture taken at different steps in the procedure: (1) prior to labeling, (2) after labeling with ^{FITC}LPLTG using ^{Cd}SrtA^{3M}, and (3) after labeling with A546-LPATG using ^{Sa}SrtA (0.25/2 h incubations). Panels show as indicated fluorescence gel imaging to detect FITC and A546 fluorophores using 488/530 (green channel) and 532/605 nm (red channel) wavelengths for excitation/emission, respectively, and the merged image of the gels demonstrating dual labeling. In the first panel, the same gel was visualized by coomassie staining.

The bioconjugation chemistry catalyzed by ^{Cd}SrtA^{3M} enables site-specific lysine labeling of a protein, creating an isopeptide linkage that may be less susceptible to proteolysis than conventional peptide bonds. An attractive feature of ^{Cd}SrtA^{3M} is its high degree of specificity for the ε-amine nucleophile within the pilin motif, which enables selective labeling. Transglutaminases can also modify protein lysine residues, but unlike ^{Cd}SrtA^{3M}, these enzymes exhibit minimal substrate specifity.[35,36] Similarly, ^{Sa}SrtA can modify lysines as a side reaction that occurs with minimal specificity and at low efficiency because the lysine ε-amine is not ^{Sa}SrtA's natural substrate.[15,22,23] Chemical methods that modify amino acid side chains have also been developed, but they often require cysteine or non-natural amino acid incorporation into the protein and in some instances harsh reaction conditions.[37] The bioconjugation chemistry catalyzed by ^{Cd}SrtA^{3M} is functionally similar to the nonenzymatic SpyTag/SpyCatcher system,[38,39] but its enzymatic activity affords greater control making ^{Cd}SrtA^{3M} an attractive new tool to engineer proteins.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.8b05200.

Additional data and procedures (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*rclubb@mbi.ucla.edu

ORCID

Robert T. Clubb: 0000-0001-5718-3985

Notes

The authors declare no competing financial interest.

DOI: 10.1021/jacs.8b05200
J. Am. Chem. Soc. 2018, 140, 8420–8423

■ ACKNOWLEDGMENTS

This work was supported by the U.S. Department of Energy Office of Science, Office of Biological and Environmental Research program under Award Number DE-FC02-02ER63421 and National Institutes of Health Grants AI52217 (R.T.C. and H. T-T.), DE025015 (H. T-T), GM103479 (J.A.L.) and U.S. Department of Energy, Office of Biological and Environmental Research contract DE-AC02-06CH11357 (J.O.). S.A.M. was supported by a Cellular and Molecular Biology Training Grant (Ruth L. Kirschstein National Research Service Award GM007185). NMR equipment used in this research was purchased using funds from shared equipment grant NIH S10OD016336.

■ REFERENCES

(1) Proft, T. *Biotechnol. Lett.* **2010**, *32*, 1−10.
(2) Matsumoto, T.; Tanaka, T.; Kondo, A. *Biotechnol. J.* **2012**, *7*, 1137−1146.
(3) Walper, S. A.; Turner, K. B.; Medintz, I. L. *Curr. Opin. Biotechnol.* **2015**, *34*, 232−241.
(4) Krall, N.; Da Cruz, F. P.; Boutureira, O.; Bernardes, G. J. L. *Nat. Chem.* **2016**, *8*, 103−113.
(5) Rashidian, M.; Dozier, J. K.; Distefano, M. D. *Bioconjugate Chem.* **2013**, *24*, 1277−1294.
(6) Mazmanian, S. K.; Liu, G.; Ton-That, H.; Schneewind, O. *Science (Washington, DC, U. S.)* **1999**, *285*, 760−763.
(7) Antos, J. M.; Chew, G. L.; Guimaraes, C. P.; Yoder, N. C.; Grotenbreg, G. M.; Popp, M. W. L.; Ploegh, H. L. *J. Am. Chem. Soc.* **2009**, *131*, 10800−10801.
(8) Williamson, D. J.; Fascione, M. A.; Webb, M. E.; Turnbull, W. B. *Angew. Chem., Int. Ed.* **2012**, *51*, 9377−9380.
(9) Levary, D. A.; Parthasarathy, R.; Boder, E. T.; Ackerman, M. E. *PLoS One* **2011**, *6*, No. e18342.
(10) Tsukiji, S.; Nagamune, T. *ChemBioChem* **2009**, *10*, 787−798.
(11) Popp, M. W.; Antos, J. M.; Grotenbreg, G. M.; Spooner, E.; Ploegh, H. L. *Nat. Chem. Biol.* **2007**, *3*, 707−708.
(12) Mao, H.; Hart, S. A.; Schink, A.; Pollok, B. A. *J. Am. Chem. Soc.* **2004**, *126*, 2670−2671.
(13) Samantaray, S.; Marathe, U.; Dasgupta, S.; Nandicoori, V. K.; Roy, R. P. *J. Am. Chem. Soc.* **2008**, *130*, 2132−2133.
(14) Antos, J. M.; Miller, G. M.; Grotenbreg, G. M.; Ploegh, H. L. *J. Am. Chem. Soc.* **2008**, *130*, 16338−16343.
(15) Möhlmann, S.; Mahlert, C.; Greven, S.; Scholz, P.; Harrenga, A. *ChemBioChem* **2011**, *12*, 1774−1780.
(16) Wagner, K.; Kwakkenbos, M. J.; Claassen, Y. B.; Maijoor, K.; Böhne, M.; van der Sluijs, K. F.; Witte, M. D.; van Zoelen, D. J.; Cornelissen, L. A.; Beaumont, T.; Bakker, A. Q.; Ploegh, H. L.; Spits, H. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 16820−16825.
(17) Beerli, R. R.; Hell, T.; Merkel, A. S.; Grawunder, U. *PLoS One* **2015**, *10*, No. e0131177.
(18) Amer, B. R.; MacDonald, R.; Jacobitz, A. W.; Liauw, B.; Clubb, R. T. *J. Biomol. NMR* **2016**, *64*, 197−205.
(19) Dorr, B. M.; Ham, H. O.; An, C.; Chaikof, E. L.; Liu, D. R. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 13343−13348.
(20) Chen, I.; Dorr, B. M.; Liu, D. R. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 11399−11404.
(21) Antos, J. M.; Truttmann, M. C.; Ploegh, H. L. *Curr. Opin. Struct. Biol.* **2016**, *38*, 111−118.
(22) Dasgupta, S.; Samantaray, S.; Sahal, D.; Roy, R. P. *J. Biol. Chem.* **2011**, *286*, 23996−24006.
(23) Bellucci, J. J.; Bhattacharyya, J.; Chilkoti, A. *Angew. Chem., Int. Ed.* **2014**, *54*, 441−445.
(24) Danne, C.; Dramsi, S. *Res. Microbiol.* **2012**, *163*, 645−658.
(25) Spirig, T.; Weiner, E. M.; Clubb, R. T. *Mol. Microbiol.* **2011**, *82*, 1044−1059.
(26) Ton-That, H.; Schneewind, O. *Trends Microbiol.* **2004**, *12*, 228−234.
(27) Echelman, D. J.; Alegre-Cebollada, J.; Badilla, C. L.; Chang, C.; Ton-That, H.; Fernández, J. M. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 2490−2495.
(28) Yeates, T. O.; Clubb, R. T. *Science* **2007**, *318*, 1558−1559.
(29) Chang, C.; Amer, B. R.; Osipiuk, J.; McConnell, S. A.; Huang, I.-H.; Hsieh, V.; Fu, J.; Nguyen, H. H.; Muroski, J.; Flores, E.; Ogorzalek Loo, R. R.; Loo, J. A.; Putkey, J. A.; Joachimiak, A.; Das, A.; Clubb, R. T.; Ton-That, H. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, E5477−E5486.
(30) Jacobitz, A. W.; Naziga, E. B.; Yi, S. W.; McConnell, S. A.; Peterson, R.; Jung, M. E.; Clubb, R. T.; Wereszczynski, J. *J. Phys. Chem. B* **2016**, *120*, 8302−8312.
(31) Manzano, C.; Izoré, T.; Job, V.; Di Guilmi, A. M.; Dessen, A. *Biochemistry* **2009**, *48*, 10549−10557.
(32) Cozzi, R.; Zerbini, F.; Assfalg, M.; D'Onofrio, M.; Biagini, M.; Martinelli, M.; Nuccitelli, A.; Norais, N.; Telford, J. L.; Maione, D.; Rinaudo, C. D. *FASEB J.* **2013**, *27*, 3144−3154.
(33) Persson, K. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2011**, *67*, 212−217.
(34) Manzano, C.; Contreras-Martel, C.; El Mortaji, L.; Izoré, T.; Fenel, D.; Vernet, T.; Schoehn, G.; Di Guilmi, A. M.; Dessen, A. *Structure* **2008**, *16*, 1838−1848.
(35) Yokoyama, K.; Nio, N.; Kikuchi, Y. *Appl. Microbiol. Biotechnol.* **2004**, *64*, 447−454.
(36) Fontana, A.; Spolaore, B.; Mero, A.; Veronese, F. M. *Adv. Drug Delivery Rev.* **2008**, *60*, 13−28.
(37) Spicer, C. D.; Davis, B. G. *Nat. Commun.* **2014**, *5*, DOI: 10.1038/ncomms5740.
(38) Reddington, S. C.; Howarth, M. *Curr. Opin. Chem. Biol.* **2015**, *29*, 94−99.
(39) Zakeri, B.; Fierer, J. O.; Celik, E.; Chittock, E. C.; Schwarz-Linek, U.; Moy, V. T.; Howarth, M. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, E690−E697.

DOI: 10.1021/jacs.8b05200
J. Am. Chem. Soc. 2018, 140, 8420−8423

Supporting Information

Protein Labeling via a Specific Lysine-Isopeptide Bond using the Pilin Polymerizing Sortase from *Corynebacterium diphtheriae*

Scott A. McConnell, Brendan R. Amer, John Muroski, Janine Fu, Chungyu Chang, Rachel R. Ogorzalek Loo, Joseph A. Loo, Jerzy Osipiuk, Hung Ton-That and Robert T. Clubb

Table of Contents

Protein and peptide reagents. All protein constructs used in this study were expressed with pE-SUMO (LifeSensors) expression vector. Constructs harboring point mutations were generated using standard site directed mutagenesis methods, transformed to XL10 cells for amplification, and sequence confirmed. His_6-tagged proteins were purified according to a published procedure[1]. Briefly, *E. coli* BL21 (DE3) cells harboring either pG_5-$SUMO_{PM}$, pSUMO-$^{Cd}SrtA^{WT}$, pSUMO-$^{Cd}SrtA^{2M}$, pSUMO-$^{Cd}SrtA^{3M}$, pSUMO-^{Sa}SrtA 4M, pSUMO-^{Sa}SrtA WT, pSUMO-GFP-LPLTGGSGRI, pSUMO-SpaA(30-500) pSUMO-^{N}SpaA(30-194), or pSUMO -^{C}SpaA(350-500) were grown in LB supplemented with kanamycin at 500 µg/ml at 37^{0}C until OD_{600} of ~ 0.6. Cells were induced with 1 mM IPTG and protein expression was allowed to proceed overnight at 17°C. Cells were then harvested by centrifugation (7,000 x g for 10 min). All proteins were purified as a His_6-SUMO- fusion using HisPure Co^{2+} IMAC resin (Thermo Scientific) per the manufacturer's instructions. Briefly, cell pellets were resuspended in 50 mM Tris-HCl pH 8.0, 300 mM NaCl (lysis buffer) and lysed by high pressure homogenization. The cell lysate was then fractionated by centrifugation (15,000 x g for 40 min) and the supernatant was loaded onto HisPure Co^{2+} IMAC resin. Proteins were then eluted from the resin using lysis buffer supplemented with 200 mM Imidazole. The His6x-SUMO tag was removed by adding His_6-Ulp1 protease, and subsequent HisPure Co^{2+} purification. Protein purity was determined by SDS-PAGE analysis.

Peptides used in this study were synthesized by Peptide2.0. These include the sorting signal containing peptides (FITC-KNAGFELPLTGGSGRI and unlabeled KNAGFELPLTGGSGRI), as well as a pilin motif containing peptide (DGWLQDVHVYPKHQALS). The AlexaFluor546-CNAGFELPATGGSRI was created by expressing pSUMO-CNAGFELPATGGSRI in BL21 (DE3) cells and purifying the His-tagged fusion protein with HisPure Co^{2+} IMAC resin, as described above. The desired peptide was obtained by cleaving it from SUMO by treatment with His_6-ULP1 protease, followed by passage of the mixture over a Co^{2+} IMAC column. The isolated

peptide was then reduced by adding 1 mM tris(2-carboxyethyl)phosphine (TCEP) and then selectively modified at its N-terminal cysteine residue with AlexaFluor546-maliemide (Invitrogen). The fluorescent peptide was then purified by reversed phase HPLC using a Waters XSelect HSS C18 reversed phase column. For the conjugation experiments, all peptides were dissolved into assay buffer (below) and stored as 2 mM stock solutions.

^{Cd}SrtA catalyzed conjugation reactions. The *in vitro* reactions demonstrating ^{Cd}SrtA-catalyzed ligation of the ^{C}SpaA and ^{N}SpaA domains were performed at room temperature as previously described[2]. Briefly, all proteins were dissolved in assay buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 5 mM DTT). In the reactions the following protein concentrations were used: 100 μM ^{Cd}SrtA enzyme (wild-type or enzyme mutants) and 300 μM ^{C}SpaA and ^{N}SpaA. Reactions were incubated at room temperature for 0, 24 and 48 hours, and stirred gently by continuous rotation. The reaction aliquots were quenched by adding two volume equivalents of SDS-loading dye. The reaction components were then separated using a pre-cast NuPAGE 12% linear Bis-Tris protein gel (Thermo Scientific). Gels were stained with coomassie, scanned, and analyzed using ImageJ (NIH) software[3].

The following procedure was used for ^{Cd}SrtA^{3M} catalyzed peptide labeling reactions. All proteins were diluted in assay buffer. In all reactions, 50 μl reaction volumes consisting of 100 μM ^{Cd}SrtA^{3M} were incubated with 1 mM peptide and 100 μM of either ^{N}SpaA or ^{C}SpaA. The reactions were allowed to proceed for 24 hours. Three distinct reactions were performed to investigate enzyme's substrate specificity. In reaction #1, ^{Cd}SrtA^{3M} was incubated with the pilin motif peptide (DGWLQDVHVYPKHQALS) and a sorting signal peptide (KNAGFELPLTGGSGRI). In reaction #2, ^{Cd}SrtA^{3M} was incubated with ^{C}SpaA and the pilin motif peptide. In reaction #3, ^{Cd}SrtA^{3M} was incubated with ^{N}SpaA and the sorting signal peptide. After 24 hours, each reaction was diluted 10-fold by adding ddH_2O, mixed with 2,5-dihydroxybenzoic

acid matrix (DHB) and analyzed by matrix-assisted laser desorption ionization mass spectrometry (MALDI-MS).

Orthogonal labeling procedure. For these studies, a protein construct containing two distinct nucleophiles (called G_5-$SUMO_{PM}$) was employed. G_5-$SUMO_{PM}$ was constructed by engineering a pentaglycine motif to the N-terminus of a fusion protein consisting of the Small Ubiquitin-like Modifier (SUMO) protein with PM ($SpaA_{30-194}$) fused to C-terminus. Orthogonal labeling was achieved using a three step following procedure. First, G_5-$SUMO_{PM}$ was diluted to 100 μM in assay buffer and then incubated with 100 μM $^{Cd}SrtA^{3M}$ and 1 mM $^{FITC}LPLTG$ for 16 hours. Second, excess $^{FITC}LPLTG$ was removed from the G_5-$SUMO_{PM}$,-FITC product (species 2, Fig. 3) using a Zeba Desalting Column (Thermo Scientific). Third, G_5-$SUMO_{PM}$,-FITC was incubated with 25 μM $^{Sa}SrtA$ and 500 μM AlexaFluor-LPATG peptide for either 15 minutes to achieve partial labeling or 2 hours for complete labeling. The $^{Sa}SrtA$ variant enzyme employed is the previously described "4M" variant that contains four activating mutations (P94S/D160N/D165A/K196T)[4]. The reaction was terminated at various steps in the procedure for analytical purposes. This was achieved by adding two volume equivalents of SDS-loading dye to each aliquot, followed by separation of the reaction components at different time points on a pre-cast NuPAGE 12% linear Bis-Tris protein gel (Thermo Scientific). Electrophoresis was performed at 170V for 70 min to achieve separation between apo-$^{N}SpaA$ and $^{N}SpaA$ modified by $^{FITC}LPLT$. Fluorescence data was acquired using a Pharos FX gel imager (BioRad). Fluorescein isothiocyanate (FITC) was detected by excitation with a 488nm laser line and detection with a 515-545nm emission filter. AlexaFluor546 was detected by excitation with a 532nm laser line and detection by an emission filter at 580-630nm. Afterwards, the SDS-PAGE gels were stained with Coomassie to detect all proteins (fluorescent and unlabeled) (**Fig. 3B**). In separate studies $^{Cd}SrtA^{3M}$ was used to singly label $^{N}SpaA$ with the $^{FITC}LPLTG$ peptide (**Fig. 2B**).

These studies were performed using reaction conditions that were to identical to those used for first step of the orthogonal labeling method.

HPLC-based hydrolysis assay. *In vitro* hydrolysis reactions were performed using the method developed by Kruger et al[5]. Briefly, 50 μM sortase enzyme (either wild-type ^{Cd}SrtA and ^{Sa}SrtA, or their mutants) was incubated with 500 μM KNAGFELPLTGGSGRI ($LPLTG_{pep}$) or KNAGFELPATGGSGRI ($LPATG_{pep}$) and 5 mM DTT in 100 μl reactions at 37°C for 24 and 48 hours. The reactions were then quenched by adding 50 μl of 1 M HCl and separated by reverse phase HPLC using a Waters XBridge Peptide BEH C18 column. Peptides were eluted by applying a gradient from 5 to 51% acetonitrile (in 0.1% trifluoroacetic acid) over 30 minutes at a flow rate of 1 ml/min. Elution of the peptides was monitored by absorbance at 215 nm. For the quantitative analysis peaks in the HPLC data were integrated using the program Graphical Analysis (Vernier).

Tandem mass spectrometry. Protein digestion and isopeptide bond identification were performed according to a previously described protocol[2,6,7]. Specifically, proteins entrapped in gel bands were reduced with 10 mM dithiothreitol (Sigma) at 60°C for an hour and then alkylated with 50 mM iodoacetamide (Sigma) at 45°C for a few minutes in the dark. Samples were then digested with 200 ng trypsin (Thermo Scientific) at 37°C overnight. At the end of trypsin digestion, zinc acetate was added to the solution to a final concentration of 2.5 mM. Asp-N endoproteinase (200ng; Thermo Scientific) was then added for another overnight incubation. Digested peptides were extracted from the gel bands in 50% acetonitrile/49.9% water/0.1% trifluoroacetic acid (TFA) and cleaned with C18 StageTip[8] before mass spectrometry analysis. Digested peptides were separated on an EASY-Spray column (25 cm × 75 μm ID, PepMap RSLC C18, 2 μm, Thermo Scientific) connected to an EASY-nLC 1000 nUPLC (Thermo Scientific) using a gradient of 5 - 35% acetonitrile in 0.1% formic acid, and a flow rate of 300

nl/min (total time 30 minutes). Tandem mass spectra were acquired in a data-dependent manner with an Orbitrap Q Exactive mass spectrometer (Thermo Fisher Scientific) interfaced to a nanoelectrospray ionization source. The raw MS/MS data were converted into MGF format by Thermo Proteome Discoverer 1.4 (Thermo Scientific). An in-house program was used to search for the isopeptides. The approach was based on the observation of published spectra[6] as well as our own on the presence of ions specific for the fragments of ELPLT (m/z 215.138, 225.122, 243.132, 294.180, 312.190, 322.174 and 340.186).

Sortase Protein Sequences:

$^{Cd}SrtA^{WT}$ (37-257)

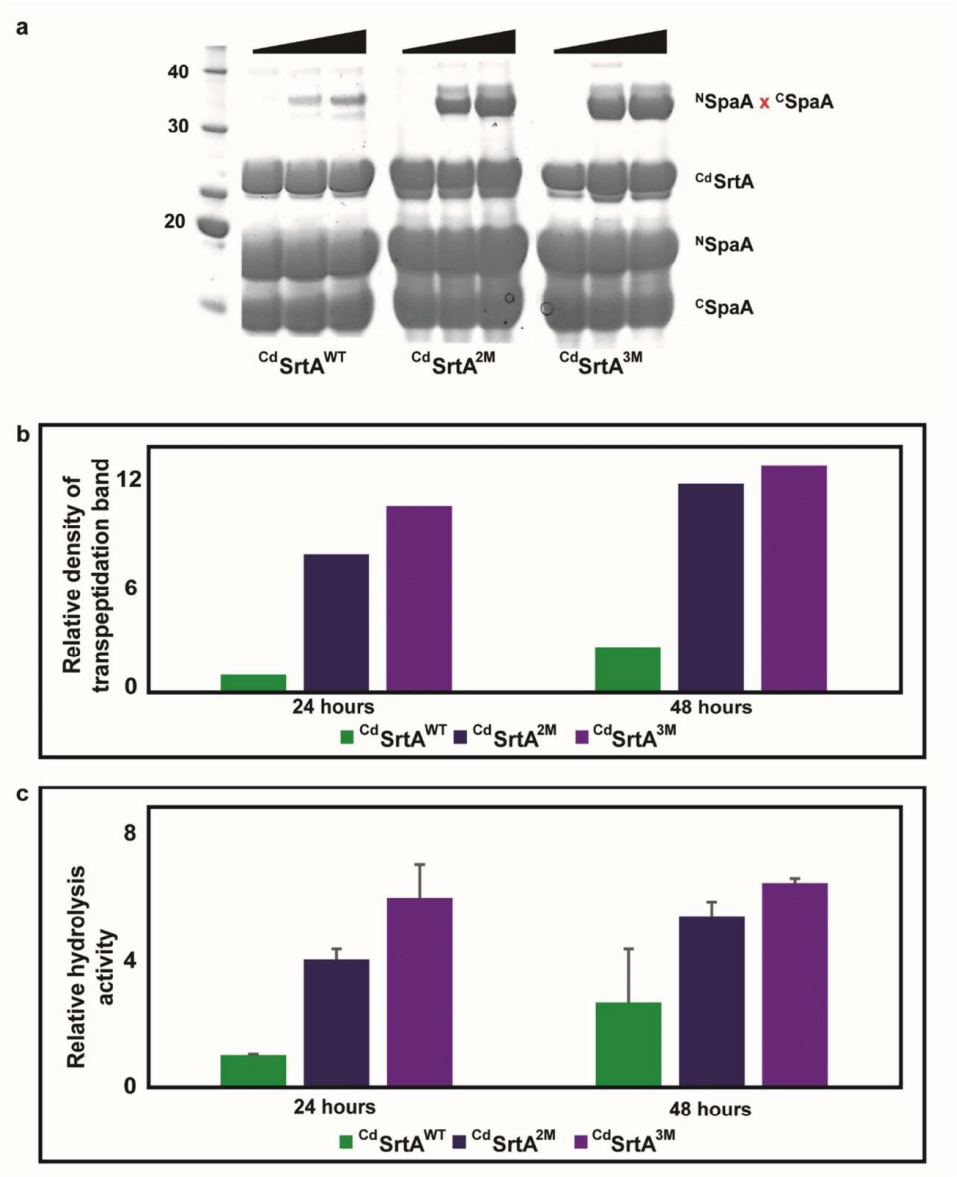
NNARQARVAQSYENSYEVDSPAVRDSVLEAARQYNTSVVGFPILDPWLNRASKNSGPYLDYLQQLNPQRA
ERPVIASISIPTIDAHLPIYHGTDTATLEHGLGHLYGSALPVGGTGTHPVITGHSGLANATLFDNLEDVK
EHDPIYITVQGETLKYEVDAINVVLPEDTKLLAPDPNKDQITLITCTPYAVNSHRLLVRAHRVDLDPNDP
NLTQTGTKIWQ

$^{Cd}SrtA^{2M}$ (37-257)

NNARQARVAQSYENSYEVDSPAVRDSVLEAARQYNTSVVGFPIL**G**P**G**LNRASKNSGPYLDYLQQLNPQRA
ERPVIASISIPTIDAHLPIYHGTDTATLEHGLGHLYGSALPVGGTGTHPVITGHSGLANATLFDNLEDVK
EHDPIYITVQGETLKYEVDAINVVLPEDTKLLAPDPNKDQITLITCTPYAVNSHRLLVRAHRVDLDPNDP
NLTQTGTKIWQ

$^{Cd}SrtA^{3M}$ (37-257)

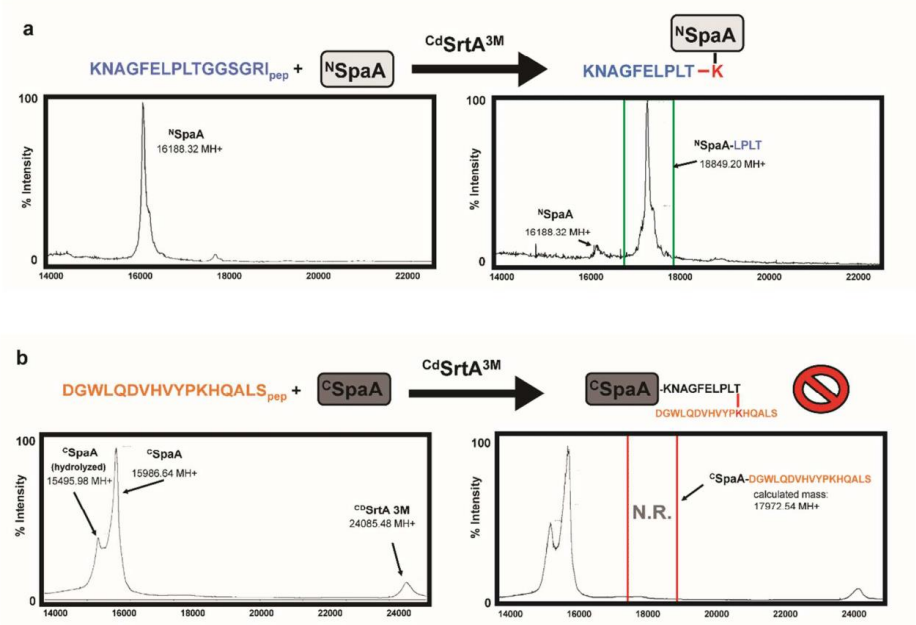
NNARQARVAQSYENSYEVDSPAVRDSVLEAARQYNTSVVGFPIL**G**P**G**L**A**RASKNSGPYLDYLQQLNPQRA
ERPVIASISIPTIDAHLPIYHGTDTATLEHGLGHLYGSALPVGGTGTHPVITGHSGLANATLFDNLEDVK
EHDPIYITVQGETLKYEVDAINVVLPEDTKLLAPDPNKDQITLITCTPYAVNSHRLLVRAHRVDLDPNDP
NLTQTGTKIWQ


a
40
30
20
NSpaA x CSpaA
CdSrtA
NSpaA
CSpaA
CdSrtAWT
CdSrtA2M
CdSrtA3M
b
Relative density of transpeptidation band
12
6
0
24 hours
48 hours
CdSrtAWT
CdSrtA2M
CdSrtA3M
c
Relative hydrolysis activity
8
4
0
24 hours
48 hours
CdSrtAWT
CdSrtA2M
CdSrtA3M

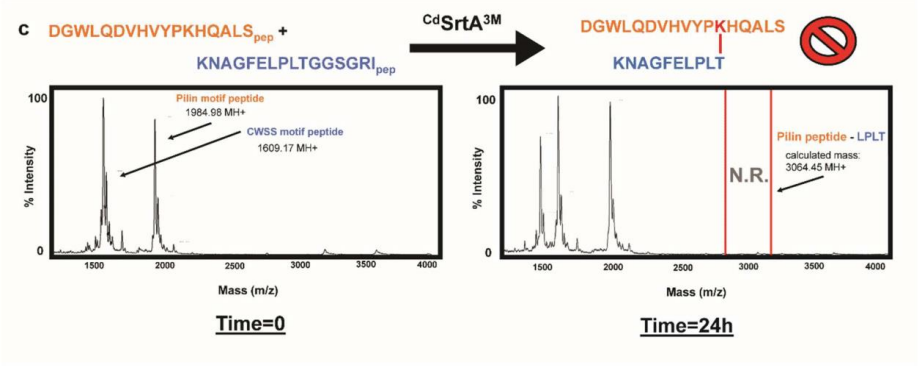
S8

Supplementary Figure 1:

Relative activities of each ^{Cd}SrtA mutant. **A)** The three ^{Cd}SrtA mutants tested demonstrate differential transpeptidation activities upon incubation with SpaA pilin domains. Each ^{Cd}SrtA mutant (WT, D81G/W83G, D81G/W83G/N85A) was incubated with ^{N}SpaA and ^{C}SpaA for 24 and 48 hours, and the reactions were separated by SDS-PAGE. Notably, the WT construct has negligible transpeptidation activity *in vitro,* presumably because of substrate occlusion from the active site by the lid structure. **B)** Quantification of the relative amounts of ^{N}SpaA-^{C}SpaA isopeptide-linked dimer produced was determined by densitometry analysis of the gel images in ImageJ[3]. Densitometry values for all reactions were plotted as factor of the transpeptidation product formed by ^{Cd}SrtA WT after 24hrs. **C)** Peptide hydrolysis data indicates ^{Cd}SrtA^{3M} forms acylation product faster than the other mutants tested. This HPLC-based assay measures the cleavage and acylation of the cell wall sorting signal peptide, which is the first step of sortase catalysis. Each ^{Cd}SrtA construct was incubated with a 10x molar excess of LPLTG peptide (KNAGFELPLTGGSRI) for 24 and 48hrs to measure their respective hydrolysis activities. Integration of the peak corresponding the full length LPLTG peptide remaining after incubation with sortase reports on the amount of peptide processed by each construct. Those integrals were determined and normalized activity (relative to the extent of peptide processing by ^{Cd}SrtA WT after 24 hours) is presented. For both bar graphs (S1B,S1C): ^{Cd}SrtA WT (green), ^{Cd}SrtA^{2M} (dark blue), and ^{Cd}SrtA^{3M} (purple).

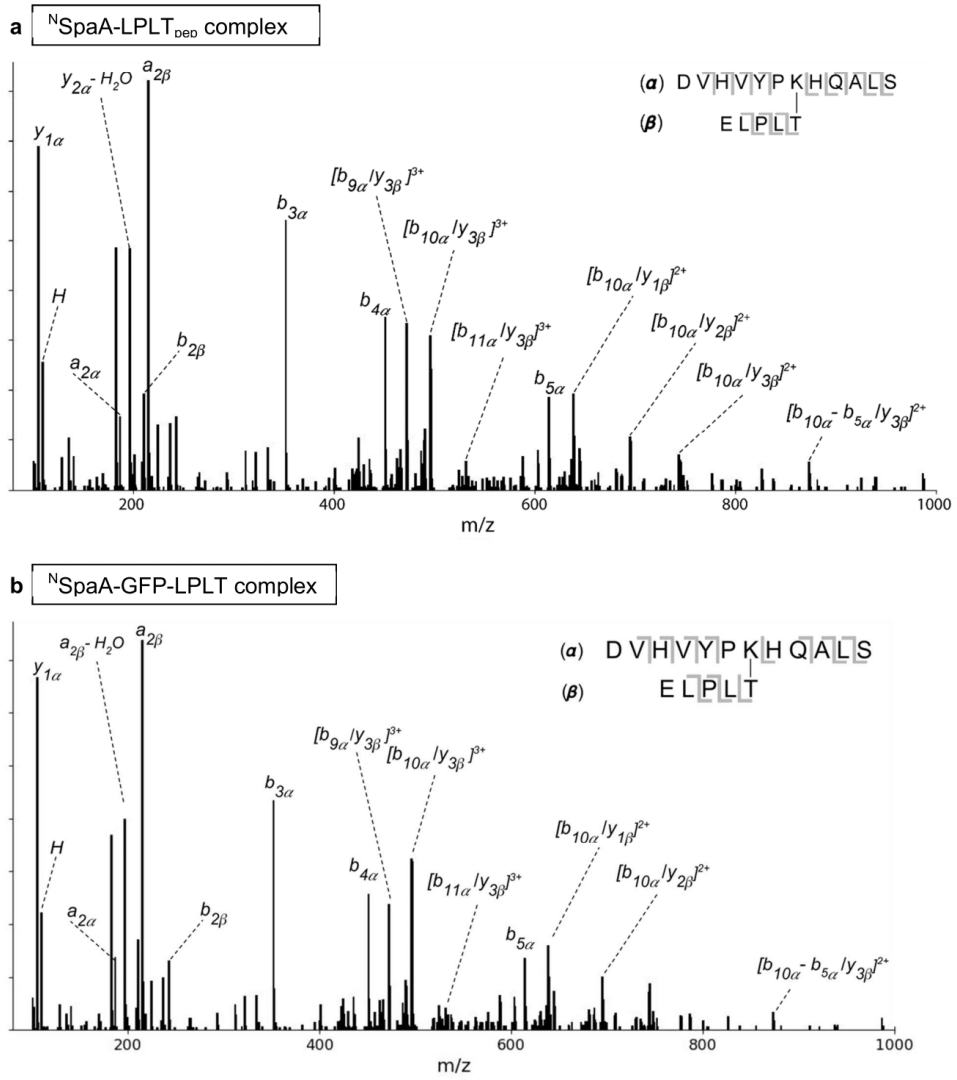


Supplementary Figure 2:



Substrate recognition determinants of $^{Cd}SrtA^{3M}$ are determined by MALDI-MS. **A)** Incubation of equimolar $^{N}SpaA$ and $^{Cd}SrtA^{3M}$ with a 10x molar excess of 'sort-tag' peptide (KNAGFELPLTGGSGRI) results in robust and specific modification of Lys190 of $^{N}SpaA$, as shown by the characteristic mass shift in the MALDI spectra. **B)** Incubation of $^{Cd}SrtA^{3M}$ with

equimolar amounts of ^{C}SpaA and 10 molar equivalents of pilin motif peptide (DGWLQDVHVYPKHQALS) does not result in any detectable ligation product as determined by MALDI-MS. **C)** Incubation of ^{Cd}SrtA^{3M} with 10 molar equivalents of pilin motif peptide and LPLTG peptide also does not yield a detectable linkage by MALDI. For each spectra, the calculated mass of the theoretical transpeptidation linkage for each reaction is highlighted by a dashed red box. N.R. = no reaction detected by MALDI-MS



Supplementary Figure 3:

Tandem mass spectrometry (MS/MS) data used to characterize the linkage formed between ^{N}SpaA and a synthetic LPLTG peptide catalyzed by ^{Cd}SrtA^{3M}, in the ^{N}SpaA-LPLT complex **(A)** and the GFP-LPLT-^{N}SpaA conjugate **(B)**. The linkage occurs between the side chain ε-amine of Lys190 in the pilin motif and the carboxyl terminus of the threonine in the synthetic peptide or

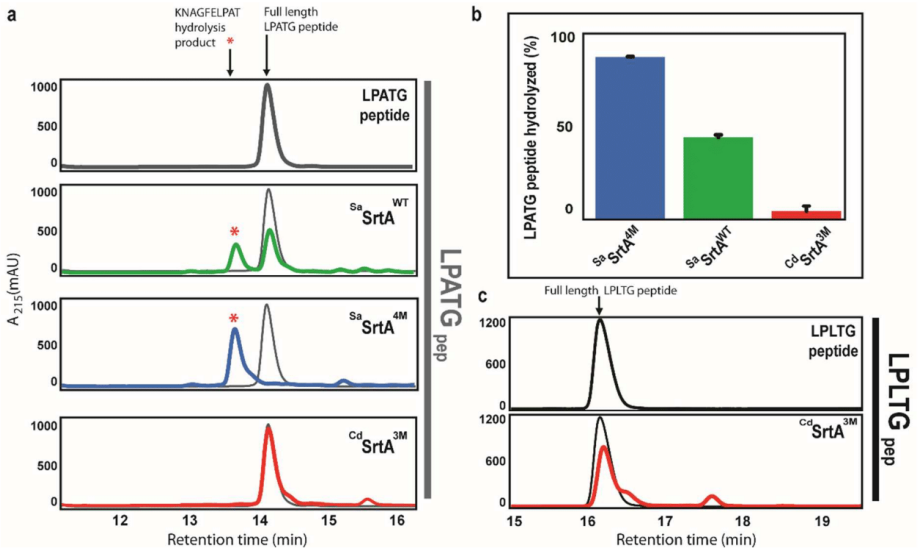
genetically appended C-terminal sorting signal of GFP-LPLTG, as previously reported for the linkage between ^{C}SpaA and ^{N}SpaA domains[2]. Shown is the mass-to-charge ratio (m/z) tandem mass spectrum of the linked peptide (sequence shown in inset; *α* is the fragment sequence from ^{N}SpaA and *β* is the fragment sequence from the synthetic LPLTG peptide or C-terminus of GFP-LPLTG). Detected fragment ion masses are reported in Supplemental Tables 1 and 2 and their origins are labeled in the inset. The MS/MS product ion labels list the fragment origins from the two polypeptide sources. For example, the $b_{3\alpha}$ product ion is a b_3 product from the *α* peptide (sequence DVH), whereas the $b_{9\alpha}/y_{3\beta}$ product ion is a $b_{9\alpha}$ product joined to a $y_{3\beta}$ product through the K-T linkage.

Table 1: Tandem mass spectrometry analysis of ^{N}SpaA-LPLT complex

Fragment Ion	Charge	Observed *m/z*	Theoretical *m/z*	Difference between observed and theoretical (ppm)
$y_{1\alpha}$	+1	106.051	106.050	-9.430
histidine immonium	+1	110.072	110.072	0
$a_{2\alpha}$	+1	187.109	187.108	5.345
$a_{2\beta}$ -H_2O	+1	197.130	197.128	10.146
$b_{2\alpha}$	+1	215.105	215.103	9.298
$a_{2\beta}$	+1	215.141	215.139	9.296
$y_{2\alpha}$	+1	219.135	219.134	4.563
$b_{2\beta}$	+1	243.136	243.134	8.226
$y_{3\beta}$-H_2O	+1	312.194	312.191	9.610
$b_{3\beta}$	+1	340.189	340.187	5.879
$b_{3\alpha}$	+1	352.164	352.162	5.679
$y_{4\alpha}$	+1	418.231	418.230	2.391
$b_{4\alpha}$	+1	451.233	451.230	6.648
$b_{4\beta}$	+1	453.269	453.271	-4.412
$b_{9\alpha}/y_{3\beta}$	+3	472.589	472.586	6.348
$b_{10\alpha}/y_{3\beta}$	+3	496.268	496.265	6.045
$b_{11\alpha}/y_{3\beta}$	+3	533.963	533.960	5.618
$y_{5\alpha}$	+1	555.294	555.289	9.004
$b_{5\alpha}$	+1	614.297	614.293	6.512
$b_{10\alpha}/y_{1\beta}$	+2	638.829	638.825	6.261
$b_{10\alpha}/y_{2\beta}$	+2	695.372	695.367	7.190
$b_{10\alpha}/y_{3\beta}$	+2	743.900	743.893	9.410
$b_{10\alpha}/y_{4\beta}$	+2	800.439	800.436	3.748
$(b_{10\alpha}-b_{5\alpha})/y_{3\beta}$	+1	873.499	873.494	5.724
$(b_{11\alpha}-b_{5\alpha})/y_{3\beta}$	+1	986.586	986.578	8.109

Table 2: Tandem mass spectrometry analysis of GFP-LPLT-^{N}SpaA complex

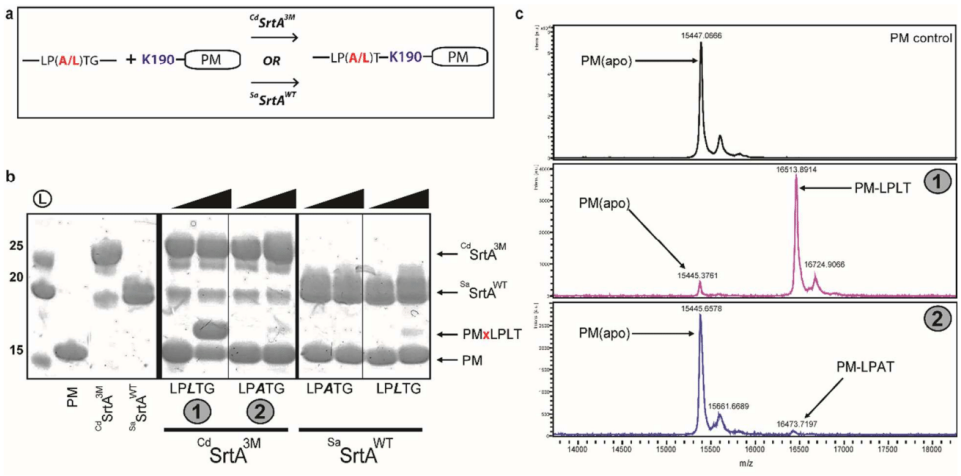
Fragment Ion	Charge	Observed *m/z*	Theoretical *m/z*	Difference between observed and theoretical (ppm)
$y_{1\alpha}$	+1	106.050	106.050	0
histidine immonium	+1	110.071	110.072	9.085
$a_{2\alpha}$	+1	187.107	187.108	5.345
$a_{2\beta}$ -H_2O	+1	197.128	197.128	0
$b_{2\alpha}$	+1	215.103	215.103	0
$a_{2\beta}$	+1	215.138	215.139	4.648
$y_{2\alpha}$	+1	219.134	219.134	0
$b_{2\beta}$	+1	243.133	243.134	4.113
$y_{3\beta}$-H_2O	+1	312.190	312.191	3.203
$b_{3\beta}$	+1	340.186	340.187	2.940
$b_{3\alpha}$	+1	352.160	352.162	5.679
$b_{4\alpha}$	+1	451.228	451.230	4.432
$b_{9\alpha}/y_{3\beta}$	+3	472.584	472.586	4.232
$b_{10\alpha}/y_{3\beta}$	+3	496.263	496.265	4.030
$b_{11\alpha}/y_{3\beta}$	+3	533.956	533.960	7.491
$y_{5\alpha}$	+1	555.283	555.289	10.805
$b_{5\alpha}$	+1	614.290	614.293	4.884
$b_{10\alpha}/y_{1\beta}$	+2	638.822	638.825	4.696
$b_{10\alpha}/y_{2\beta}$	+2	695.365	695.367	2.876
$b_{10\alpha}/y_{3\beta}$	+2	743.889	743.893	5.377
$b_{10\alpha}/y_{4\beta}$	+2	800.434	800.436	2.499
$(b_{11\alpha}-b_{5\alpha})/y_{3\beta}$	+1	986.572	986.578	6.082



Supplementary Figure 4:

In the hydrolysis reaction $^{Cd}SrtA^{3M}$ exhibits specificity for a leucine amino acid at the 'X' position within the LPXTG substrate. Incubation of $^{Cd}SrtA^{3M}$ and $^{Sa}SrtA$ with a short peptide containing an LPATG motif (KNAGFELPATGGSGRI) indicates that $^{Cd}SrtA^{3M}$ is unable to process a significant amount of peptide, while $^{Sa}SrtA$ robustly hydrolyzes the LPATG peptide, as expected. The improved variant, $^{Sa}SrtA\ 4M^{4}$, almost entirely hydrolyzes the full length peptide after 24 hours. This data also indicates that $^{Cd}SrtA^{3M}$ can hydrolyze a peptide that contains a leucine at X positon, indicating that it recognizes this site. **A)** Representative HPLC chromatographs showing the hydrolytic cleavage of a LPATG containing peptide by $^{Cd}SrtA^{3M}$ and $^{Sa}SrtA$. The traces show the peptide control before adding enzyme (grey), and after the addition of $^{Cd}SrtA^{3M}$ (red), $^{Sa}SrtA$ WT (green) and $^{Sa}SrtA$ 4M (blue). In each chromatogram, the trace of unhydrolyzed LPATG peptide is superimposed for comparison. The retention time of full length LPATG peptide and hydrolysis product peptide peaks are indicated with arrows at the top of the chromatographs for clarity. **B)** Quantification of the hydrolysis activity data from panel A. The activity of each enzyme is plotted as the percentage of LPATG peptide hydrolyzed after 24 hours. This was

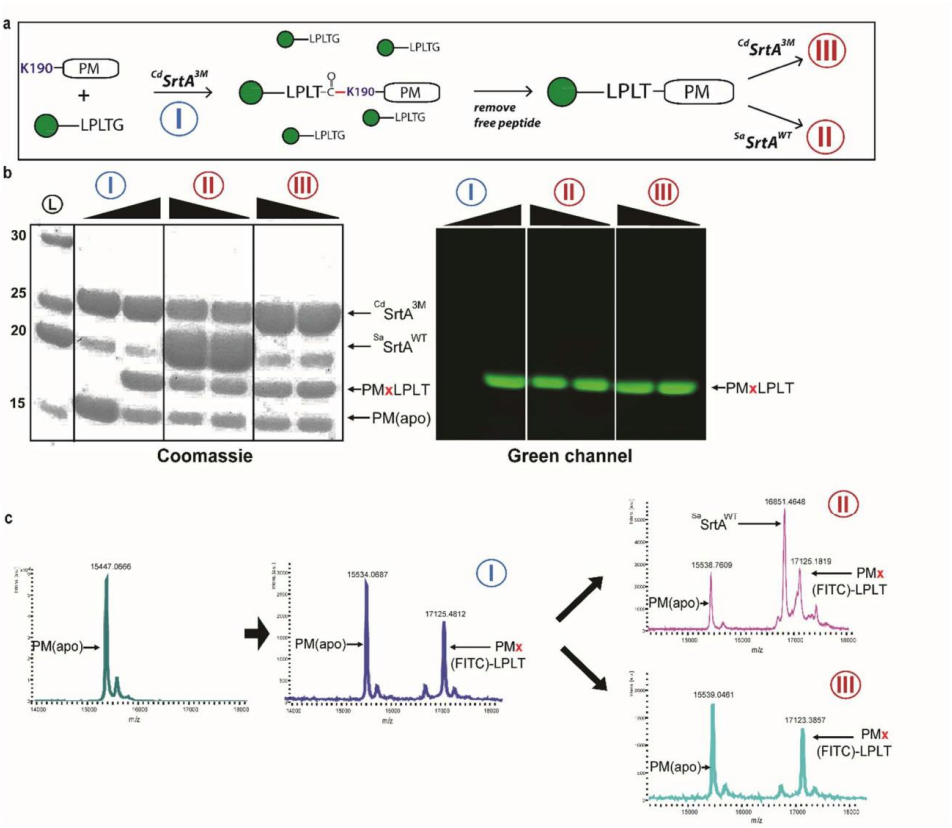
determined by integrating the peak corresponding to the full length peptide before and after treatment with each enzyme. **C)** Data showing that $^{Cd}SrtA^{3M}$ can hydrolyze a peptide containing the sequence (KNAGFELPLTGGSGRI, (LPLTG peptide)). The enzyme and peptide were incubated for 24 hours and then separated by HPLC. The chromatographs show the uncleaved peptide in the absence (top, black) and presence of $^{Cd}SrtA^{3M}$ (bottom, red). A reduction in peak intensity in the presence of $^{Cd}SrtA^{3M}$ demonstrates that the enzyme cleaves the peptide.



Supplementary Figure 5:

In the transpeptidation reaction ^{Cd}SrtA^{3M} exhibits specificity for a leucine amino acid at the 'X' position within the LPXTG substrate. **A)** Schematic outlining the reactions used to generate the data shown in panels B and C. In separate reactions, the ^{N}SpaA domain (called PM) was incubated with either ^{Cd}SrtA^{3M} or ^{Sa}SrtA, and peptides containing the sequence LPLTG or LPATG (KNAGFELPLTGGSGRI and KNAGFELPATGGSGRI, respectively). The components were incubated for 16 hours and then separated by SDS-PAGE or analyzed directly by MS to determine the amount of unmodified-PM from PM-LP(A/L)T isopeptide linked conjugates. Reaction 1: ^{Cd}SrtA^{3M} + PM + LPLTG$_{pep,}$ Reaction 2: ^{Cd}SrtA^{3M} + PM + LPATG$_{pep,}$ Reaction 3: ^{Sa}SrtA WT + PM + LPATG$_{pep,}$ Reaction 4: ^{Sa}SrtA WT + PM + LPLTG$_{pep}$. Reaction conditions are similar to those used for Figure 2. **B)** SDS-PAGE analysis of the reactions conducted for 0 and 24 hours. A protein band corresponding to the PM-LPLT conjugate is only observed for Reaction 1 (PMxLPXT), indicating that ^{Cd}SrtA^{3M} only uses the LPLTG peptide as a substrate in the transpeptidation reaction. For reference, SDS-PAGE data for the isolated enzymes and PM substrate are shown on the left. **C)** Analysis of reactions 1 and 2 by MALDI-MS. The circled

numbers correspond to the reaction number in panel A. The mass spectrum of the isolated PM is displayed for comparison (labeled PM(apo), top). The PM-LPLT conjugate has an expected mass shift of 1089 Da. The enzyme exhibits specificity for leucine at the 'X' position, as a strong peak at this value is detected for reaction 1, while only an extremely weak peak at this position is detected in reaction 2.



Supplementary Figure 6:

The isopeptide bond created by ^{Cd}SrtA^{3M} is not hydrolyzed significantly by either ^{Sa}SrtA or ^{Cd}SrtA^{3M}. **A)** A schematic depicting the experiments used to test whether the isopeptide bond is cleaved by ^{Sa}SrtA or ^{Cd}SrtA^{3M}. ^{N}SpaA (called PM) was incubated with ^{Cd}SrtA^{3M} and FITC-LPLTG for 24 hours to create the isopeptide-linked FITC-LPLTG-PM (reaction I, conditions as in Figure 2). Free peptide was then removed using a desalting column. To determine if the isopeptide bond is susceptible to proteolysis, either ^{Sa}SrtA WT (reaction II) or ^{Cd}SrtA^{3M} (reaction III) was

then added to a final concentration of 200 μM. The reactions were performed for either 0 or 24 hours, and the integrity of the isopeptide linkage was then determined using either SDS-PAGE or MS. **B)** SDS-PAGE analysis. The reaction mixtures were separated by SDS-PAGE and visualized by both Coomassie staining and fluorescent imaging (excitation at 488 nm and detection with a 515-545 nm emission filter). Reaction I in both panels shows that the FITC-LPLT-PM conjugate is produced and that it can be distinguished from apo-PM by its reduced electrophoretic mobility and its fluorescence. After exposure to either ^{Sa}SrtA (reaction II) or ^{Cd}SrtA^{3M} (reaction III) the intensity of the band corresponding to FITC-LPLT-PM is not reduced, indicating that the isopeptide linkage is stable and not hydrolyzed by the sortases. In the gel visualized with fluorescent imaging (right), lanes 1 to 2 correspond to the transpeptidation reaction mixture sampled at time zero and 24 hours, respectively. Lanes 3/4 and 5/6 sample the reaction after adding ^{Sa}SrtA or ^{Cd}SrtA^{3M}, either immediately after removal of peptide and addition of each sortase, or after 24 hours. The Coomassie stained gel (left) shows the same data, but also includes a protein molecular weight ladder (far left). **C)** MALDI-MS analysis of the reactions. The initial conjugation reaction with ^{Cd}SrtA^{3M} proceeds to ~40% modification as estimated by MALDI-MS results. Conversion of PM to FITC-LPLT-PM is indicated by a mass shift of 1592 Da (central panel). After prolonged incubation of the FITC-LPLT-PM conjugate with either ^{Sa}SrtA WT or ^{Cd}SrtA^{3M}, the isopeptide conjugate species persists, as evidenced by the constant relative intensity ratios of the PM vs FITC-LPLT-PM peaks in the MALDI spectra. In reaction II, the peak corresponding to ^{Sa}SrtA^{WT} can be observed in the middle of the spectrum, at 16851 Da. The identities of the reaction components are labeled in panels B and C.

Supplementary References:

(1) Bornhorst, J. A.; Falke, J. J. *Methods Enzymol.* **2000**, *326*, 245–254.

(2) Chang, C.; Amer, B. R.; Osipiuk, J.; McConnell, S. A.; Huang, I.-H.; Hsieh, V.; Fu, J.; Nguyen, H. H.; Muroski, J.; Flores, E.; Ogorzalek Loo, R. R.; Loo, J. A.; Putkey, J. A.; Joachimiak, A.; Das, A.; Clubb, R. T.; Ton-That, H. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, E5477–E5486.

(3) Schneider, C. A.; Rasband, W. S.; Eliceiri, K. W. *Nat. Methods* **2012**, *9*, 671–675.

(4) Chen, I.; Dorr, B. M.; Liu, D. R. *Proc. Natl. Acad. Sci.* **2011**, *108*, 11399–11404.

(5) Kruger, R. G.; Dostal, P.; McCafferty, D. G. *Anal. Biochem.* **2004**, *326*, 42–48.

(6) Kang, H. J.; Paterson, N. G.; Gaspar, A. H.; Ton-That, H.; Baker, E. N. *Proc. Natl. Acad. Sci.* **2009**, *106*, 16967–16971.

(7) Thevis, M.; Ogorzalek Loo, R. R.; Loo, J. A. *J. Proteome Res.* **2003**, *2*, 163–172.

(8) Rappsilber, J.; Mann, M.; Ishihama, Y. *Nat. Protoc.* **2007**, *2*, 1896–1906.

Appendix III: N-terminal chemistry as a means of identifying isopeptide bond formation

As an extension of the work of identifying the isopeptide bonds used for sortase, an attempt was made to use N-terminal chemistry to assist in the identification. Specifically, phenylisothiocyanate (PITC) was added to the N-terminal of the isopeptide bond. The rationale was two-fold. By modifying the N-terminal of an isopeptide bond, we would expect that a mass shift of 2 PITC additions may be identified, whereas a standard peptide would only have one. Also, PITC results in the formation of very strong b_1 ions upon collisionally induced dissociation (1). By using this, we may be able to more easily identify isopeptide bonds. The samples obtained by the Clubb lab from the work performed for the work published in McConnell *et al*. (2) were modified with PITC by the methods described in Wang *et al.*(1). Peptides were digested with AspN and trypsin. The spectrum obtained and assignments are presented. Future work would seek to optimize a protocol to ensure efficient modification of two N-terminal peptides. Using the strong presence of two b_1 ions to identify non-specific isopeptide bonding is an appealing idea. Further optimization as to the protocol and the use of similar PITC analogues such as 4-sulfophenyl isothiocyanate and optimization of proteases used for cleavage would be necessary.

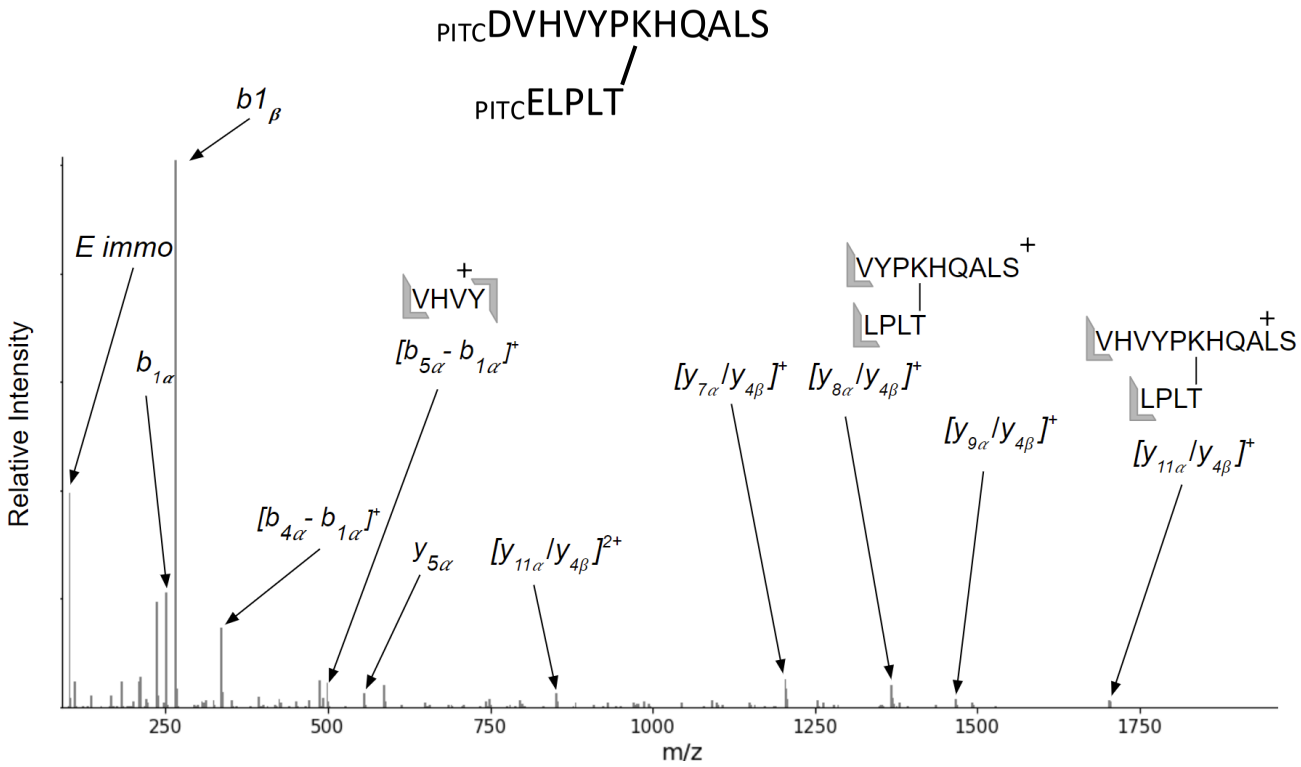


Figure 1. Spectrum of PITC derived isopeptide bond. The iospeptide bond between the DVHVYPKHQALS sequence (α chain) and ELPLT (β chain). Several fragments are displayed above their annotated peaks to assist with nomenclature.

.

Ion Identified	Charge	Observed *m/z*	Theoretical *m/z*	Difference between observed and theoretical (ppm)
Glutamate immonium	+1	102.056	102.055	-2.352
$y_{2\alpha}$	+1	219.134	219.134	0.091
$a_{1\beta}$	+1	237.069	237.069	-0.422
$b_{1\alpha}$	+1	251.048	251.049	-0.757
$b_{1\beta}$	+1	265.064	265.064	-1.170
$[b_{4\alpha}\text{-}\ b_{1\alpha}]$	+1	336.203	336.203	-0.892
$y_{4\alpha}$	+1	418.230	418.230	0.669
$[b_{5\alpha}\text{-}\ b_{1\alpha}]$	+1	499.266	499.266	-1.282
$y_{5\alpha}$	+1	555.289	555.289	0.540
$[y_{11\alpha}/y_{4\beta}]$	+2	851.986	851.986	0.246
$[y_{7\alpha}/y_{4\beta}]$	+1	1204.703	1204.705	-1.154
$[y_{8\alpha}/y_{4\beta}]$	+1	1367.766	1367.768	-1.704
$[y_{9\alpha}/y_{4\beta}]$	+1	1466.843	1466.837	4.609
$[y_{11\alpha}/y_{4\beta}]$	+1	1702.945	1702.964	-11.010
M*	+2	1109.035	1109.035	0.451
M*	+3	739.693	739.692	1.163

Table 1. Ions identified in the spectra. Ions assigned in the spectra of PITC associated N-terminal isopeptide linkage. (*) Denotes precursor ions were identified in MS^1 spectra, not the fragmentation spectra.

APPENDIX IV: Kinetics and Optimization of the Lysine–Isopeptide Bond Forming Sortase Enzyme from *Corynebacterium diphtheriae*

Bioconjugate Chemistry

pubs.acs.org/bc Article

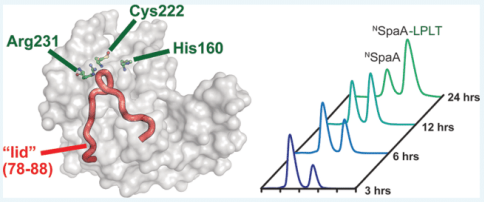
Kinetics and Optimization of the Lysine–Isopeptide Bond Forming Sortase Enzyme from *Corynebacterium diphtheriae*

Christopher K. Sue, Scott A. McConnell, Ken Ellis-Guardiola, John M. Muroski, Rachel A. McAllister, Justin Yu, Ana I. Alvarez, Chungyu Chang, Rachel R. Ogorzalek Loo, Joseph A. Loo, Hung Ton-That, and Robert T. Clubb*

Cite This: https://dx.doi.org/10.1021/acs.bioconjchem.0c00163 Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Site-specifically modified protein bioconjugates have important applications in biology, chemistry, and medicine. Functionalizing specific protein side chains with enzymes using mild reaction conditions is of significant interest, but remains challenging. Recently, the lysine–isopeptide bond forming activity of the sortase enzyme that builds surface pili in *Corynebacterium diphtheriae* (^{Cd}SrtA) has been reconstituted *in vitro*. A mutationally activated form of ^{Cd}SrtA was shown to be a promising bioconjugating enzyme that can attach Leu-Pro-Leu-Thr-Gly peptide fluorophores to a specific lysine residue within the N-terminal domain of the SpaA protein (^{N}SpaA), enabling the labeling of target proteins that are fused to ^{N}SpaA. Here we present a detailed analysis of the ^{Cd}SrtA catalyzed protein labeling reaction. We show that the first step in catalysis is rate limiting, which is the formation of the ^{Cd}SrtA-peptide thioacyl intermediate that subsequently reacts with a lysine ϵ-amine in ^{N}SpaA. This intermediate is surprisingly stable, limiting spurious proteolysis of the peptide substrate. We report the discovery of a new enzyme variant (^{Cd}SrtA$^{\Delta}$) that has significantly improved transpeptidation activity, because it completely lacks an inhibitory polypeptide appendage ("lid") that normally masks the active site. We show that the presence of the lid primarily impairs formation of the thioacyl intermediate and not the recognition of the ^{N}SpaA substrate. Quantitative measurements reveal that ^{Cd}SrtA$^{\Delta}$ generates its cross-linked product with a catalytic turnover number of 1.4 ± 0.004 h^{-1} and that it has apparent K_M values of 0.16 ± 0.04 and 1.6 ± 0.3 mM for its ^{N}SpaA and peptide substrates, respectively. ^{Cd}SrtA$^{\Delta}$ is 7-fold more active than previously studied variants, labeling >90% of ^{N}SpaA with peptide within 6 h. The results of this study further improve the utility of ^{Cd}SrtA as a protein labeling tool and provide insight into the enzyme catalyzed reaction that underpins protein labeling and pilus biogenesis.



INTRODUCTION

New methods are needed to create protein bioconjugates that can be used as therapeutics, imaging tools, diagnostic reagents, and materials.[1−5] Labeling specific sites on the protein is often preferred as it enables the construction of well-defined antibody–drug conjugates, small molecule- and fluorophore-labeled proteins for biophysical experiments, orientation-specific protein immobilization and the preparation of ordered, multifunctional protein complexes.[6−8] A variety of protein modification strategies have been developed that exhibit varying degrees of site-selectivity, efficiency, and ease of use. They range from chemical approaches that leverage the reactivity of amino acid specific functional groups (e.g., cysteine and lysine modifications) to highly selective, but less facile methods that require the incorporation of non-natural amino acids to facilitate bio-orthogonal conjugation chemistries (e.g., azide or alkyne-containing residues for click chemistry).[9,10] Bioconjugating enzymes (e.g., ligase, transferases, etc.) are particularly attractive for site-specific protein labeling, because they can be employed using mild reaction conditions, and in principle can be highly selective.[11,12] The sortase A enzyme from *Staphylococcus aureus* (^{Sa}SrtA) is one of the most widely used bioconjugating enzymes.[12−14] It has been successfully deployed to catalyze protein–protein ligations and backbone cyclization, and to modify proteins with peptides, lipids, sugars, and small molecules.[15−21] However, ^{Sa}SrtA bioconjugations do not readily modify protein side chains and are almost exclusively restricted to altering only the N- or C-terminus of a protein.

Recently, we demonstrated that the pilus-specific sortase from *Corynebacterium diphtheriae* (^{Cd}SrtA) can be used to

Received: March 25, 2020
Revised: May 7, 2020
Published: May 12, 2020

attach a peptide fluorophore via an isopeptide bond to a specific lysine residue within a protein.[22] Although ^{Cd}SrtA and ^{Sa}SrtA are both members of the sortase-superfamily of cysteine transpeptidases, they have distinct substrate specificities.[23] ^{Sa}SrtA is a class A sortase that catalyzes formation of backbone−backbone peptide bonds, whereas ^{Cd}SrtA is a class C sortase that joins molecules together via lysine−isopeptide bonds.[24,25] ^{Cd}SrtA assembles the SpaA pilus in *C. diphtheriae* by cross-linking SpaA "pilin" subunits via a lysine−isopeptide bond.[26,27] In this process, a lysine residue (K190) on one SpaA pilin is joined to the C-terminal LPLTG sorting signal located in a second SpaA pilin.[26] Repetition of this reaction forms the SpaA pilus, which is approximately 1−2 μm in length and further elaborated with unique tip and basal pilin proteins (Figure 1A).[27,28] *In vitro*, the native ^{Cd}SrtA enzyme is enzymatically inactive because it contains a polypeptide appendage that occludes its active site, called a "lid" (Figure 1C). However, ^{Cd}SrtA variants containing destabilizing amino acid substitutions in the lid exhibit *in vitro* activity.[28] The most active form of the enzyme thus far discovered is ^{Cd}SrtA^{3M}, which contains residues N37−Q257 in ^{Cd}SrtA and D81G/W83G/N85A substitutions in the lid.[22] ^{Cd}SrtA^{3M} is a promising bioconjugation tool, as it can be used to selectively modify proteins harboring the N-terminal domain of SpaA (^{N}SpaA) with peptide fluorophores. Modification via lysine−isopeptide bonds is attractive, as these linkages may be less susceptible to proteolysis and enzymatic reversibility.

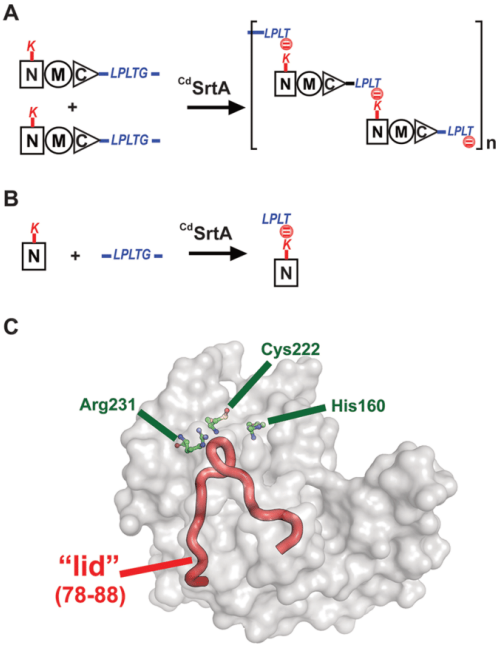


Figure 1. *C. diphtheriae* ^{Cd}SrtA pilin sortase catalyzes lysine isopeptide bond formation. (A) Schematic showing the pilin polymerization reaction catalyzed by ^{Cd}SrtA. The enzyme creates the SpaA pilus by polymerizing SpaA pilin proteins. In the reaction, it recognizes lysine (K190) side chain nucleophile within the N-terminal domain of SpaA (^{N}SpaA), and it joins the backbone threonine carbonyl carbon atom located in the C-terminal LPLTG sorting signal located within another SpaA protein. This reaction is repeated to construct the SpaA pilus that mediates bacterial adhesion. (B) Schematic of the reaction used to monitor lysine isopeptide bond formation. In this assay, the ^{Cd}SrtA enzyme ligates the isolated ^{N}SpaA domain to the peptide containing the LPLTG sorting signal (FELPLTGGSG). (C) Structure of ^{Cd}SrtA showing H160, C222, and R231 active site residues. The "Lid" is highlighted in red (residues P77 to S89).

In this study, we developed an HPLC-based assay to measure for the first time the kinetics of catalysis, and we have used the assay to identify a new ^{Cd}SrtA variant that has improved bioconjugation activity. In particular, we show that (i) the bioconjugation reaction rate is limited by the formation of an enzyme−acyl intermediate with the LPLTG sorting signal, (ii) the enzyme preferentially recognizes nonpolar amino acids at the X position within the sorting signal, (iii) unlike ^{Sa}SrtA, ^{Cd}SrtA exhibits minimal proteolytic activity, (iv) amino acid substitutions introduced into the lid accelerate catalysis by facilitating enzyme-acyl intermediate formation, and (v) complete removal of the lid further activates the enzyme. These results increase the *in vitro* utility of ^{Cd}SrtA as a bioconjugation tool to modify proteins and provides new insight into the enzymatic reaction that underpins the construction of surface pili in Gram-positive bacteria.

RESULTS AND DISCUSSION

Kinetics of Lysine−Isopeptide Bond Formation. Previously, we monitored the lysine−isopeptide bond forming activity of ^{Cd}SrtA^{3M} using SDS-PAGE.[22] However, the kinetics of this process could not be accurately determined because the reactants and products were difficult to separate and quantify. To overcome this problem, we developed an HPLC-based transpeptidation assay that monitors the ability of ^{Cd}SrtA^{3M} to join the N-terminal domain from SpaA (^{N}SpaA, residues E30-S195 of SpaA) to a peptide containing its C-terminal sorting signal (FELPLTGGSG, hereafter called LPLTG peptide). This reaction represents an isolated ligation event in the polymerization reaction by producing a ^{N}SpaA-LPLT product in which the K190 side chain in ^{N}SpaA is joined via an isopeptide bond to the threonine carbonyl group in the FELPLT peptide (Figure 1B). The reactants and products are readily separated by reverse-phase HPLC (Figure 2A). Moreover, this procedure enables facile monitoring of the time-dependent conversion of the protein substrate (^{N}SpaA) into the cross-linked protein−peptide product (^{N}SpaA-LPLT) (Figure 2B). The identity of the product and the location of its isopeptide linkage were previously confirmed by LC-MS/MS.[22,28] Initially, for each substrate (^{N}SpaA and the LPLTG peptide) the dependence of the reaction velocity on substrate concentration was determined (Figure 2C). This analysis reveals that ^{N}SpaA and peptide substrates do not saturate the enzyme even when they are present at concentrations of 500 μM (Figure 2C, insert) and 4 mM (Figure 2C, main), respectively. Because using higher concentrations of each substrate is not practical, we determined apparent steady-state parameters using subsaturating amounts of each substrate (representative data is shown in Figure 2D). Two sets of Michaelis−Menten parameters were obtained. Either the concentration of ^{N}SpaA was varied from 62.5 to 500 μM with the amount of LPLTG peptide held constant at 1 mM, or the concentration of the LPLTG peptide was varied from 0.5 to 4 mM, while the concentration of ^{N}SpaA was held fixed at 500 μM. These measurements revealed that ^{Cd}SrtA^{3M} catalyzes isopeptide

https://dx.doi.org/10.1021/acs.bioconjchem.0c00163
Bioconjugate Chem. XXXX, XXX, XXX−XXX

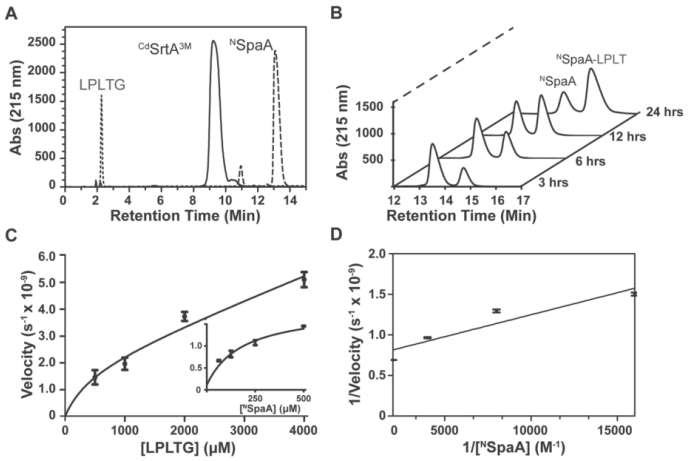


Figure 2. ^{Cd}SrtA transpeptidation assay: (A) Superimposed reversed-phase HPLC traces showing the separation of the sorting signal peptide (LPLTG, short dashes), ^{Cd}SrtA^{3M} (solid line), and ^{N}SpaA (long dashes). (B) Representative HPLC traces that track the progress of the reaction. Peaks corresponding to ^{N}SpaA and ^{N}SpaA-LPLT are shown. The reactions were sampled at 3, 6, 12, and 24 h, and contained 100 μM of enzyme, 200 μM of ^{N}SpaA, and 1 mM of LPLTG-peptide. (C) Plots showing the measured velocity versus substrate concentration for the LPLTG peptide (the sorting signal) and ^{N}SpaA (inset). A concentration range of 500 μM to 4 mM, and 62.5 μM to 500 μM were used for the LPLTG peptide and ^{N}SpaA, respectively. Initial velocities were determined after 3 h, as described in the Methods section. (D) Lineweaver–Burk graph graphing showing kinetics data for ^{Cd}SrtA^{3M}. The k_{cat} and K_M values were determined from a linear fit of this data.

Table 1. Kinetics of ^{Cd}SrtA Catalyzed Lysine–Isopeptide Formation[a]

	$k_{cat} \times 10^{-5}$ (s^{-1})[b]	$^{N}K_M \times 10^{-4}$ (M)	$^{S}K_M \times 10^{-4}$ (M)	$k_{cat}/^{N}K_M$ (s^{-1} M^{-1})
^{Cd}SrtA	n.d.[c]	n.d.	n.d.	n.d.
^{Cd}SrtA^{3M}	5.6 ± 0.8	0.7 ± 0.1	20 ± 10	0.7 ± 0.1
H160A	3.1 ± 0.4	0.43 ± 0.05	—	0.72 ± 0.08
C222A	n.d.	n.d.	—	n.d.
R231A	n.d.	n.d.	—	n.d.
^{Cd}SrtA$^{\Delta}$	40 ± 0.1	1.6 ± 0.4	16 ± 3	2.5 ± 0.6
H160A	2.5 ± 0.6	0.70 ± 0.02	—	0.36 ± 0.09
C222A	n.d.	n.d.	—	n.d.
R231A	n.d.	n.d.	—	n.d.
^{Sa}SrtA[d]	1600 ± 100	1.8 ± 0.1	73.3 ± 10.1[e]	86 ± 5

[a]All kinetics are approximations as saturating concentrations were not able to be measured. [b]Transpeptidation steady-state kinetic parameters for ^{Cd}SrtA were determine by the monitoring rate at which the enzyme ligated an FELPLTGGSG peptide to the ^{N}SpaA domain via a lysine–isopeptide bond. [c]n.d., not determined because an insufficient amount of product was detectable. [d]Transpeptidation steady-state kinetic parameters for ^{Sa}SrtA were determined by monitoring the rate at which the enzyme ligated GGG and FELPLTGGSG peptides via a backbone peptide bond. Reported values are the average from three measurements, and the error is the standard deviation. [e]Values are reported from Frankel et al. (2005) and measures reactions between a Abz-LPETG-Dap(Dnp)-NH_2 and pentaglycine.[29] ^{N}Refers to transpeptidation kinetics measure when ^{N}SpaA is varied and FELPLTGGSG concentration is held constant. ^{S}Refers to when FELPLTGGSG peptide is varied and ^{N}SpaA is held constant.

bond formation with apparent K_M values for ^{N}SpaA ($^{N}K_M$) and the LPLTG peptide ($^{S}K_M$) of 70 ± 10 μM and 2 ± 1 mM, respectively (Table 1). Similar turnover numbers were measured in each experiment, with a maximal value of 26 ± 10 ($\times 10^{-5}$ s) (obtained when ^{N}SpaA is held constant at 500 μM and the LPLTG peptide is varied).

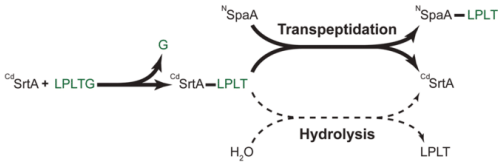
Formation of the ^{Cd}SrtA-LPLT Thioacyl Intermediate Is a Rate Limiting Step in Catalysis and Can Be Accelerated by Completely Removing the Lid. ^{Cd}SrtA^{3M} is the most active form of ^{Cd}SrtA thus far discovered and contains three substitutions in the inhibitory lid structure (D81G, W83G, N85A).[22] We reasoned that an enzyme variant with the entire lid deleted might exhibit even higher transpeptidation activity by completely unmasking the active site. Inspection of the crystal structure ^{Cd}SrtA suggests that lid residues I78 to A88 can be deleted without disrupting its tertiary structure, as the remaining P77 and S89 amino acids are positioned adjacent to one another in three-dimensional space.[28] Indeed, the measured steady-state kinetic parameters for a lid-deletion ^{Cd}SrtA$^{\Delta}$ (residues N37–Q257 of ^{Cd}SrtA in which the amino acids I78 to A88 are deleted) reveal that it is more active than ^{Cd}SrtA^{3M}; there is a 7-fold improvement in the apparent k_{cat}, while the ^{N}K$_M$ and ^{S}K$_M$ values are only modestly affected (Table 1).

To address why ^{Cd}SrtA$^{\Delta}$ is catalytically more active than ^{Cd}SrtA^{3M}, we investigated how lid removal affected catalysis. By analogy to the prototypical ^{Sa}SrtA enzyme, ^{Cd}SrtA presumably catalyzes lysine–isopeptide bond formation via a two-step

https://dx.doi.org/10.1021/acs.bioconjchem.0c00163
Bioconjugate Chem. XXXX, XXX, XXX–XXX

process (Scheme 1).[23,29−31] In the transpeptidation mechanism, ^{Cd}SrtA's C222 thiol presumably functions as a

Scheme 1. Schematic Showing the Overall Mechanism of ^{Cd}SrtA-Catalyzed Isopeptide Formation (Top) and a Potential Hydrolytic Side Reaction (Bottom)



nucleophile to attack the threonine carbonyl carbon in the LPLTG sorting signal to form a ^{Cd}SrtA-LPLT thioacyl intermediate. Then, the enzyme recognizes the K190 side chain amine group located within ^{N}SpaA, which resolves the thioacyl intermediate to form the lysine−isopeptide linked ^{N}SpaA-LPLT product. In ^{Sa}SrtA, a side reaction also occurs in which a water molecule attacks the thioacyl intermediate instead of a primary amine, resulting in the hydrolysis of the intermediate to release the LPLT peptide.[29] In this side reaction, the enzyme effectively functions as a protease, cleaving the LPLTG peptide substrate at the peptide bond that joins the threonine and glycine residues. ^{Cd}SrtA has also been shown to proteolyze its LPLTG peptide substrate, but the kinetics and extent of proteolysis have not been rigorously measured.[22,28]

To determine if differences in the rate of the hydrolytic side reaction cause the ^{Cd}SrtA$^{\Delta}$ and ^{Cd}SrtA^{3M} enzymes to produce differing amounts of transpeptidation product, we measured the ability of each enzyme to proteolyze the LPLTG peptide substrate using reversed-phase HPLC. In these reactions, only the enzyme and LPLTG peptide are present. Interestingly, even though ^{Cd}SrtA$^{\Delta}$ and ^{Cd}SrtA^{3M} catalyze transpeptidation at 25 °C, the rate at which the hydrolytic side reaction occurs at this temperature is very slow for both enzymes, with less than 5% of the LPLTG peptide substrate consumed by the enzyme (Figure 3A, left). In fact, because this assay employed excess amounts of peptide relative to enzyme (50 μM enzyme and 500 μM peptide), much of the observed peptide consumption can be attributed to formation of the thioacyl intermediate and not to repeated rounds of proteolysis. This finding is in marked contrast to the archetypal ^{Sa}SrtA enzyme, which proteolyzes more than 60% of its sorting signal peptide substrate when identical reaction conditions are used (Figure 3A, right). Thus, we conclude that for both the ^{Cd}SrtA$^{\Delta}$ and ^{Cd}SrtA^{3M} enzymes the transpeptidation pathway is dominant and the hydrolytic side reaction occurs only to a minor extent.

We wondered whether the superior transpeptidation activity of ^{Cd}SrtA$^{\Delta}$ originated from its ability to form the enzyme−LPLT thioacyl intermediate more rapidly than ^{Cd}SrtA^{3M} (Scheme 1). For both the ^{Cd}SrtA$^{\Delta}$ and ^{Cd}SrtA^{3M} enzymes, this intermediate is readily observable in LC-MS mass spectra when they are incubated with the LPLTG peptide (Figure 3B). This finding is consistent with the low proteolytic activity of each enzyme and suggests that the two steps of catalysis are independent of one another—each enzyme can form and

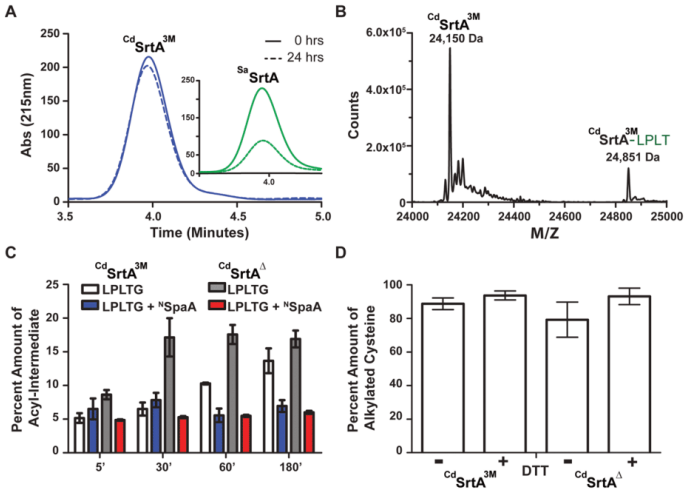


Figure 3. Characterization of ^{Cd}SrtA^{3M} and ^{Cd}SrtA$^{\Delta}$. (A) Representative reverse phase HPLC trace showing the change in the concentration of LPLTG peptide due to proteolysis by ^{Cd}SrtA^{3M} (blue) and ^{Sa}SrtA^{WT} (green) at 25 °C after 0 (solid lines) and 24 h (dashed lines). (B) Mass deconvolution of LC-MS data of ^{Cd}SrtA^{3M}. The acyl-intermediate at 24,851 Da is approximately 700 Da higher than where the enzyme is at 24,150 Da. (C) Comparison of the amounts of acyl-intermediate in the presence and absence of ^{N}SpaA measured over a period of three hours. ^{Cd}SrtA^{3M} without ^{N}SpaA (white) in comparison to ^{Cd}SrtA$^{\Delta}$ without ^{N}SpaA (gray) shows the faster formation of the ^{Cd}SrtA$^{\Delta}$ acyl-intermediate. In contrast, both enzymes form these intermediates at similar rates when ^{N}SpaA is present (^{Cd}SrtA^{3M} (blue) and ^{Cd}SrtA$^{\Delta}$ (red)). (D) Data showing the amount of reduced active site C222 thiol in the ^{Cd}SrtA^{3M} and ^{Cd}SrtA$^{\Delta}$ enzymes. Freshly produced samples of each enzyme were either treated with an excess amount of DTT or a buffer control for one hour. The proteins were then digested with trypsin, reacted with iodoacetamide, and the extent of cysteine alkylation determined by mass spectrometry. The experiments were performed in triplicate.

https://dx.doi.org/10.1021/acs.bioconjchem.0c00163
Bioconjugate Chem. XXXX, XXX, XXX−XXX

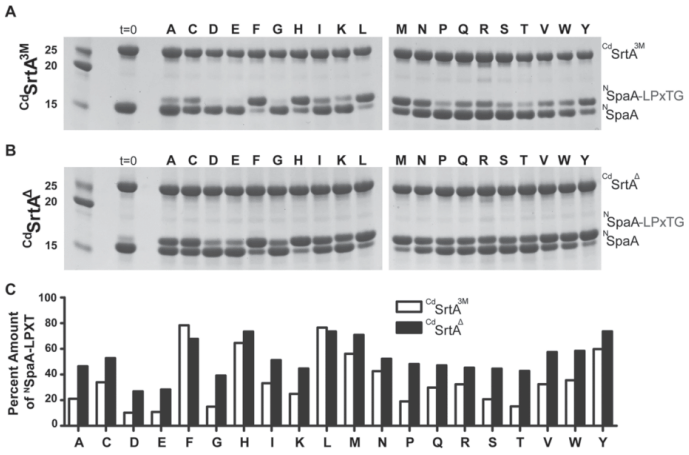


Figure 4. Role of the X residue in transpeptidation. (A) SDS-PAGE analysis of product formation when ^{Cd}SrtA^{3M} was reacted with ^{N}SpaA, and a series of LPXTG peptides in which the identity of the amino acid at the X position was systematically varied. Reactions (200 μM enzyme, 200 μM ^{N}SpaA, 5 mM DTT, and 1 mM Peptide) were measured after 24 h. The most reactive sorting signal peptides contained nonpolar X residues, while those containing polar residues were less reactive. (B) As in panel (A), but ^{Cd}SrtA$^{\Delta}$ was used instead of ^{Cd}SrtA^{3M}. Similar trends in activity are observed. (C) Histogram plot showing the amount of product created for each peptide in the library. The fraction of ^{N}SpaA converted to ^{N}SpaA-LPXT is shown. The data was obtained by analyzing the SDS-PAGE data (panels A and B) using the program ImageJ.

maintain the enzyme−acyl intermediate in the absence of the ^{N}SpaA nucleophile. Moreover, it is compatible with our previously published cellular studies of class C sortases in which long-lived and stable acyl−enzyme intermediates were established to be important for catalysis.[32] To determine if ^{Cd}SrtA$^{\Delta}$ and ^{Cd}SrtA^{3M} differed in their ability to perform the first step of catalysis, we tracked thioacyl intermediate formation at various times after mixing the enzymes with the LPLTG peptide (25 μM and 1 mM of enzyme and LPLTG peptide, respectively). The intermediate formation was followed over a 3 h period; the same duration was used to measure transpeptidation activity. This analysis revealed that ^{Cd}SrtA$^{\Delta}$ forms the thioacyl intermediate more rapidly than ^{Cd}SrtA^{3M}, with ∼17 ± 2% of ^{Cd}SrtA$^{\Delta}$ joined to the peptide after 30 min, while it takes up to 3 h for ^{Cd}SrtA^{3M} to form similar levels of this reaction intermediate (Figure 3C). Interestingly, when similar time course experiments are performed in the presence of both ^{N}SpaA and LPLTG substrates, significantly lower amounts of thioacyl intermediate are observed for each type of enzyme (∼5% for both ^{Cd}SrtA$^{\Delta}$ and ^{Cd}SrtA^{3M}). As the reaction conditions are identical to those used to measure transpeptidation activity, this data suggests that the formation of the acyl-intermediate is rate-limiting. Thus, we conclude that improved transpeptidation activity of ^{Cd}SrtA$^{\Delta}$ results from its ability to form the acyl-intermediate at a faster rate than ^{Cd}SrtA^{3M}.

The Identity of the "X" Residue within the LPXTG Sorting Signal Affects the Rate of Transpeptidation. Structural and computational studies of class A and B sortases bound to their respective sorting signals have revealed that they do not recognize the side chain of the X residue within their respective LPXTG-type sorting signal substrates because it projects away from the enzyme into the solvent.[23,33,34] The structures are consistent with detailed substrate specificity analyses of the class A ^{Sa}SrtA enzyme, which revealed that LPXTG sorting signals containing any amino acid at the X position can be used as substrates.[35] However, ^{Cd}SrtA and other class C sortases are unique, because they contain a lid structure whose proximity to the active site could affect recognition of the LPXTG sorting signal (Figure 1C).[33] Indeed, our prior studies of ^{Cd}SrtA^{3M} revealed that a leucine to alanine substitution at the X position of the LPLTG sorting signal slowed transpeptidation.[22] To investigate this issue in greater detail, the ^{Cd}SrtA$^{\Delta}$ and ^{Cd}SrtA^{3M} enzymes were tested for their ability to utilize as transpeptidation substrates 20 distinct LPXTG peptides in which the X position was varied. For these studies, the ^{Cd}SrtA$^{\Delta}$ and ^{Cd}SrtA^{3M} enzymes were incubated with ^{N}SpaA and each member of the peptide library, and the amount of cross-linked product was then determined by SDS-PAGE (Figure 4A,B). Significant variation in reactivity is observed for the different library members. However, in nearly all cases, ^{Cd}SrtA$^{\Delta}$ is more active than ^{Cd}SrtA^{3M}, consistent with the steady-state kinetic measurements that employed the LPLTG peptide (Table 1). Interestingly, both enzymes exhibit similar sequence preferences. In particular, they preferentially use sorting signals containing phenylalanine, histidine, methionine, tyrosine, and leucine at the X position, while their least reactive substrates contain negatively charged side chains at this site. In all cases, peptides containing leucine at the X position are very reactive, explaining why this amino acid is present in the native LPLTG substrate present in SpaA. The X position-dependent activity of ^{Cd}SrtA is distinct from what has been observed for ^{Sa}SrtA, as similar peptide library studies have shown that after 24 h exposure all peptides in the library are processed to a similar extent by ^{Sa}SrtA.[35] The molecular basis underlying the observed variation in peptide reactivity is not known, but it is not caused by the presence of the lid as similar trends in activity are observed for ^{Cd}SrtA^{3M} and ^{Cd}SrtA$^{\Delta}$.

Cysteine and Arginine Active Site Residues Are Required for Catalysis. Based on sequence homology with

https://dx.doi.org/10.1021/acs.bioconjchem.0c00163
Bioconjugate Chem. XXXX, XXX, XXX−XXX

the well-studied ^{Sa}SrtA sortase, three conserved residues in ^{Cd}SrtA are presumably required for catalysis: H160, C222, and R231 (Figure 1B).[36,37] However, their role in catalyzing lysine−isopeptide bond transpeptidation *in vitro* has not been determined. We therefore used the HPLC-assay to measure the transpeptidation activities of ^{Cd}SrtA^{3M} and ^{Cd}SrtA$^{\Delta}$ enzymes containing alanine substitutions at these sites. For both enzyme variants, C222A and R231A substitutions completely abrogate transpeptidation activity, demonstrating that they have critical functions in catalysis (Table 1). However, ^{Cd}SrtA^{3M} and ^{Cd}SrtA$^{\Delta}$ enzymes harboring H160A substitutions retain some enzyme activity, exhibiting similar turnover numbers that are reduced by 44% and 94% as compared to their native forms, respectively. In the well-studied ^{Sa}SrtA transpeptidation reaction the analogous histidine, cysteine, and arginine residues are essential for catalysis *in vitro* (R197, C184, and H120 in the ^{Sa}SrtA sequence).[36] The cysteine thiol functions as a nucleophile, while the arginine guanidinium group (R197 in ^{Sa}SrtA) has been proposed to facilitate catalysis by stabilizing oxyanion tetrahedral intermediates.[23] Our finding that the C222A and R231A sortase variants are completely inactive is consistent with these residues having similar functions.[28] It is also compatible with pH dependence of the transpeptidation reaction, which occurs most rapidly at pH values between 7.5 and 8. We were surprised that ^{Cd}SrtA containing a H160A substitution retained some activity, since the analogous alteration in ^{Sa}SrtA disrupts transpeptidation *in vitro*.[36] In ^{Sa}SrtA, the histidine side chain has been proposed to function as a general acid and base, protonating the amine of the leaving group glycine residue as the threonine−glycine peptide bond in the LPLTG sorting signal is broken, and facilitating the last step of catalysis by deprotonating the incoming amine nucleophile that resolves the thioacyl intermediate to produce the final transpeptidation product (Scheme 1).[29,38] Whole cell studies have shown that ^{Cd}SrtA containing H160A are incapable of assembling surface pili.[28] The residual activity observed in the H160A enzyme suggests that residues in addition to H160 in ^{Cd}SrtA may facilitate these steps *in vitro*, albeit less efficiently.

We wondered whether the superior activity of ^{Cd}SrtA$^{\Delta}$ relative to ^{Cd}SrtA^{3M} could be attributed to differences in the oxidation state of the C222 sulfhydryl group that is influenced by the presence of the lid. This is because the sulfhydryl group can in principle become oxidized to unreactive disulfide, sulfenic, sulfinic, and sulfonic forms that are non-nucleophilic.[39] No significant differences in disulfide formation were observed for the two proteins in nonreducing SDS-PAGE and MALDI experiments. To investigate whether they formed more oxidized states, freshly purified ^{Cd}SrtA$^{\Delta}$ and ^{Cd}SrtA^{3M} enzymes were incubated for one hour in buffer A (50 mM Tris-HCl, 300 mM NaCl, at a pH of 8.0) with, and without, a reducing agent (5 mM DTT). The amount of C222 (the protein's only cysteine) present in the active thiol form was then determined by adding iodoacetamide, digesting with trypsin, and analyzing by mass spectrometry. The percent of reduced thiol was calculated by measuring the amount of C222 in each enzyme that was derivatized with carboxyamidomethyl as compared to the total amount of C222 in its various oxidation states. In the absence of reducing agent, 89 ± 4 and 79 ± 10% of ^{Cd}SrtA^{3M} and ^{Cd}SrtA$^{\Delta}$ contain a reactive thiol, respectively. Moreover, only small increases in the amount of reactive thiol are observed when the DTT reducing agent is present; under these conditions, 94 ± 3% and 93 ± 5% ^{Cd}SrtA^{3M} and ^{Cd}SrtA$^{\Delta}$ are reactive, respectively (Figure 3D). Thus, the C222 sulfhydryl in freshly purified ^{Cd}SrtA^{3M} and ^{Cd}SrtA$^{\Delta}$ primarily exists in a reduced, transpeptidation competent state. A caveat is that ionization efficiencies may differ for alkylated and nonalkylated forms of a cysteine-containing peptide. Nevertheless, the similarity between the +/−DTT values implies that C222 is primarily reduced. It should be noted that we have found that C222 can become oxidized and unreactive if the enzymes are stored for more than several weeks, and their activity can be restored by incubating them with DTT. Therefore, including DTT in the labeling reactions is recommended as a precaution.

Improved Protein Lysine Labeling Using ^{Cd}SrtA$^{\Delta}$. Having defined substrate and reaction conditions that are optimal for activity, we directly compared the peptide labeling efficiencies of ^{Cd}SrtA^{3M} with the newer ^{Cd}SrtA$^{\Delta}$ enzyme. Consistent with our steady-state kinetic analyses, a temporal analysis using identical conditions clearly shows that ^{Cd}SrtA$^{\Delta}$ (blue trace) produces more cross-linked product than

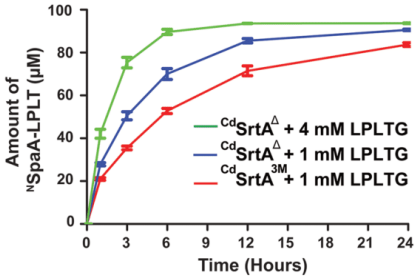


Figure 5. Labeling activity of ^{Cd}SrtA^{3M} and ^{Cd}SrtA$^{\Delta}$. Reactions containing each enzyme were performed using identical conditions: 100 μM enzyme, 100 μM ^{N}SpaA, and either 1 or 4 mM LPLTG peptide. ^{Cd}SrtA$^{\Delta}$ exhibits superior labeling activity at all measured time points. The experiments were performed in triplicate.

^{Cd}SrtA^{3M} (red trace) (Figure 5) (100 μM enzyme, 100 μM ^{N}SpaA and 1 mM LPLTG peptide at 25 °C). In particular, when ^{Cd}SrtA$^{\Delta}$ is used, ∼90% of ^{N}SpaA is modified with peptide within 12 h, a higher amount than is achieved with ^{Cd}SrtA^{3M} even after 24 h. Notably, even faster labeling can be achieved using ^{Cd}SrtA$^{\Delta}$ by increasing the concentration of LPLTG peptide in the reaction to 4 mM, which enables 90 ± 2% of ^{N}SpaA to be modified within 6 h (Figure 5, green trace). This is finding is compatible with the relatively high $^{S}K_{M}$ (2 ± 1 mM), which makes it challenging to saturate the enzyme with sorting signal substrate (Table 1). As ^{Cd}SrtA^{3M} has previously been shown to be capable of labeling ^{N}SpaA with a peptide fluorophore, ^{Cd}SrtA$^{\Delta}$ can be expected to more active in these labeling reactions as well.[22] Similar comparisons were performed at 37 °C instead of 25 °C. ^{Cd}SrtA$^{\Delta}$ becomes less active at elevated temperatures because it is less thermostable than ^{Cd}SrtA^{3M}. However, both the rate and yield of product formed by ^{Cd}SrtA$^{\Delta}$ at 25 °C are superior to what ^{Cd}SrtA^{3M} produces at 37 °C. Thus, ^{Cd}SrtA$^{\Delta}$ should be used for labeling reactions that are performed at room temperature.

Comparison with the Prototypical ^{Sa}SrtA Sortase. ^{Sa}SrtA is the best-studied member of the sortase superfamily

https://dx.doi.org/10.1021/acs.bioconjchem.0c00163
Bioconjugate Chem. XXXX, XXX, XXX−XXX

and is routinely used to modify the N- and C-termini of proteins.[13,23] ^{Cd}SrtA and ^{Sa}SrtA catalyze mechanistically related transpeptidation reactions, but differ in the type of nucleophile that they use, i.e., a lysine ε-amine within ^{N}SpaA and the N-terminal amine group within an oligoglycine peptide, respectively. In order to directly compare their activities, an established HPLC assay was used to measure the rate at which ^{Sa}SrtA joins the LPLTG peptide previously used in studies of ^{Cd}SrtA to its triglycine substrate (Gly_3), a reaction that creates a FELPTGGGG peptide product (Table 1).[29,40] The steady-state kinetic values measured from this analysis are generally consistent with published values that used a much shorter fluorogenic labeled peptide.[29] Our direct comparison reveals that ^{Sa}SrtA forms backbone peptide bonds ~40-times faster than ^{Cd}SrtA$^{\Delta}$ creates lysine−isopeptide bonds (Table 1). Interestingly, both enzymes exhibit generally similar K_M values for their substrates, with each exhibiting millimolar K_M values for the sorting signal substrate (^{Cd}SrtA$^{\Delta}$, 1.6 ± 0.3 mM; ^{Sa}SrtA, 7.33 ± 1.01 mM), and approximately ~10-fold lower K_M values for their respective amine nucleophiles (^{Cd}SrtA$^{\Delta}$, 160 ± 30 μM; ^{Sa}SrtA, 180 ± 10 μM).[29] In addition, based on prior studies by the McCafferty group and the studies presented here (Figure 3), for both ^{Sa}SrtA and ^{Cd}SrtA the first step of catalysis is rate limiting—the formation of the thioacyl intermediate (Scheme 1).[29,40] It is unclear why this step is so slow in both enzymes. However, it has been shown that ^{Sa}SrtA primarily exists in an inactive form in which only a small fraction (ca. 0.06%) of enzyme contains catalytically capable Cys184 thiolate and His120 imidazolium forms that are capable of reacting with the sorting signal to form the thioacyl intermediate.[30,41,42] Whether the active site of ^{Cd}SrtA also primarily exists in a catalytically dormant state remains to be determined. Startlingly, we found that ^{Cd}SrtA is quite inefficient at proteolyzing the LPLTG substrate, which is in stark contrast to previously reported studies of ^{Sa}SrtA that have shown that it catalyzes this reaction with a k_{cat} of 0.28 ± 0.02 s^{-1} (Figure 3A).[29] Presumably this difference originates from distinct active site features that enable ^{Sa}SrtA to use water as a nucleophile much more efficiently than ^{Cd}SrtA (Scheme 1). This idea is substantiated by our observation that the thioacyl intermediate in the ^{Cd}SrtA reaction forms to an appreciable extent in the absence of the nucleophile (Figure 3C). The reduced proteolytic activity of ^{Cd}SrtA is presumably advantageous, limiting the release of partially assembled pili from the bacterial cell surface. Similar to ^{Cd}SrtA, under certain conditions ^{Sa}SrtA can catalyze formation of a lysine isopeptide bond.[43−45] A thorough study by Dasgupta and colleagues demonstrated that ^{Sa}SrtA can join the sorting signal and lysine−containing peptides together via an isopeptide linkage.[44] However, the reaction was inefficient and exhibited poor substrate specificity; even after 12 h, the ligation reaction was incomplete and a range of distinct isopeptide-linked products were generated. Moreover, during these reactions a significant amount of the sorting signal was proteolyzed, suggesting that ^{Sa}SrtA does not discriminate between lysine and water nucleophiles. This is in contrast to ^{Cd}SrtA$^{\Delta}$ and ^{Cd}SrtA^{3M}, which produce lysine isopeptide-linked products at high yields with only limited proteolysis of the sorting signal substrate. A current limitation of the ^{Cd}SrtA-labeling system compared to ^{Sa}SrtA is that it requires the target protein to be expressed as a fusion to ^{N}SpaA to achieve maximum labeling. However, because ^{N}SpaA is a small protein its presence is unlikely to affect the function of the target protein.

In conclusion, we have characterized the *in vitro* kinetics and mechanism of lysine−isopeptide bond forming activity of ^{Cd}SrtA, and discovered ^{Cd}SrtA$^{\Delta}$ which is ~7-fold more active than previously reported enzyme variants. This bioconjugation activity is beginning to rival that of microbial transglutaminases from *Streptomyces mobaraensis*, which can be used to join biomolecules isopeptide linkages between glutamine and lysine side chains.[46] Although promising, these enzymes have not gained wide usage in site-specific protein labeling, presumably because of their penchant to catalyze spurious ligations.[46−49] In contrast, ^{Cd}SrtA exhibits a high level of specificity for its substrates, ligating peptides containing a LPXTG sequence to a specific lysine residue within the ^{N}SpaA domain. The molecular basis of specificity for the K190 side chain remains unknown, but presumably originates from protein−protein interactions between the ^{N}SpaA and ^{Cd}SrtA-LPLT thioacyl intermediate that function to properly position the nucleophile for catalysis. It may also arise from ^{N}SpaA structural features that provide an environment for K190 that lower its pK_a. Our mechanistic analysis also provides insight into the function of the lid, which is widely conserved in sortase enzymes that assemble pili. We show that *in vitro* its presence primarily affects the rate of thioacyl intermediate formation, and that it does not have a significant role in recognizing ^{N}SpaA or the X residue within the LP(X)TG sorting signal. Further improvements in ^{Cd}SrtA$^{\Delta}$-mediated labeling activity may also be possible, as we estimate that current versions of the enzyme catalyze *in vitro* transpeptidations ~10^2−10^3 times more slowly than the native enzyme when it is located on the cell surface. A number of approaches could be used to improve the kinetics of protein labeling, including discovering variants of ^{Cd}SrtA$^{\Delta}$ that have superior thermostability and employing strategies that increase the effective substrate concentration by either immobilizing the reactants or fusing the enzyme to its substrates.[50−52] Finally, further improvements may also be achieved by obtaining a greater understanding of the process of substrate recognition and catalysis.

METHODS

Protein Reagents. Purified ^{Cd}SrtA^{3M} pilin sortase (residues N37−Q257 of SrtA from *C. diphtheriae*) and enzyme variants were expressed and purified as described previously.[22] Briefly, proteins were expressed from a pE-SUMO (Lifesensors) plasmid in *E. coli* BL21 (DE3) cells. The cells were grown up in LB supplemented with 500 μg/mL of Kanamycin at 37 °C until they reached an OD_{600} of ~0.6. The cells were induced with 1 M IPTG and then left to express at 17 °C for 8 to 12 h. After, the cells were removed from the incubator and pelleted at 8670*g* for 15 min. The pellets were then dissolved in a buffer of 50 mM Tris-HCl, 300 mM NaCl, at a pH of 8.0 (lysis buffer). Subsequently, the cells were lysed using high-pressure emulsification and then fractionated via centrifugation at 22,720*g* for 50 min. Afterward, the cell lysate was purified via IMAC-Co^{2+} purification. Proteins were eluted from the resin using a lysis buffer supplemented with 200 mM Imidazole. The His_{6x}-SUMO tags were removed via treatment by His_{6x}-Ulp1 protease at 1 mg/mL and subsequent purification by IMAC-Co^{2+}. Afterward, the protein was purified by size exclusion chromatography via the AKTA Pure (GE) and with Superdex 75pg resin. Protein purity was confirmed by SDS-PAGE. pE-SUMO expression plasmids encoding ^{Cd}SrtA^{3M} variants were

https://dx.doi.org/10.1021/acs.bioconjchem.0c00163
Bioconjugate Chem. XXXX, XXX, XXX−XXX

created using standard molecular biology methods and confirmed by nucleotide sequencing. ^{N}SpaA (residues E30 to S195) and ^{Sa}SrtA (*S. aureus* Sortase A, residues Q60−K206) were purified as described previously.[28,53] All purified enzymes were stored at −20 °C in buffer A (50 mM Tris-HCl, 300 mM NaCl, at a pH of 8.0) supplemented with 40% glycerol. The FELPLTGGSG peptide (LPLTG peptide) used in the transpeptidation and hydrolysis assays was synthesized by Peptide 2.0.

Transpeptidation Assays. An HPLC-based assay was developed to quantify the kinetics of ^{Cd}SrtA catalyzed lysine−isopeptide bond formation. In this assay, the ^{N}SpaA protein containing the reactive lysine is ligated to a FELPLTGGSG peptide (LPLTG, where the underlined residues correspond to the sorting signal) by the pilin sortase, followed by quantification using a HPLC C4 column. Reactions were performed in 100 μL volumes and contained 25 μM of pilin sortase (either ^{Cd}SrtA^{3M} or ^{Cd}SrtA$^{\Delta}$), DTT (5 mM), either constant or variable amounts of LPLTG peptide (1 mM or 0.5 to 4 mM), and either constant or variable amounts of ^{N}SpaA (500 μM, 62.5 to 500 μM). All components were dissolved in buffer A. At these substrate concentrations, an estimated K_M for ^{N}SpaA and LPLTG was determined. Reactions were initiated by adding the pilin sortase from a 2 mM stock solution, incubated for 3 h at 25 °C and then flash-frozen with liquid N_2 and stored at −20 °C. The reactions were analyzed using a Phenomenex C4 column (5 μm, 4.6 × 150 mm) and with an initial dwell time of 3 min at 36% CH_3CN/0.1% TFA followed by a linear gradient from 36% to 46% CH_3CN/0.1% TFA for 10 min at 1 mL/min was applied. The column was subsequently flushed with high concentrations of CH_3CN/0.1% at 1 mL/min. ^{N}SpaA containing peaks were detected at 215 nm and the amount of substrate converted to product was calculated by integrating the area under the HPLC traces. The identity of each peak in the HPLC chromatogram was confirmed via MALDI-TOF MS. The activity of ^{Cd}SrtA was compared to ^{Sa}SrtA, which catalyzes a transpeptidation reaction that forms a backbone−backbone peptide bond between LPXTG and oligoglycine peptides. ^{Sa}SrtA transpeptidation activity was measured as described previously.[40] These reactions were performed in an identical manner to the ^{Cd}SrtA reaction described above, except that they contained 25 μM ^{Sa}SrtA instead of ^{Cd}SrtA and were supplemented with 10 mM calcium in Buffer A, and triglycine (Gly_3 peptide) (62.5 μM to 1 mM) instead of ^{N}SpaA. Reactions were quenched 15 min after mixing by adding an equal amount of 1 N hydrochloric acid (HCl). The reaction products were separated by HPLC using a Phenomenex C18 column (10 μm, 4.6 × 150 mm) and a linear gradient from 26% to 30% CH_3CN/0.1% TFA over 8 min (1 mL/min). All HPLC experiments were performed on an Agilent 1100 HPLC. For both the ^{Sa}SrtA and ^{Cd}SrtA reactions kinetic parameters were obtained by fitting the data with Sigmaplot 12.0.

Two types of transpeptidation assays were used to investigate sorting signal specificity of the ^{Cd}SrtA^{3M} and ^{Cd}SrtA$^{\Delta}$ enzymes for the X residue within the LPXTG sorting signal. A total of 20 peptides were tested in which the X residues in the FELPXTGGSG was varied (Peptide 2.0). Reactions were performed in buffer A with a total volume of 35 μL. The reactions contained: either ^{Cd}SrtA^{3M} or ^{Cd}SrtA$^{\Delta}$ (200 μM), DTT (5 mM), ^{N}SpaA (200 μM), and one of the FELPXTGGSG peptides (1 mM, Peptide 2.0). Transpeptidation reactions performed at 25 °C for 24 h, and then quenched by flash freezing in liquid N_2 and stored at −20 °C. Five microliters of each reaction was diluted 4 times in SDS loading buffer and separated using a 12% SDS-PAGE gel, then visualized by Coomassie staining. The resulting bands were analyzed with ImageJ with the zero hour time point used as a control for activity.

Hydrolysis and Cysteine Oxidation Measurements. The hydrolytic activity of ^{Sa}SrtA and ^{Cd}SrtA variants was determined using an HPLC-based assay that monitors the ability of each enzyme to cleave the LPLTG peptide between the threonine and glycine residues. Reactions were performed in buffer A supplemented with 10 mM calcium and contained a total volume of 100 μL: sortase (50 μM), LPLTG (500 μM), and DTT (5 mM). Reactions were incubated at 25 °C for 24 h, and then quenched by adding an equal volume of 1 N HCl. Reaction products were separated on a Waters C18 Column (10 μm, 4.6 × 150 mm) using a linear gradient from 26% to 30% CH_3CN/0.1% TFA over 8 min at 1 mL/min. The reaction was monitored at 215 nm and the identity of each peak in the HPLC chromatogram was confirmed via LC-MS.

The oxidation status of the active site cysteine by monitoring susceptibility to iodoacetamide alkylation. ^{Cd}SrtA^{3M} and ^{Cd}SrtA$^{\Delta}$ were purified and stored in buffer A. Enzymes were then either treated to a final concentration of 5 mM of DTT or a buffer control for 1 h before being frozen at −20 °C. Subsequently, the proteins were defrosted and alkylated in 25 mM iodoacetamide and exchanged four times into 100 mM ammonium bicarbonate buffer using Amicon 3 kDa centrifugal filters.[54] Trypsin digestion was performed overnight at 37 °C. C18 Stage Tips were used for desalting prior to tandem mass spectrometry. Peptides were separated and measured on an EASY-Spray HPLC column (25 cm × 75 μm ID packed with PepMap RSLC C18, 2 μm particles, Thermo Scientific) with an online Easy-nLC 100 chromatography system to a Orbitrap mass spectrometer (Q-Exactive Orbitrap, Thermo Scientific). Precursor ions were selected using data-dependent acquisition (top 10) and fragmented using collision induced dissociation (CID) at a normalized collision energy of 27. Raw MS/MS files were converted to mgf format (Thermo Proteome Discoverer, Thermo Scientific ver. 1.4) and were searched against a sequence database using MASCOT (Matrix Science). Searches employed variable cysteine carbamidomethylation and methionine oxidation. The precursor mass accuracy was set to 10 ppm, while that for product ions was set to 0.02 da. Once identified, the fully digested peptides' intensities were quantified from the area under the curves. Intensities of carbamidomethylated peptides were normalized by dividing against the summed intensity of the three most abundant peptides identified.

Acyl-Intermediate Detection with LC-MS. LC-MS reactions to compare acyl formation with and without the presence of ^{N}SpaA were performed in buffer A and contained a total volume of 100 μL: sortase (25 μM), DTT (5 mM), LPLTG (1 mM), and ^{N}SpaA (250 μM) or buffer. All reactions were incubated at 25 °C and 10 μL time points were removed and frozen with liquid N_2 before being stored at −20 °C. An experiment containing only the enzyme (no peptide added) was performed to provide an external standard for the amount of unmodified enzyme in the mass spectrum. Because the ^{Cd}SrtA-LPLT acyl intermediate cannot readily be separated from unmodified ^{Cd}SrtA by chromatography, its amount in the assay is estimated from the mass spectrometry data by assuming that the acyl-intermediate and unmodified enzyme

https://dx.doi.org/10.1021/acs.bioconjchem.0c00163
Bioconjugate Chem. XXXX, XXX, XXX−XXX

ionize to similar extents. Before being run on the LC-MS system, samples were then diluted with 90 μL of 200 mM L-tryptophan (internal standard) and then measured on a Zorbax 300SB-C3 (3.5 μm, 3.0 × 150 mm) with an Agilent 6530 Q-TOF and Agilent 1260 Infinity HPLC with a gradient of 30−99% over 6 min at 0.8 mL/min. The data was analyzed with Agilent MassHunter Qualitative Analysis. Note that, although the sequence of ^{N}SpaA has been verified, the mass of ^{N}SpaA and the acyl intermediate in the spectrum is 88 Da larger than the mass predicted based on the primary sequence. The origin of this difference is not known. The amount of enzyme and acyl-intermediate was calculated by integrating the area under the curve for each peak.

■ AUTHOR INFORMATION

Corresponding Author

Robert T. Clubb − *Department of Chemistry and Biochemistry, UCLA-DOE Institute for Genomics and Proteomics, and Molecular Biology Institute, University of California, Los Angeles, Los Angeles, California 90095, United States;* orcid.org/0000-0001-5718-3985; Phone: (+1) 310 206 2334; Email: rclubb@mbi.ucla.edu; Fax: (+1) 310 206 4779

Authors

Christopher K. Sue − *Department of Chemistry and Biochemistry and UCLA-DOE Institute for Genomics and Proteomics, University of California, Los Angeles, Los Angeles, California 90095, United States*

Scott A. McConnell − *Department of Chemistry and Biochemistry and UCLA-DOE Institute for Genomics and Proteomics, University of California, Los Angeles, Los Angeles, California 90095, United States*

Ken Ellis-Guardiola − *Department of Chemistry and Biochemistry and UCLA-DOE Institute for Genomics and Proteomics, University of California, Los Angeles, Los Angeles, California 90095, United States*

John M. Muroski − *Department of Chemistry and Biochemistry and UCLA-DOE Institute for Genomics and Proteomics, University of California, Los Angeles, Los Angeles, California 90095, United States*

Rachel A. McAllister − *Department of Chemistry and Biochemistry and UCLA-DOE Institute for Genomics and Proteomics, University of California, Los Angeles, Los Angeles, California 90095, United States*

Justin Yu − *Department of Chemistry and Biochemistry and UCLA-DOE Institute for Genomics and Proteomics, University of California, Los Angeles, Los Angeles, California 90095, United States*

Ana I. Alvarez − *Department of Chemistry and Biochemistry and UCLA-DOE Institute for Genomics and Proteomics, University of California, Los Angeles, Los Angeles, California 90095, United States*

Chungyu Chang − *Molecular Biology Institute and the Division of Oral Biology and Medicine, School of Dentistry, University of California, Los Angeles, Los Angeles, California 90095, United States*

Rachel R. Ogorzalek Loo − *Department of Chemistry and Biochemistry and UCLA-DOE Institute for Genomics and Proteomics, University of California, Los Angeles, Los Angeles, California 90095, United States*

Joseph A. Loo − *Department of Chemistry and Biochemistry and Molecular Biology Institute, University of California, Los Angeles, Los Angeles, California 90095, United States;* orcid.org/0000-0001-9989-1437

Hung Ton-That − *Molecular Biology Institute and the Division of Oral Biology and Medicine, School of Dentistry, University of California, Los Angeles, Los Angeles, California 90095, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.bioconjchem.0c00163

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the U.S. Department of Energy Office of Science, Office of Biological and Environmental Research program under Award Number DE-FC02-02ER63421 and National Institutes of Health Grants AI52217 (R.T.C. and H.T.-T.), DE025015 (H.T.-T.) and GM103479 (J.A.L.). C.K.S. and S.A.M. were supported by a Cellular and Molecular Biology Training Grant (Ruth L. Kirschstein National Research Service Award GM007185). J.M. was supported by a UCLA Molecular Biology Institute Whitcome Fellowship (S10OD016336) NMR equipment used in this research was purchased using funds from shared equipment grant NIH S10OD016336.

■ REFERENCES

(1) Agarwal, P., and Bertozzi, C. R. (2015) Site-Specific Antibody "Drug Conjugates: The Nexus of Bioorthogonal Chemistry, Protein Engineering, and Drug Development. *Bioconjugate Chem. 26* (2), 176−192.

(2) Chudasama, V., Maruani, A., and Caddick, S. (2016) Recent advances in the construction of antibody drug conjugates. *Nat. Chem. 8* (2), 114−119.

(3) Hoyt, E. A., Cal, P. M. S. D., Oliveira, B. L., and Bernardes, G. a. J. L. (2019) Contemporary approaches to site-selective protein modification. *Nature Reviews Chemistry 3* (3), 147−171.

(4) Lagasse, H. A. D., Alexaki, A., Simhadri, V. L., Katagiri, N. H., Jankowski, W., Sauna, Z. E., and Kimchi-Sarfaty, C. (2017) Recent advances in (therapeutic protein) drug development. *F1000Research 6*, 113−113.

(5) Specht, E. A., Braselmann, E., and Palmer, A. E. (2017) A Critical and Comparative Review of Fluorescent Tools for Live-Cell Imaging. *Annu. Rev. Physiol. 79* (1), 93−117.

(6) Sochaj, A. M., Swiderska, K. W., and Otlewski, J. (2015) Current methods for the synthesis of homogeneous antibody drug conjugates. *Biotechnol. Adv. 33* (6), 775−784.

(7) Matsumoto, T., Tanaka, T., and Kondo, A. (2012) Enzyme-mediated methodologies for protein modification and bioconjugate synthesis. *Biotechnol. J. 7* (9), 1137−1146.

(8) Mohamad, N. R., Marzuki, N. H. C., Buang, N. A., Huyop, F., and Wahab, R. A. (2015) An overview of technologies for immobilization of enzymes and surface analysis techniques for immobilized enzymes. *Biotechnol. Biotechnol. Equip. 29* (2), 205−220.

(9) Krall, N., da Cruz, F. P., Boutureira, O., and Bernardes, G. a. J. L. (2016) Site-selective protein-modification chemistry for basic biology and drug development. *Nat. Chem. 8* (2), 103−113.

(10) Spicer, C. D., and Davis, B. G. (2014) Selective chemical protein modification. *Nat. Commun. 5* (1), 4740.

(11) Li, X., Fang, T., and Boons, G.-J. (2014) Preparation of Well-Defined Antibody Drug Conjugates through Glycan Remodeling and Strain-Promoted Azide-Alkyne Cycloadditions. *Angew. Chem., Int. Ed. 53* (28), 7179−7182.

(12) Zhang, Y., Park, K.-Y., Suazo, K. F., and Distefano, M. D. (2018) Recent progress in enzymatic protein labelling techniques and their applications. *Chem. Soc. Rev. 47* (24), 9106−9136.

(13) Antos, J. M., Truttmann, M. C., and Ploegh, H. L. (2016) Recent advances in sortase-catalyzed ligation methodology. *Curr. Opin. Struct. Biol. 38*, 111−118.
(14) Schmohl, L., and Schwarzer, D. (2014) Sortase-mediated ligations for the site-specific modification of proteins. *Curr. Opin. Chem. Biol. 22*, 122−128.
(15) Popp, M. W.-L., and Ploegh, H. L. (2011) Making and Breaking Peptide Bonds: Protein Engineering Using Sortase. *Angew. Chem., Int. Ed. 50* (22), 5024−5032.
(16) Samantaray, S., Marathe, U., Dasgupta, S., Nandicoori, V. K., and Roy, R. P. (2008) Peptideâ'Sugar Ligation Catalyzed by Transpeptidase Sortase:A Facile Approach to Neoglycoconjugate Synthesis. *J. Am. Chem. Soc. 130* (7), 2132−2133.
(17) Mazmanian, S. K., Liu, G., Ton-That, H., and Schneewind, O. (1999) *Staphylococcus aureus* sortase, an enzyme that anchors surface proteins to the cell wall. *Science 285* (5428), 760−3.
(18) Antos, J. M., Chew, G.-L., Guimaraes, C. P., Yoder, N. C., Grotenbreg, G. M., Popp, M. W.-L., and Ploegh, H. L. (2009) Site-specific N- and C-terminal labeling of a single polypeptide using sortases of different specificity. *J. Am. Chem. Soc. 131* (31), 10800−10801.
(19) Popp, M. W., Dougan, S. K., Chuang, T. Y., Spooner, E., and Ploegh, H. L. (2011) Sortase-catalyzed transformations that improve the properties of cytokines. *Proc. Natl. Acad. Sci. U. S. A. 108* (8), 3169−74.
(20) Tsukiji, S., and Nagamune, T. (2009) Sortase-mediated ligation: a gift from Gram-positive bacteria to protein engineering. *ChemBioChem 10* (5), 787−98.
(21) Fottner, M., Brunner, A.-D., Bittl, V., Horn-Ghetko, D., Jussupow, A., Kaila, V. R. I., Bremm, A., and Lang, K. (2019) Site-specific ubiquitylation and SUMOylation using genetic-code expansion and sortase. *Nat. Chem. Biol. 15* (3), 276−284.
(22) McConnell, S. A., Amer, B. R., Muroski, J., Fu, J., Chang, C., Ogorzalek Loo, R. R., Loo, J. A., Osipiuk, J., Ton-That, H., and Clubb, R. T. (2018) Protein Labeling via a Specific Lysine-Isopeptide Bond Using the Pilin Polymerizing Sortase from Corynebacterium diphtheriae. *J. Am. Chem. Soc. 140* (27), 8420−8423.
(23) Jacobitz, A. W., Kattke, M. D., Wereszczynski, J., and Clubb, R. T. (2017) Sortase Transpeptidases: Structural Biology and Catalytic Mechanism. *Adv. Protein Chem. Struct. Biol. 109*, 223−264.
(24) Kang, H. J., and Baker, E. N. (2009) Intramolecular isopeptide bonds give thermodynamic and proteolytic stability to the major pilin protein of *Streptococcus pyogenes*. *J. Biol. Chem. 284* (31), 20729−37.
(25) Comfort, D., and Clubb, R. T. (2004) A comparative genome analysis identifies distinct sorting pathways in gram-positive bacteria. *Infect. Immun. 72* (5), 2710−22.
(26) Echelman, D. J., Alegre-Cebollada, J., Badilla, C. L., Chang, C., Ton-That, H., and Fernandez, J. M. (2016) CnaA domains in bacterial pili are efficient dissipaters of large mechanical shocks. *Proc. Natl. Acad. Sci. U. S. A. 113* (9), 2490−5.
(27) Ton-That, H., and Schneewind, O. (2003) Assembly of pili on the surface of *Corynebacterium diphtheriae*. *Mol. Microbiol. 50* (4), 1429−38.
(28) Chang, C., Amer, B. R., Osipiuk, J., McConnell, S. A., Huang, I. H., Hsieh, V., Fu, J., Nguyen, H. H., Muroski, J., Flores, E., Ogorzalek Loo, R. R., Loo, J. A., Putkey, J. A., Joachimiak, A., Das, A., Clubb, R. T., and Ton-That, H. (2018) In vitro reconstitution of sortase-catalyzed pilus polymerization reveals structural elements involved in pilin cross-linking. *Proc. Natl. Acad. Sci. U. S. A. 115* (24), E5477−E5486.
(29) Frankel, B. A., Kruger, R. G., Robinson, D. E., Kelleher, N. L., and McCafferty, D. G. (2005) *Staphylococcus aureus* sortase transpeptidase SrtA: insight into the kinetic mechanism and evidence for a reverse protonation catalytic mechanism. *Biochemistry 44* (33), 11188−200.
(30) Clancy, K. W., Melvin, J. A., and McCafferty, D. G. (2010) Sortase transpeptidases: insights into mechanism, substrate specificity, and inhibition. *Biopolymers 94* (4), 385−96.
(31) Bradshaw, W. J., Davies, A. H., Chambers, C. J., Roberts, A. K., Shone, C. C., and Acharya, K. R. (2015) Molecular features of the sortase enzyme family. *FEBS J. 282* (11), 2097−114.
(32) Guttilla, I. K., Gaspar, A. H., Swierczynski, A., Swaminathan, A., Dwivedi, P., Das, A., and Ton-That, H. (2009) Acyl enzyme intermediates in sortase-catalyzed pilus morphogenesis in gram-positive bacteria. *J. Bacteriol. 191* (18), 5603−12.
(33) Jacobitz, A. W., Naziga, E. B., Yi, S. W., McConnell, S. A., Peterson, R., Jung, M. E., Clubb, R. T., and Wereszczynski, J. (2016) The "Lid" in the Streptococcus pneumoniae SrtC1 Sortase Adopts a Rigid Structure that Regulates Substrate Access to the Active Site. *J. Phys. Chem. B 120*, 8302.
(34) Chan, A. H., Yi, S. W., Terwilliger, A. L., Maresso, A. W., Jung, M. E., and Clubb, R. T. (2015) Structure of the Bacillus anthracis Sortase A Enzyme Bound to Its Sorting Signal: A Flexible Amino-Terminal Appenage Modulates Substrate Access. *J. Biol. Chem. 290* (42), 25461−74.
(35) Kruger, R. G., Otvos, B., Frankel, B. A., Bentley, M., Dostal, P., and McCafferty, D. G. (2004) Analysis of the Substrate Specificity of the Staphylococcus aureus Sortase Transpeptidase SrtA. *Biochemistry 43* (6), 1541−1551.
(36) Frankel, B. A., Tong, Y., Bentley, M. L., Fitzgerald, M. C., and McCafferty, D. G. (2007) Mutational analysis of active site residues in the *Staphylococcus aureus* transpeptidase SrtA. *Biochemistry 46* (24), 7269−78.
(37) Zong, Y., Bice, T. W., Ton-That, H., Schneewind, O., and Narayana, S. V. (2004) Crystal structures of *Staphylococcus aureus* sortase A and its substrate complex. *J. Biol. Chem. 279* (30), 31383−9.
(38) Huang, X., Aulabaugh, A., Ding, W., Kapoor, B., Alksne, L., Tabei, K., and Ellestad, G. (2003) Kinetic mechanism of *Staphylococcus aureus* sortase SrtA. *Biochemistry 42* (38), 11307−15.
(39) Alcock, L. J., Perkins, M. V., and Chalker, J. M. (2018) Chemical methods for mapping cysteine oxidation. *Chem. Soc. Rev. 47* (1), 231−268.
(40) Kruger, R. G., Dostal, P., and McCafferty, D. G. (2004) Development of a high-performance liquid chromatography assay and revision of kinetic parameters for the Staphylococcus aureus sortase transpeptidase SrtA. *Anal. Biochem. 326* (1), 42−8.
(41) Weiner, E. M., Robson, S., Marohn, M., and Clubb, R. T. (2010) The Sortase A enzyme that attaches proteins to the cell wall of *Bacillus anthracis* contains an unusual active site architecture. *J. Biol. Chem. 285* (30), 23433−43.
(42) Connolly, K. M., Smith, B. T., Pilpa, R., Ilangovan, U., Jung, M. E., and Clubb, R. T. (2003) Sortase from Staphylococcus aureus does not contain a thiolate-imidazolium ion pair in its active site. *J. Biol. Chem. 278* (36), 34061−5.
(43) Bellucci, J. J., Bhattacharyya, J., and Chilkoti, A. (2014) A Noncanonical Function of Sortase Enables Site-Specific Conjugation of Small Molecules to Lysine Residues in Proteins. *Angew. Chem., Int. Ed. 54* (2), 441−445.
(44) Dasgupta, S., Samantaray, S., Sahal, D., and Roy, R. P. (2011) Isopeptide ligation catalyzed by quintessential sortase A: mechanistic cues from cyclic and branched oligomers of indolicidin. *J. Biol. Chem. 286* (27), 23996−24006.
(45) Mohlmann, S., Mahlert, C., Greven, S., Scholz, P., and Harrenga, A. (2011) In vitro Sortagging of an Antibody Fab Fragment: Overcoming Unproductive Reactions of Sortase with Water and Lysine Side Chains. *ChemBioChem 12* (11), 1774−1780.
(46) Deweid, L., Avrutina, O., and Kolmar, H. (2019) Microbial transglutaminase for biotechnological and biomedical engineering. *Biol. Chem. 400*, 257.
(47) Malešević, M., Migge, A., Hertel, T. C., and Pietzsch, M. (2015) A Fluorescence-Based Array Screen for Transglutaminase Substrates. *ChemBioChem 16* (8), 1169−1174.
(48) Jeger, S., Zimmermann, K., Blanc, A., Grünberg, J., Honer, M., Hunziker, P., Struthers, H., and Schibli, R. (2010) Site-Specific and Stoichiometric Modification of Antibodies by Bacterial Transglutaminase. *Angew. Chem., Int. Ed. 49* (51), 9995−9997.

https://dx.doi.org/10.1021/acs.bioconjchem.0c00163
Bioconjugate Chem. XXXX, XXX, XXX−XXX

APPENDIX V: N-terminal autoprocessing and acetylation of multifunctional-autoprocessing repeats-in-toxins (MARTX) Makes Caterpillars Floppy-like effector is stimulated by adenosine diphosphate (ADP)-Ribosylation Factor 1 in advance of Golgi fragmentation

Received: 15 August 2019 | Revised: 26 September 2019 | Accepted: 11 October 2019
DOI: 10.1111/cmi.13133

RESEARCH ARTICLE

WILEY

N-terminal autoprocessing and acetylation of multifunctional-autoprocessing repeats-in-toxins (MARTX) Makes Caterpillars Floppy-like effector is stimulated by adenosine diphosphate (ADP)-Ribosylation Factor 1 in advance of Golgi fragmentation

Alfa Herrera[1] | John Muroski[2] | Ranjan Sengupta[3] | Hong Hanh Nguyen[2] | Shivangi Agarwal[1] | Rachel R. Ogorzalek Loo[5,6] | Seema Mattoo[3,4] | Joseph A. Loo[2,5] | Karla J. F. Satchell[1]

[1] Department of Microbiology-Immunology, Northwestern University Feinberg School of Medicine, Chicago, Illinois

[2] Department of Chemistry and Biochemistry, University of California-Los Angeles, Los Angeles, California

[3] Department of Biological Sciences, Purdue University, West Lafayette, Indiana

[4] Purdue Institute for Inflammation, Immunology and Infectious Diseases, Purdue University, West Lafayette, Indiana

[5] Department of Biological Chemistry, David Geffen School of Medicine, UCLA Molecular Biology Institute, University of California-Los Angeles, Los Angeles, California

[6] UCLA/DOE Institute of Genomics and Proteomics, University of California-Los Angeles, Los Angeles, California

Correspondence
Karla J. F. Satchell, Department of Microbiology-Immunology, Northwestern University Feinberg School of Medicine, Chicago, IL 60611.
Email: k-satchell@northwestern.edu

Present Address
Ranjan Sengupta, Department of Medicinal Chemistry and Molecular Pharmacology, Purdue University, West Lafayette, Indiana.

Hong Hanh Nguyen, IndoChina Center of Excellence, Transmed Co., Ltd, Ho Chi Minh City, Vietnam.

Shivangi Agarwal, Department of Internal Medicine, Rush University Medical Center, Chicago, Illinois.

Funding information
National Institute of Allergy and Infectious Diseases, Grant/Award Numbers: AI092825 and T32AI007476; National Institute of General Medical Sciences, Grant/Award Numbers: GM103479, GM104610 and R01GM10092; Purdue University Institute for Inflammation, Immunology and Infectious Diseases, Grant/Award Numbers: PI4D- 575 209263, GM103479, GM104610 and PI4D-209263; NCI, Grant/Award Number: P30 CA060553; UCLA Molecular Biology Institute Whitcome

Abstract

Studies have successfully elucidated the mechanism of action of several effector domains that comprise the multifunctional-autoprocessing repeats-in-toxins (MARTX) toxins of *Vibrio vulnificus*. However, the biochemical linkage between the cysteine proteolytic activity of Makes Caterpillars Floppy (MCF)-like effector and its cellular effects remains unknown. In this study, we identify the host cell factors that activate in vivo and in vitro MCF autoprocessing as adenosine diphosphate (ADP)-Ribosylation Factor 1 (ARF1) and ADP-Ribosylation Factor 3 (ARF3). Autoprocessing activity is enhanced when ARF1 is in its active [guanosine triphosphate (GTP)-bound] form compared to the inactive [guanosine diphosphate (GDP)-bound] form. Subsequent to auto-cleavage, MCF is acetylated on its exposed N-terminal glycine residue. Acetylation apparently does not dictate subcellular localization as MCF is found localized throughout the cell. However, the cleaved form of MCF gains the ability to bind to the specialized lipid phosphatidylinositol 5-phosphate enriched in Golgi and other membranes necessary for endocytic trafficking, suggesting that a fraction of MCF may be subcellularly localized. Traditional thin-section electron microscopy, high-resolution cryoAPEX localization, and fluorescent microscopy show that MCF causes Golgi dispersal resulting in extensive vesiculation. In addition, host mitochondria are disrupted and fragmented. Mass spectrometry analysis found no reproducible modifications of ARF1 suggesting that ARF1 is not post-translationally modified by MCF. Further, catalytically active MCF does not stably

Fellowship; Ruth L. Kirschstein Institutional National Research Service Award, Grant/Award Number: 5T32AI007476; Indiana Clinical and Translational Sciences Institute, Grant/Award Number: CTSI-106564; National Institutes of Health, Grant/Award Numbers: R01GM10092 and AI092825

associate with ARF1. Our data indicate not only that ARF1 is a cross-kingdom activator of MCF, but also that MCF may mediate cytotoxicity by directly targeting another yet to be identified protein. This study begins to elucidate the biochemical activity of this important domain and gives insight into how it may promote disease progression.

1 | INTRODUCTION

Vibrio vulnificus is a Gram-negative pathogen found in warm marine environments that is capable of causing life-threatening gastrointestinal and wound infections, particularly from handling or consumption of seafood. These infections are rare, although the incidence of *V. vulnificus* is increasing in the US, in part due to expanding geographical distribution due to warming seawaters (Baker-Austin et al., 2012; Deeb, Tufford, Scott, Moore, & Dow, 2018; King et al., 2019). The rising incidence of *V. vulnificus* infections is extremely concerning given its severity, as it often leads to fatal sepsis (Horseman & Surani, 2011). About 90% of patients are hospitalized and more than 50% die, often as fast as 48 hr after symptom onset (CDC, 2014; Gulig, Bourdage, & Starks, 2005). These infections are costly due to high mortality rates and negative impact on economies that depend on recreation and seafood harvesting (Hoffman, Maculloch, & Batz, 2015; Morgan, Stevens, & Degner, 2010). Currently in the US, *V. vulnificus* accounts for only 0.001% of infections, but 3% of all food-related deaths (Mead et al., 1999). According to the US Department of Agriculture, *V. vulnificus* exerts the highest per-case economic burden of any food-borne disease and its total burden exceeds those of *Shigella* and *Escherichia coli* O157:H7 (Hoffman et al., 2015).

For both wound and intestinal infections, the primary *V. vulnificus* virulence factor associated with sepsis and subsequent death is the composite multifunctional-autoprocessing repeats-in-toxins (MARTX) toxin encoded by the gene *rtxA1* (Satchell, 2015). Expression of the MARTX toxin can increase lethality of *V. vulnificus* up to 2,600-fold (Kwak, Jeong, & Satchell, 2011). The *V. vulnificus* MARTX toxin is secreted from the bacterium as a single large polypeptide. It is composed of conserved glycine-rich repeats at both the N- and C-termini between which there are variable effector domains and a cysteine protease domain (CPD). The repeat regions are thought to recognize eukaryotic host cells via a yet undetermined mechanism and then form a pore in the plasma membrane. The effectors and CPD are then translocated across the membrane through this pore. Within the cell, the CPD binds inositol hexakisphosphate, which activates the CPD to auto-cleave at multiple sites to release each of the arrayed effectors individually into the host cytoplasm. The MARTX toxin thus serves as an effector delivery platform for translocation of multiple toxic virulence factors inside cells as a single bolus (Gavin & Satchell, 2015; Kim, 2018; Satchell, 2015).

Across distinct *V. vulnificus* isolates, the effector content can vary from two to five effectors selected from a total of nine known effector domains (Kwak et al., 2011; Ziolo et al., 2014). The effector repertoire can be exchanged by horizontal gene transfer and recombination events such that even closely related strains can have different MARTX toxin variants (Roig, Gonzalez-Candelas, & Amaro, 2011). The effector domain most commonly present in *V. vulnificus* MARTX toxins is the Makes Caterpillars Floppy-like (MCF) effector domain (Agarwal, Agarwal, Biancucci, & Satchell, 2015; Agarwal, Zhu, Gius, & Satchell, 2015). It is found in all Biotype 1 and 2 strains that cause human infections and is even duplicated in the MARTX toxin of some strains (Kwak et al., 2011; Roig et al., 2011).

The MCF effector domain is 376 aa in size. In representative strain CMCP6, MCF is encoded by gene *rtxA1* (*vv2_0479*) nucleotides 9610–10,737 to comprise residues 3204–3579 of the full-length MARTX toxin. Once within cells, MCF has been shown to induce the intrinsic apoptosis pathway, resulting in cell rounding, vesiculation, and loss of cell proliferation (Agarwal, Agarwal, et al., 2015; Agarwal, Zhu, et al., 2015). MCF has been shown to activate caspase-9, -7, and -3 and PARP-γ, and to induce nuclear shrinkage and fragmentation. It also upregulates the proapoptotic proteins Bax, Bak, and Bad to induce loss of mitochondrial transmembrane potential resulting in increased release of cytochrome *c* into the cytosol.

Sequence homology places MCF into the C58 family of cysteine peptidases (Rawlings, Waller, Barrett, & Bateman, 2014). Site-directed mutagenesis revealed that the conserved Cys-148 residue (aa 3351 in full-length toxin) is essential for MCF cytotoxicity. However, even though other putative catalytic residues His-55 (aa 3258 in full-length toxin) and Asp-142 (aa 3345 in full-length toxin) are conserved by sequence alignment with the C58 peptidases that use a Cys-His-Asp (CHD) catalytic triad, alanine substitution of these residues reduced, but did not eliminate, cell rounding (Agarwal, Agarwal, et al., 2015). Instead, MCF has an Arg-Cys-Asp (RCD) peptide triad at residues 147–149 (aa 3350–3352 in full-length toxin) that is necessary for its cytotoxic effects. A homology search revealed numerous known and putative toxin proteins that contain domains with sequence similarity to MCF that also contain the RCD/Y motif (Agarwal, Agarwal, et al., 2015). We suggest that these proteins form a new subfamily of the C58 peptidases.

Upon ectopic expression in host cells, the MCF undergoes autoprocessing at its N-terminus, cleaving itself between Lys-15 and Gly-16, dependent upon an intact RCD motif (Agarwal, Agarwal, et al., 2015). This autoprocessing could be reproduced with recombinant MCF, but only upon addition of eukaryotic cell lysate to the reaction. In this study, we identify the cross-kingdom autoprocessing-

stimulating host factor as Class I and II ADP-ribosylation factors (ARFs), and most prominently ARF1. We further show that subsequent to autoprocessing, the exposed N-terminal Gly-16 is acetylated. After, but not before cleavage, MCF binds phosphatidylinositol-5-phosphate (PtdIns5P), a lipid enriched on the Golgi apparatus. Cells treated with MCF show increased vesiculation caused by dissolution of the Golgi apparatus. Destruction of the Golgi, and also the mitochondria, suggests a mechanism for induction of caspases leading to MARTX toxin dependent cellular apoptosis (Lee, Choi, & Kim, 2008).

2 | RESULTS

2.1 | Purified ARF1 and ARF3 guanosine triphosphate (GTP)ases induce autoprocessing of recombinant MCF protein

MCF is known to be induced to auto-cleave by an unidentified eukaryotic host cell factor. To identify this host factor, MCF encoding gene sequences were truncated to mimic the cleaved form of MCF (cMCF, aa 16–376) and then modified to alter the codon for the catalytic Cys-148 to alanine (cMCF^{CA}). The modified gene was ectopically expressed in human embryonic kidney (HEK) 293T cells fused with an N-terminal FLAG®-tag and C-terminal hemagglutinin (HA)tag to express dual-tagged cMCF^{CA} (dt-cMCF^{CA}). Anti-HA immunoprecipitation revealed a protein of approximately 20 kDa that bound to dt-cMCF^{CA} (Figure 1a). Peptide sequencing of the excised band revealed eight unique peptides. Six mapped to the human protein ADP-Ribosylation Factor (ARF1) and the nearly identical protein ADP-Ribosylation Factor 3 (ARF3). One peptide matched exclusively to ADP-Ribosylation Factor 4 (ARF4) and another to ADP-Ribosylation Factor 6 (ARF6) (Figure 1b). To confirm the identification of ARF1 as a protein that binds MCF, the pull-down was repeated and immunoblotting with antibody against ARF1 detected the band (Figure 1c and Supplemental Figure S1).

To show that ARF1 is also the host factor that induces autoprocessing, full-length Makes Caterpillars Floppy (flMCF) with and without its catalytic Cys-148 (flMCF^{CA}), cMCF, and catalytically inactive cMCF (cMCF^{CA}) were purified as 6xHis-tagged recombinant proteins. Truncated Δ17ARF1 fused with glutathione-S-transferase (GST-Δ17ARF1) was purified as removal of the flexible N-terminus is known to increase protein solubility (Kahn et al., 1992). Both catalytically active and inactive flMCF and cMCF forms co-precipitated with

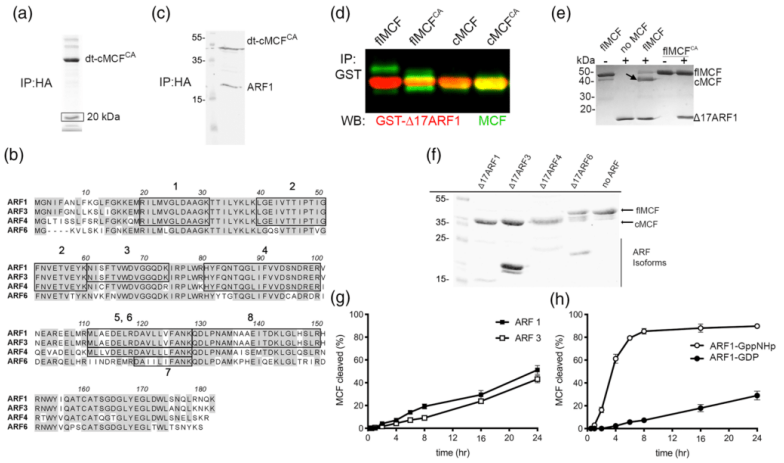


FIGURE 1 Makes Caterpillars Floppy (MCF) autoproteolytic cleavage is induced by adenosine diphosphate (ADP)-Ribosylation Factor (ARF) guanosine triphosphate (GTP)ases. (a) dt-cMCF^{CA} was immunoprecipitated (IP) from HEK293T cell lysate using anti-hyaluronic acid (HA) antibody. Box indicates band excised for mass spectrometry peptide sequencing. (b) Amino acid sequence alignment of major isoforms of ARFs with boxes indicating identified peptides. Peptides marked 6 and 7 are unique to ADP-Ribosylation Factor 4 (ARF4) and ADP-Ribosylation Factor 6 (ARF6), respectively. (c) Immunoblot showing endogenous ADP-Ribosylation Factor 1 (ARF1) pulls-down with dual-tagged cMCF^{CA} (dt-cMCF^{CA}) recovered from whole cell lysate as detected by anti-ARF1 antibody and anti-HA. Replicate pull-downs of independent transfections can be found in Supplemental Figure S1. (d) Purified glutathione-S-transferase (GST)-Δ17ARF1 was incubated with either MCF, Catalytically inactive MCF (MCF^{CA}), cleaved form of MCF (cMCF), or catalytically inactive cleaved MCF (cMCF^{CA}) at 37°C. Western blot on samples following anti-GST IP using anti-MCF (green bands) and anti-ARF1 (red bands) showing catalytically active and inactive full-length MCF (flMCF) and cleaved MCF (cMCF) (green bands) can bind ARF1 (red bands) in vitro. (e) Recombinant flMCF and flMCF^{CA} incubated with ΔARF1 indicate induced autoprocessing is dependent on MCF catalytic active site. Arrow indicates cleaved MCF band. (f) Auto-cleavage induced by purified ARF isoforms incubated with MCF for 24 hr. (g, h) Autoprocessing of flMCF at 37°C for time indicated induced by full-length ARF1 or ARF3 (g) or with ARF1 pre-loaded with GDP or non-hydrolyzable GTP (GppNHp) (h). Representative gels (n > 3), with percent cleaved MCF was determined by densitometry of bands on Coomassie stained gels (% = (cMCF/(cMCF + flMCF) * 100)) from three independent reactions.

GST-Δ17ARF1, confirming that MCF and ARF1 interact and that the binding does not involve the flexible N-terminus of ARF1 (Figure 1d).

The flMCF proteins were next incubated with Δ17ARF1 purified as a 6×-tagged recombinant protein to test stimulation of autoprocessing. Δ17ARF1 stimulated flMCF cleavage and the processing was autocatalytic as flMCF^{CA} was not processed even in the presence of Δ17ARF1 (Figure 1e). Δ17ARF3 and Δ17ARF4 also stimulated autoprocessing, while Δ17ARF6 did not (Figure 1f). Full-length ARF1 and ARF3 with the N-terminus intact were also able to stimulate autoprocessing with similar kinetics (Figure 1g).

ARFs are all members of the Ras superfamily of small GTPases that cycle between an active (GTP-bound) and inactive (GDP-bound) form. ARF1 pre-loaded with GTP-analog GppNHp induced more rapid and more complete MCF auto-cleavage (85 ± 3%) compared to ARF1 pre-loaded with GDP (29 ± 13%) (Figure 1h).

Altogether, these data demonstrate that ARF1 is a host factor that binds and stimulates autoprocessing of MCF. The nearly identical protein Class I ARF3 and the Class II ARF4 can also serve as cross-kingdom host factors capable of inducing MCF auto-cleavage. Further, there is a strong preference for induction of autoprocessing by the active (GTP-bound) form of ARFs.

2.2 | MCF^{CA} functions as a trap for ARF1 in cells

To identify which of the major ARF isoforms are important within cells for induced autoprocessing, dt-cMCF^{CA} was co-expressed in HEK293T cells with Myc-tagged ARF proteins. ARF1, but not ARF4 or ARF6, co-immunoprecipitated with cMCF^{CA} (Figure 2a). We were unable to reproducibly pulldown ARF3 with dt-cMCF^{CA} across independent experiments and thus were not able to confirm this interaction. Thus, although multiple ARF isoforms stimulate MCF autoprocessing in vitro, ARF1 is the only isoform that consistently associates with dt-cMCF^{CA} in cells.

Consistent with its function as a host factor for induction of autoprocessing of flMCF, ARF1 was also co-immunoprecipitated with dt-flMCF^{CA}, indicating interaction between the two proteins occurs prior to the autoprocessing event. By contrast, ARF1 did not co-immunoprecipitate with either dt-flMCF or dt-cMCF if the catalytic Cys-148 was not modified to alanine (Figure 2b). This is not due to an inability of the proteins to associate because, as shown above, GST-Δ17ARF1 co-precipitated with both catalytically active and inactive flMCF and cMCF in vitro (Figure 1d). These data indicate that, in cells, MCF autoprocessing results in release and/or loss of the ability to bind to ARFs such that only catalytically inactive MCF functions as a protein trap. Further, since overexpression or delivery of catalytically inactive MCF to cells does not itself cause cytotoxicity (Agarwal, Agarwal, et al., 2015; Agarwal, Zhu, et al., 2015), trapping of ARFs is not sufficient for MCF-mediated cytotoxicity.

Repeated attempts to identify an MCF-specific processing or post-translational modification of ARF1 or ARF3 in vitro or on proteins recovered from cells were not successful. We did not observe any visible changes to the size of ARF1-Myc when co-transfected with either full length or cMCF or the catalytically inactive versions by western blot (Figure 2a,b). We also did not identify any MCF-specific post-translational modifications on ARF1 or ARF3 using a bottom-up proteomics strategy (Supplemental Figure S2). We thus speculate the preferred association of MCF with ARF1-GTP functions to localize MCF to membranes with active ARF1 where MCF is activated by autoprocessing, and then may access another target at that location.

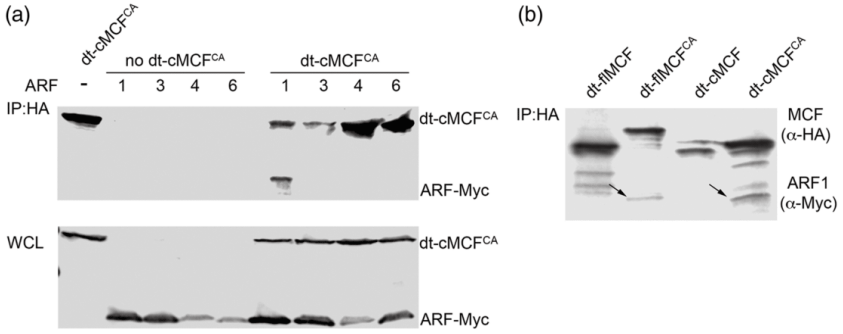


FIGURE 2 Catalytically inactive Makes Caterpillars Floppy (MCF^{CA}) stably associates adenosine diphosphate (ADP)-Ribosylation Factor 1 (ARF1) in HEK293T cells. (a) HEK293T cells were co-transfected with dual-tagged cMCF^{CA} (dt-cMCF^{CA}) and either Myc-tagged ARF1, 3, 4, or 6. Western blot of 40 μg of total protein in whole cell lysates (bottom) recovered from cells. Western blot of 600 μg of total protein from whole cell lysate after immunoprecipitated (IP) using anti-hyaluronic acid (HA) beads (top). MCF and ARFs detected by anti-HA and anti-Myc antibodies respectively with representative gel shown ($n > 3$). (b) HEK293T cells were co-transfected with ARF1-Myc and either dual-tagged full length MCF (dt-flMCF), dual-tagged full length MCF^{CA} (dt-flMCF^{CA}), dual-tagged cleaved MCF (dt-cMCF), or dual-tagged cleaved MCF^{CA} (dt-cMCF^{CA}). Western blot on cell lysates recovered from these cells following anti-HA IP as in (a) followed by simultaneous detection of ARF1 using anti-Myc antibody and MCF using anti-HA antibody. Arrows indicated ARF1-myc. Note there was no pull-down of ARF1 when catalytic cysteine is intact.

2.3 | After autoprocessing, MCF is N-terminally acetylated

In cells, MCF undergoes additional modifications. Previously, attempts to identify the autoprocessing site on MCF from protein recovered from cells revealed the N-terminus was blocked to Edman degradation. The N-terminal block does not occur in vitro. The Edman degradation analysis of immunoprecipitated MCF was repeated for this study and confirmed to be blocked (Supplemental Figure S3A–B). Further, two-dimensional gel analysis of MCF expressed in cells showed that dt-flMCF undergoes in vivo autoprocessing to release its N-terminal FLAG tag, coincident with a shift in electrophoretic mobility, which was detected by western blotting using anti-HA antibody directed against the C-terminal tag. It was further noted that processed flMCF separated into discrete populations that differ in isoelectric point. Catalytically inactive dt-flMCF^{CA} does not shift in size, consistent with the autoprocessing being autocatalytic, or separate into different populations (Figure 3a).

To identify potential post-translational modifications on MCF, the four individual populations were excised and analyzed by bottom-up mass proteomics (Figure 3b and Supplemental Figure S4). The analysis revealed that Gly-16, which is exposed following auto-cleavage, was acetylated (Figure 3c,d and Supplemental Figures S4 and S5). The acetylation was identified in populations 2 and 3, but not in population 1 or 4 (Figure 3d and Supplemental Figures S4 and S5). Acetylation of other lysine residues in populations 2, 3, and 4 was also observed, but the modification of any specific lysine across different experiments or replicates of the same population was not consistent (Figure 3c,d and Supplemental Figures S4 and S5). These observations indicate that only the acetylation of the N-terminal glycine of MCF is specific, while other non-reproducible acetylation events are likely stochastic modifications that arose from overexpression of a protein with many surface exposed lysine residues.

2.4 | MCF binds Golgi enriched lipids and causes Golgi dispersion

ARF1 and ARF3 most commonly localize to the Golgi apparatus (Hofmann & Munro, 2006; Tsai, Adamik, Haun, Moss, & Vaughan, 1993) and ARF association with the Golgi is dependent upon its activation by GTP. As MCF preferentially associates with active ARFs, it may facilitate targeting to subcellular organelles. In addition, many proteins are directed to subcellular organelles by binding specific lipids. Lipid

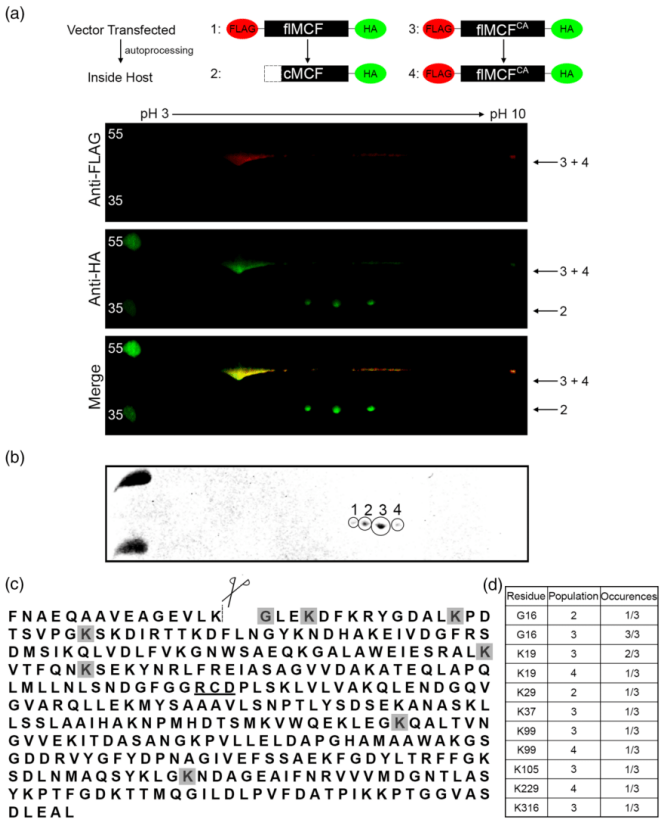


FIGURE 3 Makes Caterpillars Floppy (MCF) is N-terminally acetylated following auto-cleavage. (a) Anti-hyaluronic acid (HA) immunoprecipitated (IP) on cell lysates of HEK293T cells ectopically expressing dual-tagged full length MCF (dt-flMCF) or dual-tagged full length MCF^{CA} (dt-flMCF^{CA}) was individually completed. Samples were then mixed and run together for two-dimensional western analysis using immobilized pH gradient (IPG) strips pH 3–10 with western blot completed using anti-fludarabine, cytarabine, and G-CSF (FLAG) (red) and anti-HA (green) antibodies. The mixing of samples helped to orient cleaved vs uncleaved fragments for the dt-flMCF sample. Red (top panel) and green (middle panel) channels were imaged separately and then merged (bottom panel). Numbers beside panels indicate which version of MCF in the schematic is found in each spot. (b) Two-dimensional analysis with Coomassie gel was completed for only sample of dt-flMCF. Gel shows populations that were individually excised and analyzed by bottom up mass spectrometry. (c) Protein sequence of MCF designating auto-cleavage site (dashed-line), and catalytic residues important for auto-proteolytic activity (underlined). Shaded residues are those found to be acetylated in populations 1–4. (d) Table summarizing specific residues found to be modified and the number of times the modification was detected in each population.

Spot #	PIP Strips * and PIP MicroStrips †
1	Lysophosphatidic acid
2	Lysophosphatidylcholine
3	Phosphatidylinositol (PtdIns)
4	PtdIns(3)P
5	PtdIns(4)P
6	PtdIns(5)P
7	Phosphatidylethanolamine
8	Phosphatidylcholine
9	Sphingosine 1-phosphate
10	PtdIns(3,4)P_2
11	PtdIns(3,5)P_2
12	PtdIns(4,5)P_2
13	PtdIns(3,4,5)P_3
14	Phosphatidic acid
15	Phosphatidylserine
16	Blank

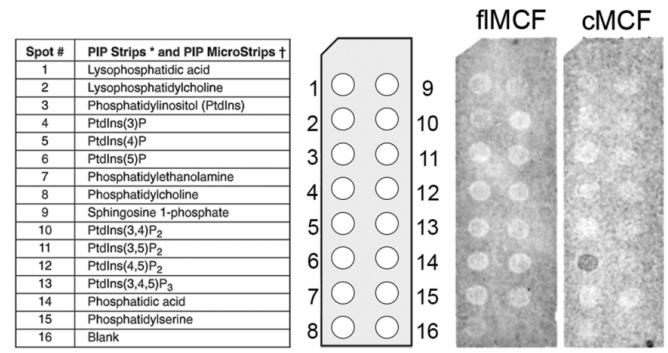


FIGURE 4 Makes Caterpillars Floppy (MCF) associates with Golgi associated lipids. Lipid overlay assay was conducted using PIP Strips™ with either purified full length MCF (flMCF) or cleaved MCF (cMCF). Blot shows cMCF binds phosphatidylinositol-5-phosphate.

overlay assays showed that cMCF, but not flMCF, preferentially bound to PtdIns5P (Figure 4). PtdIns5P is not well described, but has a suggested role in membrane trafficking and is enriched in Golgi membranes (Backer, 2010). Thus, while GTP-bound ARFs may initially direct MCF to the Golgi for activation; after cleavage, we speculate that stable retention of MCF at the Golgi may instead depend on PtdIns5P binding, even as cMCF dissociates from ARFs.

The co-localization of MCF with the Golgi was thus assessed by immunofluorescence (IF). In cells transfected to express only green fluorescent protein (eGFP), the human Golgi matrix protein GM130, a marker for the *cis*-Golgi, showed the Golgi apparatus as tight and condensed at its expected position near the nuclei (Figure 5a). Catalytically inactive flMCF^{CA} frequently co-localized with GM130, although protein also diffused throughout the cell (Figure 5a).

However, the co-localization of flMCF^{CA} does not inform about possible localization of catalytically active flMCF. Thus, we examined whether flMCF with an intact cysteine, that undergoes natural autoprocessing, retains its association with the Golgi or other membranes. The Golgi in cells expressing flMCF-eGFP was found to be disrupted and dispersed in nearly all cells examined, making attempts to co-localize MCF inconclusive. Indeed, GM130 in these cells was not visible unless the image was overexposed (Figure 5a and Supplemental Figure S6A). Similar dispersion was not observed for the endoplasmic reticulum (ER)-associated protein calreticulin in cells

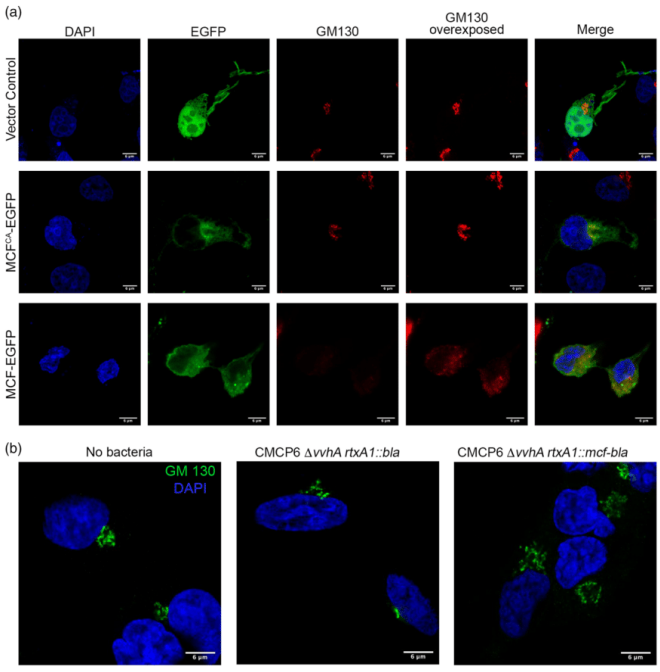


FIGURE 5 Makes Caterpillars Floppy (MCF) induces Golgi dispersion. (a) Cos7 cells were transfected with empty vector, catalytically inactive MCF (MCF^{CA})-express only enhanced green fluorescent protein (eGFP), or MCF-eGFP (green) for 18 hr, fixed, and stained for 4′,6′-diamidino-2-phenylindole (DAPI) (blue), and *cis*-Golgi marker (GM130) (red). (b) Cos7 cells were infected with CMCP6 *rtxA1::bla* or CMCP6 *rtxA1::mcf-bla* at a MOI of 5 for an hour, fixed, and stained for DAPI (blue), and *cis*-Golgi marker (GM130) (green). Histograms quantifying Golgi dispersal for transfected and intoxicated cells in Supplemental Figure S6.

transfected to express MCF-eGFP (Supplemental Figure S7). Thus, catalytically active MCF was found to disrupt the Golgi of host cells and the resulting damage prevented resolution of its intracellular localization using IF.

2.5 | Golgi dispersion and mitochondrial fragmentation occurs when MCF is delivered to cells by the bacterium

To ensure that Golgi disruption is not an artifact of overexpression of MCF by transient transfection, cells treated with live *V. vulnificus* were imaged. Sequences corresponding to the MARTX toxin effectors had been previously deleted from the *rtxA1* gene in strain CMCP6 Δ*vvhA* and replaced with in-frame beta-lactamase gene sequences to generate a fusion gene *rtxA1::bla* that encodes an effector-less MARTX toxin but delivers beta-lactamase as a marker for intact protein and translocation function. Sequences for MCF were then introduced to generate *rtxA1::mcf-bla* that expresses a MARTX toxin that delivers only the MCF effector (Agarwal, Zhu, et al., 2015).

Treatment of cells with CMCP6 Δ*vvhA rtxA1::bla* showed the Golgi remained perinuclear and condensed similar to non-intoxicated cells. In contrast, cells intoxicated with CMCP6 Δ*vvhA rtxA1::mcf-bla* had significantly more dispersed Golgi spreading away from the nucleus (Figure 5b and Supplemental Figure S6B). In addition, the impact of MCF on mitochondria was also visualized using MitoTracker Green FM. The mitochondrial network of cells treated with the control strain CMCP6 Δ*vvhA rtxA1::bla* appeared as long dynamic fragments throughout the cell (Figure 6 and Supplemental Figure S8). In contrast, the mitochondria of cells intoxicated with CMCP6 Δ*vvhA rtxA1::mcf-bla* were shortened and fragmented in comparison. The mitochondrial network in these cells was composed of more short round fragments than long branches and this phenotype was reproducible across multiple experiments (Figure 6 and Supplemental Figure S8).

These data show that MCF delivered to cells by bacteria induces intracellular damage of both the Golgi and the mitochondria.

2.6 | Transmission electron microscopy shows that MCF induces extensive vesiculation of the Golgi apparatus

To gain an ultrastructural understanding of how catalytically active MCF induces Golgi dispersion, transmission electron microscopy was employed. As expected, at 10 hr after transfection, PBS mock transfected HeLa cells showed a Golgi with typical stacked morphology (Figure 7a, green G). By contrast, in cells ectopically expressing MCF, there are a significant number of vesicles in the cytoplasm localized near the Golgi, with some enmeshed within the Golgi complex (Figure 7b and Supplemental Figure S9A). At higher magnification, it

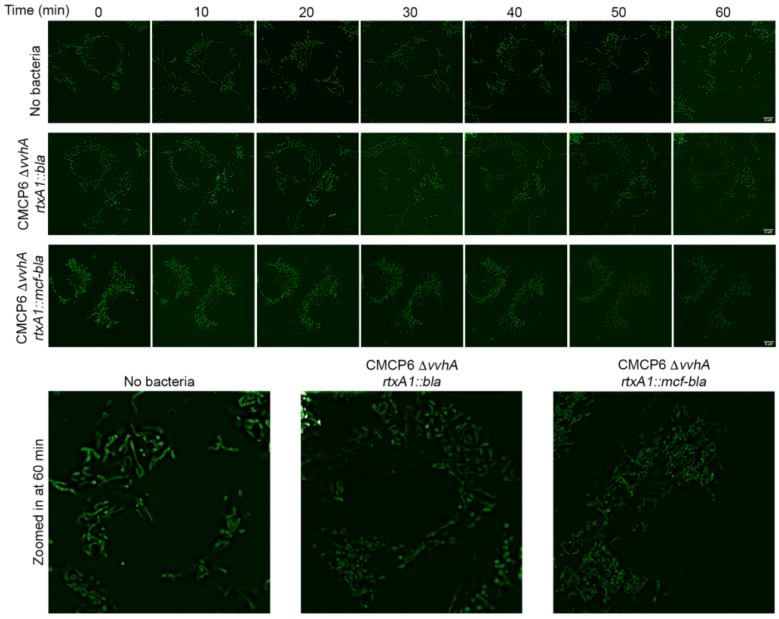


FIGURE 6 Makes Caterpillars Floppy (MCF) disrupts the mitochondrial network. Cos7 cells were incubated with MitoTracker™ Green FM to visualize mitochondria prior to infection with CMCP6 *rtxA1::bla* or CMCP6 *rtxA1::mcf-bla* at a MOI of 5 for an hour. Images taken prior to intoxication and every 10 min after. Insets show enlarged image of one cell for each condition at 60 min. Additional cells are shown in Supplemental Figure S8.

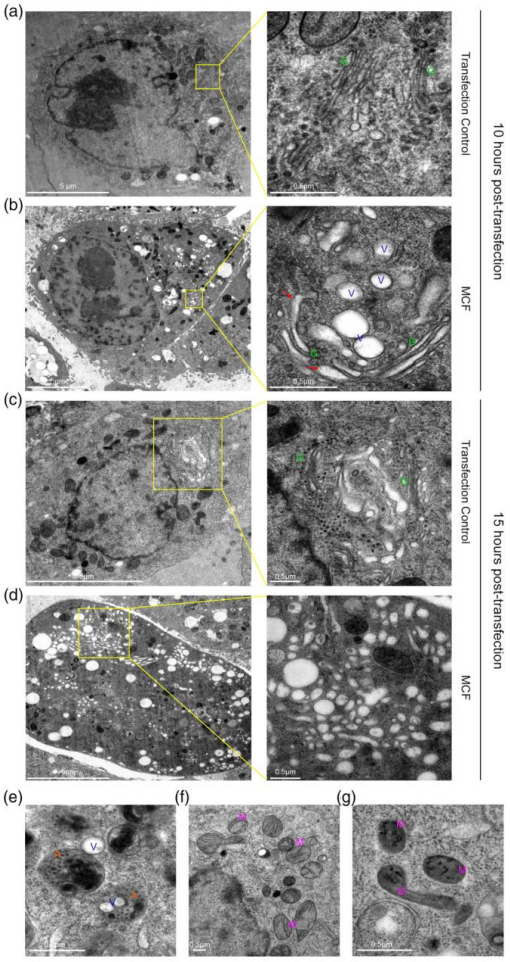


FIGURE 7 Makes Caterpillars Floppy (MCF) causes vesiculation of the Golgi apparatus as seen by transmission electron microscopy. (a)–(g). Thin-section electron microscopy of HeLa cells ectopically expressing MCF, or mock transfected for either 10 (a, b) or 15 hr (c, d) post-transfection. (a–d) On right, higher magnification of boxed region on left showing Golgi (G), herniated Golgi (red arrows), and vesicles (V). (e) Autolysomes (A) clearing MCF induced vesicles. (f, g). Mitochondria (M) of mock transfected cells (f) and MCF transfected cells (g) at 10 hr post-transfection.

was evident that, although some of the Golgi retained a classical stacked morphology some of the cisternae appeared herniated transitioning into a more vesicular morphology (Figure 7b and Supplemental Figure S9A). Additionally, an abundance of autolysosomes was also observed. Some autolysosomes contained vesicles, indicating ongoing clearance off these vesicles (Figure 7e). At 15 hr post transfection, the cytoplasm was inundated with a variety of different sized pleomorphic vesicles (Figure 7d and Supplemental Figure S9B). In these cells the typical stacked structure of the Golgi apparatus, which is seen in mock treated cells, was not observed (Figure 7c).

The electron microscopy data correlate with the IF observations showing darkened mitochondria at 10 hr post-transfection with condensed disintegrating cristae in cells ectopically expressing MCF (Figure 7g). This was in contrast to control cells where mitochondria have an oval shape, with visible cristae (Figure 7f). These data show that both the mitochondria and the Golgi are impacted by expression of MCF.

2.7 | MCF localizes to the Golgi apparatus, the membrane of the vesicular structures and the cytosol

As both N-terminal acetylation and binding to PtdIns5P could dictate subcellular localization, we investigated localization of MCF using electron microscopy based high resolution cryoAPEX technology (Sengupta, Poderycki, & Mattoo, 2019). Briefly, MCF was expressed in HeLa cells as a fusion with a soybean ascorbate peroxidase (Apex2). This tag catalyzes a peroxide reaction that converts diaminobenzidine (DAB) into an insoluble osmophilic precipitate. This precipitate could then be preferentially and specifically stained with osmium tetroxide thereby revealing the subcellular localization of MCF at a nanometer-scale resolution. MCF-Apex2 expressing cells were easily distinguished from non-expressing cells due to dark staining of the cytoplasm and presence of cytoplasmic vesicles compared to the mock transfected control (Figure 8a–c). This ubiquitous staining of the cytosol indicates a large pool of MCF is present in the cytosol. Interestingly, staining was also detected on the membrane of the Golgi stacks visible in samples from the 10-hour time-point as well as on vesicles at both time points (Figure 8b,c, blue arrows). In comparison, control cells transfected with MCF-Apex2 for 10 hr not treated with H_2O_2 exhibit a relatively clear cytoplasm and lack membrane staining of vesicles (Figure 8d). Cells expressing MCF also had darkened staining at the nuclear membrane compared to control cells, indicating a presence of the effector here as well (Figure 8b,c).

3 | DISCUSSION

ADP-ribosylation factors (ARFs) were first identified as activators of the ADP-ribosylation activity of the bacterial protein cholera toxin (Gillingham & Munro, 2007). They subsequently have been shown to have important roles as regulators of eukaryotic cell membrane trafficking (Gillingham & Munro, 2007; Muthamilarasan, Mangu, Zandkarimi, Prasad, & Baisakh, 2016). Depending on the isoform, they localize to specific subcellular locations, including the Golgi, plasma membrane, and endosomes. The Golgi apparatus is generally composed of tight perinuclear Golgi stacks and plays a central role in membrane trafficking, wherein it receives, post-translationally modifies and sorts proteins from the ER to their appropriate destinations (Liu & Storrie, 2015; Wei & Seemann, 2010). At the Golgi, ARF1 is recognized by its guanine nucleotide exchange factors (GEF), inducing a

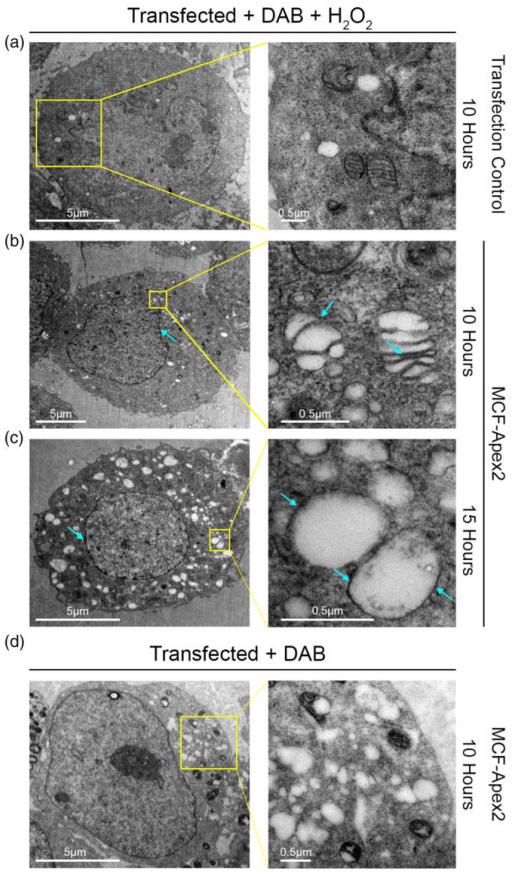


FIGURE 8 Makes Caterpillars Floppy (MCF) concentrates in the cytoplasm, Golgi cisternae, and nuclear and vesicle membranes. (a–d) Thin-section electron microscopy of resin embedded HeLa cells treated with diaminobenzidine (DAB) + H_2O_2 (a–c) or just DAB (d) ectopically expressing MCF (b, c, d), or mock transfected (a) for either 10 or 15 hr post-transfection. On right, higher magnification of boxed region on left. (b–d) Light blue arrows show darkened Apex 2 staining, indicating high concentrations of MCF, seen with H_2O_2 treatment (b, c) and not without (d).

conformational change that exposes its N-terminal amphipathic helix so that it binds GTP in place of GDP (Behnia, Panic, Whyte, & Munro, 2004). This change allows ARF1 to bind the membrane where it plays an important role in recruiting coat complex protein I, clathrin adapter proteins AP1 and AP2, and other proteins necessary for proper vesicle budding and trafficking. This membrane binding is reversed by GTPase-activating proteins (GAPs) that catalyze hydrolysis of the GTP resulting in ARF1 release from the membrane. The nearly identical protein ARF3 has similar function, behaving indistinguishably in most studies. Given the importance of ARFs and the Golgi to cellular processes, bacteria often manipulate ARF function and Golgi stability to promote bacterial infection.

In this study, we show that the pro-form of the *V. vulnificus* MARTX effector MCF is activated by autoprocessing upon binding with Class I ARF1 and ARF3. Although MCF can also be stimulated by ARF4 to autoprocess, these proteins do not associate in cells. Notably, MCF can be stimulated to auto-cleave by either the full-length forms of ARF1 and ARF3 or the more stable Δ17 forms of the proteins. This demonstrates that binding of ARF to MCF does not require the first 17 aa that are essential in vivo for membrane association.

The protein–protein interaction of MCF with ARF1 was confirmed in cells. ARF1 transitions between a GDP- and GTP-bound state dictating its position on and off of the Golgi (Behnia et al., 2004). As MCF is more rapidly and more fully induced to auto-cleave by ARF1-GTP than by ARF1-GDP, MCF is probably preferentially recruited to the Golgi, where it binds ARF1-GTP. Indeed, a low level of active ARF3-GTP at the Golgi under conditions for these experiments may explain our difficulty confirming MCF binding to ARF3 in cells (Tsai et al., 1993).

The cell data also support that once ARFs stimulate autoprocessing, they do not continue to stably associate with MCF, unless the catalytic cysteine is inactivated. This is shown in the coimmunoprecipitation experiments in which $flMCF^{CA}$ or $cMCF^{CA}$ co-precipitate with ARF1-myc, while flMCF and cMCF do not (Figure 2a,b). The lack of binding of ARFs to catalytically active MCF is in contrast to studies using purified proteins that show that ARF1 can bind to both catalytically active and inactive MCF. We originally surmised that this was evidence that ARF1 was modified in cells by catalytically active MCF and then released, although we thus far have found no evidence of cleavage or post-translational modification of ARF1 or ARF3. We therefore suggest that there may be a secondary event or transition that alters the structure of MCF after processing to release ARFs and this exchange may require Cys-148.

Catalytically inactive cMCF was by contrast able to function as a protein trap for ARFs. Yet, this protein does not cause obvious cytotoxicity when overexpressed in cells (Agarwal, Agarwal, et al., 2015; Agarwal, Zhu, et al., 2015). These data contradict potential models whereby MCF functions simply to sequester ARF from its normal function. Thus, we surmise that there is another protein that may be a target for cleavage by catalytically active cleaved MCF. Attempts to identify this host co-factor protein using catalytically inactive MCF as a substrate trap/bait for the initial immunoprecipitation were unsuccessful. It is possible that this protein may be cleaved or degraded in intoxicated cells and thus was difficult to identify using immunoprecipitation strategies.

In cells, MCF Gly-16 is exposed by natural processing and then undergoes acetylation at this newly exposed N-terminal glycine. Recombinant MCF stimulated to process in vitro was not blocked to Edman degradation, even using crude eukaryotic cell lysate as a source of the host factor (Agarwal, Agarwal, et al., 2015). This observation suggests that MCF could be acetylated only in vivo by host N-acetyltransferases when MCF is present within cells. Indeed, the ARF related protein ARFRP1 is N-terminal acetylated by NatC for recruitment to the Golgi. In our case, we speculate that recruitment

to the Golgi is initiated by association with ARF-GTP, but perhaps its retention during Golgi dispersion is aided by the acetylation, particularly as it dissociates from ARFs. This retention may be further augmented also by binding to PtdIns5P.

Our electron microscopy and immunofluorescence studies reveal that expression of active MCF in cells results in dramatic destruction of the Golgi and vesiculation of the cells. The destruction of the Golgi occurred also during natural delivery of the effector from bacteria. Unlike traditional IF techniques, using the novel cryoAPEX approach, the localization of MCF in cells could be successfully visualized. Although possible that acetylation and binding PtdIns5P initially enriched MCF at the Golgi, the overexpressed protein is broadly distributed. It is found throughout the cell, enriched also on mitochondrial membranes and possibly the nuclear membrane. This is supported by cell fractionation that show MCF present in cytosol, nuclear, membrane, and mitochondria fractions (Supplemental Figure S10). At late time points, there were many darkly stained vesicles in the cell that might contain MCF although these are not easily distinguished by the cryoAPEX method from dark staining lysosomes. The direct association with mitochondrial membranes could explain the loss of the mitochondrial membrane potential that ultimately leads to mitochondrial mediated apoptosis through activation of caspases. Likewise, Golgi dispersion is also known to induce apoptosis (Machamer, 2015). Thus, we identify the linkage of MCF to ARFs as the likely cellular process that results in MCF mediated cell death, while the precise molecular events subsequent to ARF stimulated activation of MCF remain to be resolved.

Other bacterial effectors have been identified as directly interacting with ARFs. For example, *Shigella flexneri* Type III secretion effector IpaJ is a member of the C58 family of cysteine peptidases but unlike MCF its activity depends on the canonical CHD active site. Furthermore, IpaJ does not autoprocess but instead, it directly recognizes and cleaves the myristoylated glycine of the lipid modified N-terminus of ARF1 and ARL GTPases. Similar to MCF, IpaJ preferentially targets the GTP-bound active form of ARF1. By cleaving ARF1, IpaJ releases the GTPase from the membrane, thus inducing Golgi disruption (Burnaevskiy et al., 2013). However, despite extensive effort, our negative data support that MCF does not likewise directly cause any processing or modification to ARFs as its mechanism for Golgi dispersion.

As noted above, ARFs are also allosteric activators of the ADP-ribosylating activity of cholera toxin and related heat-labile toxins from enterotoxigenic *E. coli* (Moss & Vaughan, 1991). Similar to the findings here, the N-terminal first 17 aa of ARF1 are dispensable for activation of cholera toxin (Hong, Haun, Tsai, Moss, & Vaughan, 1994). The structure of cholera toxin bound to ARF6 ultimately revealed that binding of ARF6 via its Switch 1 and Switch 2 regions to face of CTA1 ~15 Å away from the CTA1 active site. Binding of ARF6 caused a structural remodeling of the CTA1 loop regions to allow nicotinamide adenine dinucleotide to bind to the active site and to ADP-ribosylate the alpha subunit of the stimulatory G protein G_s (O'Neal, Jobling, Holmes, & Hol, 2005). Thus, there is precedence in *Vibrios* for binding of ARFs as an activator to stimulate structural remodeling of a bacterial effector, then without ARF for the effector to function on a different target protein. Similar structural studies of MCF and MCF bound to ARFs could elucidate these mechanisms for MCF.

These similarities support that stimulation by ARFs that is a unifying theme for *Vibrio* toxins. Although most *V. vulnificus* MARTX toxins have MCF, the MARTX toxin for Biotype 3 strains lack this effector (Ziolo et al., 2014). However, these toxins do deliver a distinct MARTX effector we designated DomainX (DmX). DmX is similar in size to MCF and likewise shows homology to the C58 cysteine peptidases. However, DmX is only 22% identical to MCF and lacks the signature RCD/Y motif. Further, the aligned His and Asp catalytic residues shared with the C58 peptidases were found to be essential for DmX, whereas they were dispensable in MCF (Agarwal, Agarwal, et al., 2015; Kim & Satchell, 2016). DmX was likewise found to autoprocess upon stimulation by ARF1, ARF3, and ARF4; however, all three ARFs bound consistently with DmX in cells, unlike MCF which shows a preference for ARF1. Expression of DmX in cells likewise results in dispersion of the Golgi (Kim & Satchell, 2016). However, unlike MCF, DmX does not have an N-terminus blocked to Edman degradation and is not thought to be post-translationally modified in cells. In addition, catalytically inactive DmX is cytotoxic when expressed in cells, suggesting that ARF sequestering may play a more significant role in its cytotoxic mechanism. We postulate that these two MARTX effectors have an evolutionarily shared mechanism for stimulated autoprocessing to cause Golgi dispersion, but these two effectors have diverged to target the Golgi by distinct molecular mechanisms. Still, the presence of either MCF or DmX in nearly all *V. vulnificus* MARTX toxins reveals that dissolution of vesicular trafficking is important for *V. vulnificus* pathogenesis.

4 | MATERIALS AND METHODS

4.1 | Tissue culture

Cos7 and HEK293T (generously provided by Dr. Richard Longnecker at Northwestern) cells were prepared and cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% heat-inactivated fetal bovine serum (FBS) and 1% penicillin–streptomycin at 37°C in 5% CO_2. Cells were seeded in 150 mm tissue culture dishes or 12-well tissue culture plates, and grown to 60–70% confluence.

4.2 | Bacterial strains and plasmids

gBlocks containing Δ17 ARF1, Δ17 ARF3, Δ17 ARF4, Δ17 ARF6, ARF1, ARF3, MCF, MCF^{CA}, cMCF, and $cMCF^{CA}$ were synthesized by Integrated DNA Technologies (IDT) (Supplemental Table S1). These fragments were cloned into SspI digested vector pMCSG7 using Gibson assembly (New England Biolabs E2611L) (Antic, Biancucci, Zhu, Gius, & Satchell, 2015). N-terminal FLAG and C-terminal HA tagged vectors dt-$flMCF^{CA}$, dt-$cMCF^{CA}$, dt-flMCF, dt-cMCF, and MCF-eGFP plasmids were generated as described in Agarwal, Zhu et al. (2015).

ARF-Myc isoforms were constructed as in Kim and Satchell (2016). CMCP6 Δ*vvhA rtxA1*::*bla* and CMCP6 Δ*vvhA rtxA1*::*mcf-bla V. vulnificus* strains were generated as detailed in Agarwal et al. (2015). MCF-Apex2 was made by Gibson assembly using an APEX2 gBlock synthesized by IDT (Supplemental Table S1) cloned into the dt-flMCF expression plasmid. GST-Δ17ARF1 was generously provided by Dr. Heike Folsch as described in Shteyn, Pigati, and Folsch (2011).

4.3 | Transfections and western blotting

For ectopic gene expression, cells were transfected using 1:3 μg of plasmid DNA to μl of 1 mg/ml polyethylenimine. After 4 hr, the media was exchanged with fresh media. Following an additional 14 hr, the cells were washed with phosphate buffered saline (PBS) and collected. Whole cells were lysed in lysis buffer (300 mM NaCl, 20 mM Tris–HCl pH 8, 1.1% Triton X-100) supplemented with ethylenediaminetetraacetic acid (EDTA) free protease inhibitor tablets (Thermo Scientific A32965) incubated for an hour at 4°C and centrifuged to remove cell debris. A total of 40 μg of total protein in the whole cell lysate was boiled for 5 min in 2× Laemmli sample buffer and separated by SDS-PAGE, followed by transfer to nitrocellulose membrane with the BioRad Mini TransBlot System. Western blots were then completed using anti-hemagglutinin (HA) (Sigma H6908), anti-Myc (Thermo Scientific PA1-981), anti-MCF (produced with purified MCF protein by Lampire Biological Laboratories, Pipersville PA), anti-ARF1 (Novus Biologicals NBP1-97935), anti-FLAG (Sigma F3165), anti-mouse Immunoglobulin G (IgG) (Li-Cor 926-32210 or 926-68070), or anti-rabbit IgG (Li-Cor 926-32211 or 926-68071) antibodies and visualized using the Li-Cor Odyssey Fc imaging system.

4.4 | Immunoprecipitation

A total of 600 μg of total protein of whole cell lysate was incubated overnight at 4°C with mouse IgG agarose (Sigma A0919), agarose removed by centrifugation, and lysate subsequently incubated overnight with EZview™ Red Anti-c-Myc Affinity Gel or EZview™ Red Anti-HA Affinity Gel (Sigma E6779 and E6654). Beads were washed with a lysis buffer supplemented with an EDTA-free protease inhibitor tablets four times, once with 10 mM Tris–HCl pH 8.0, and protein eluted with 2 M NaSCN in 50 mM Tris–HCl pH 8.0, 150 mM NaCl.

4.5 | Protein production and in vitro cleavage assays

Plasmids were expressed in BL21 (DE3) *E. coli* (Wu, Christendat, Dharamsi, & Pai, 2000) to an OD_{600} of 0.8 in terrific broth, induced with 1 M isopropyl β-D-1-thiogalactopyranoside and grown overnight at 28°C. Cells were pelleted, lysed with a sonicator in Buffer A (10 mM Tris–HCl pH 8.0, 500 mM NaCl) with lysozyme, and cell debris removed by centrifugation. Δ17 ARF6 was recovered from the insoluble fraction by incubation for 1 hr in Buffer A with 6 M Urea, followed by centrifugation. Protein was then purified using the AKTA protein purification system with a Ni-NTA HisTrap affinity column (GE Healthcare GE17-5248-02), followed by passage through a Superdex 75 column for size exclusion chromatography. GST-Δ17ARF1 was grown and lysate collected as above and batch purified using glutathione sepharose (GE 17-0756-01).

Purified proteins were used for cleavage assays in which 2.5 μM of specified MCF and 4 μM ARF isoform were incubated in RAX buffer (20 mM Tris pH 7.5, 150 mM NaCl, 10 mM $MgCl_2$) for various time points at 37 °C, Laemmli sample buffer added, and boiled for 5 min prior to separation by SDS-PAGE.

4.6 | Bacterial intoxication

For bacterial treatments, mid-log phase CMCP6 Δ*vvhA rtxA1*::*bla* and CMCP6 Δ*vvhA rtxA1*::*mcf-bla V. vulnificus* strains were pelleted and resuspended in PBS. Cos7 cells were washed twice with PBS, treated with CMCP6 *rtxA1*::*bla* or CMCP6 *rtxA1*::*mcf-bla* at a MOI of 5 in DMEM, centrifuged at 500 ×*g* for 3 min, and incubated for an hour at 37°C with 5% CO_2.

4.7 | Fluorescence microscopy

For immunofluorescence microscopy, images were taken at the Northwestern University Center for Advanced Microscopy using the Nikon A1R+ GaAsP Confocal Laser Microscope or the Nikon N-SIM Structured Illumination Super-resolution Microscope.

Following treatment, cells grown on glass coverslips were washed with PBS and fixed with 4% formaldehyde for 5 min and permeabilized with 0.2% Tween 20. Cells were then blocked with 10% goat serum in staining buffer (1% bovine serum albumin (BSA) plus 0.05% Tween 20 in PBS) for 1 h prior to staining with mouse anti-GM130 (BD Transduction Laboratories 610822) or mouse anti-calreticulin (Abcam ab22683) diluted 1:200 in staining buffer overnight. The coverslips were then washed five times for 5 min each time with staining buffer. Secondary goat anti-mouse Alexa Fluor 647 antibody (Thermo Scientific) was then added at 1:400 in staining buffer and incubated for 1 h. The slides were washed twice with PBS and twice with double distilled water with 2 min between each wash. The slides were mounted using ProLong diamond antifade mountant with 4′,6′-diamidino-2-phenylindole (DAPI) (Invitrogen P36962). Cells ectopically expressing MCF, as detected by GFP fluorescence in at least three fields from three different transfections, were measured for the size of their Golgi in area μm^2 in NIH Image J version 2.0.0-rc-68/1.52h. Cells intoxicated with *V. vulnificus* strains from at least three fields from three different transfections were scored based on degree of Golgi dispersion from 1 to 3, with representative images for each score depicted in Supplemental Figure S6.

For live cell imaging, cells were grown on 35/10 mm glass bottom tissue culture dishes (Greiner Bio-One 627870). Prior to bacterial intoxication, cells were washed with PBS and incubated with 150 nM MitoTracker™ Green FM (Thermo Fisher M7514) in DMEM for 45 min. The cells were then washed with PBS and incubated with 5

μg/ml Hoescht 33342 (Immunochemistry Technologies 639) in PBS. After 10 min the cells were washed, fresh DMEM added, and treated with bacteria as previously described. Images were taken every 5 min following intoxication for an hour.

4.8 | Electron microscopy and cryoAPEX localization

HeLa cells were grown in 10 cm dishes and transfected with MCF-Apex2 using Lipofectamine 3000 (Thermo Fisher). Cells were trypsinized 10 or 15 hr post-transfection, resuspended in DMEM and then pelleted at 500 ×g. For live HPF cells were centrifuged and resuspended in 20% BSA in PBS. Cells were pelleted and loaded onto copper membrane carriers (1 mm × 0.5 mm; Ted Pella Inc.) and cryofixed using the EM PACT2 high pressure freezer (Leica). Cryofixed cells were processed by FS using an AFS2 automated FS unit (Leica) using the extended FS protocol as described in Sengupta et al. (2019). For preparation of samples for cryoAPEX localization, cells were instead resuspended in 0.1% sodium cacodylate buffer containing 2% glutaraldehyde for 30 min following trypsinization, washed three times with 0.1% sodium cacodylate buffer and once with cacodylate buffer containing 1 mg/ml 3,3'-DAB (Sigma-Aldrich). Pellets were then incubated 20 min in 1 mg/ml DAB (Sigma) and 5.88 mM hydrogen peroxide (EM sciences) in cacodylate buffer, pelleted and washed twice for 5 min in cacodylate buffer, once with DMEM, once with DMEM 20% BSA, and pelleted again. These cells were loaded onto copper membrane carriers and cryofixed as described before. An extended FS protocol was optimized for the preferential osmication of the peroxidase-DAB byproduct to ascertain membrane association of MCF. Briefly, frozen pellets were incubated for 60 hr at 90 °C in acetone containing 5% water and 1% osmium tetroxide then at −20°C for a 6 hour osmication cycle. For FS of direct HPF samples, pellets were incubated for 24 hr at −90°C in acetone containing 0.2% tannic acid, washed three times for 5 min with glass distilled acetone (EM Sciences), and resuspended in acetone containing 5% water, 1% osmium tetroxide (OS) and 2% uranyl acetate (UA). Resin exchange was carried out by infiltrating the sample with a gradually increasing concentration of Durcupan ACM resin (Sigma-Aldrich) as follows: 2%, 4% and 8% for 2 hr each and then 15%, 30%, 60%, 90%, 100% and 100% + component C for 4 hr each. Samples were then embedded in resin blocks and polymerized at 60°C for 36 hr. Post-hardening, planchets were extracted by dabbing liquid nitrogen on the membrane carriers. Blocks were embedded a second time and hardened at 60°C for 36 hr. Serial sections, 90 nm thick were obtained using a Leica UC7 microtome. Serial-section ribbons were collected on formvar coated copper slot grids (EM sciences). Sample grids from HPF processing were stained for an hour with 4% UA solution in water followed by Sato's Lead. CryoAPEX samples were not post stained. CryoAPEX images are representative of between 10 and 20 serial sections of 90 nm thickness that were screened and imaged for each sample. Screening was carried out using Tecnai-T12 transmission electron microscope operating at 80 kV and images collected with a Gatan 4K camera.

4.9 | Lipid overlay assays

PIP Strip™ Membranes (Thermo Fisher P23750) were probed as per the manufacturer instructions. Briefly, the membrane was blocked for an hour using TBS-T (10 mM Tris–HCl, pH 8.0, 150 mM NaCl, 0.1% Tween-20) + 3% fatty acid-free BSA (Sigma 126609) (TBS-T + BSA) for an hour at room temperature. The His-tag was removed from purified MCF and cMCF proteins using Tobacco Etch Virus (TEV) protease. Membranes were subsequently incubated with 0.5 μg/ml of either of these proteins in TBS-T + BSA overnight at 4°C and washed 10 times with TBS-T + BSA. The membrane was then reprobed and washed as before. Bound protein was detected using anti-MCF and anti-rabbit IgG antibodies as per the western blotting protocol.

4.10 | Two dimensional western analysis

Anti-HA pulldowns of dt-flMCF^{CA}, dt-cMCF^{CA}-HA, dt-flMCF, and dt-cMCF recovered as described above were prepped using the Bio-Rad ReadyPrep 2-D Cleanup Kit (Bio-Rad 1632130). The samples were prepared separately and then mixed prior to loading onto an 11 cm pH 3–10 immobilized pH gradient (IPG) strips (Bio-Rad 1632014) and rehydrated overnight. Isoelectric focusing was run on a Bio-Rad Protean IEF cell as per the Bio-Rad recommendations. Following IEF, the IPG strip was equilibrated and analyzed by SDS-PAGE and western blotting.

4.11 | Mass spectrometry analysis

Protein embedded in gel bands was digested with 200 ng trypsin (Thermo Scientific) at 37°C overnight. Digest products were extracted from the gel bands in 50% acetonitrile/49.9% water/0.1% trifluoroacetic acid (TFA) and desalted with C18 StageTips prior to analysis by tandem mass spectrometry. Peptides were separated on an EASY-Spray HPLC column (25 cm × 75 μm ID, PepMap RSLC C18, 2 μm, Thermo Scientific) using a gradient of 5–35% acetonitrile in 0.1% formic acid and a flow rate of 300 nl/min for 45 min delivered by an EASY-nLC 1000 nUPLC (Thermo Scientific). Tandem mass spectra were acquired in a data-dependent manner with an Orbitrap Q Exactive mass spectrometer (Thermo Fisher Scientific) interfaced to a nanoelectrospray ionization source.

The raw MS/MS data were converted into MGF format by Thermo Proteome Discoverer

(VER. 1.4, Thermo Scientific). The MGF files were then analyzed by a MASCOT sequence database search. The search was performed using fixed modifications of Met oxidation and Cys carbamidomethylation and variable modifications of Lys acetylation and N-terminal acetylation. Mass accuracy of 10 ppm on precursor ions and 15 ppm of fragment ions were allotted. Peptides identified by the MASCOT search were then manually validated.

To determine the activator of MCF, pulldowns completed on dt-cMCF^{CA}, ectopically expressed in HEK293T cells were analyzed for

total protein identification by mass spectrometry at the University of Illinois at Chicago Mass Spectrometry Core.

ACKNOWLEDGEMENTS

The authors thank Heike Folsch for sharing ARF expression plasmids. We thank the Purdue CryoEM Facility at Hockmeyer Hall for access to the high-pressure freezer, freeze substitution unit, microtome and chemical hood. We thank Dr. Christopher Gilpin for access to the T12 electron microscope at the Purdue Life Sciences EM Core Facility where thin sectioning was carried out. We also thank Dr. Arvanitis and the Center for Advanced Microscopy and Nikon Imaging Center at Northwestern University for their invaluable help and resources. This work was supported by National Institutes of Health grants AI092825 (to K.J.F.S), R01GM10092 (to S.M.), as well as Indiana Clinical and Translational Sciences Institute grant CTSI-106564 (to S.M.), and Purdue University Institute for Inflammation, Immunology and Infectious Diseases grant PI4D-209263 (to S.M.), GM104610 (to R. R.O.L.), and GM103479 (to J.A.L.). A.H. was supported by a Ruth L. Kirschstein Institutional National Research Service Award (5T32AI007476). J.M. was supported by a UCLA Molecular Biology Institute Whitcome Fellowship. Immunofluorescence imaging work was performed at the Northwestern University Center for Advanced Microscopy generously supported by NCI CCSG P30 CA060553 awarded to the Robert H Lurie Comprehensive Cancer Center.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

ORCID

Alfa Herrera https://orcid.org/0000-0002-7081-4208
Rachel R. Ogorzalek Loo https://orcid.org/0000-0002-0580-2833
Seema Mattoo https://orcid.org/0000-0003-3260-0046
Joseph A. Loo https://orcid.org/0000-0001-9989-1437
Karla J. F. Satchell https://orcid.org/0000-0003-3274-7611

REFERENCES

Agarwal, S., Agarwal, S., Biancucci, M., & Satchell, K. J. (2015). Induced autoprocessing of the cytopathic Makes Caterpillars Floppy-like effector domain of the *Vibrio vulnificus* MARTX toxin. *Cellular Microbiology*, *17*(10), 1494–1509. https://doi.org/10.1111/cmi.12451

Agarwal, S., Zhu, Y., Gius, D. R., & Satchell, K. J. (2015). The Makes Caterpillars Floppy (MCF)-like domain of *Vibrio vulnificus* induces mitochondrion-mediated apoptosis. *Infection and Immunity*, *83*(11), 4392–4403. https://doi.org/10.1128/IAI.00570-15

Antic, I., Biancucci, M., Zhu, Y., Gius, D. R., & Satchell, K. J. F. (2015). Site-specific processing of Ras and Rap1 Switch I by a MARTX toxin effector domain. *Nature Communications*, *6*, 7396. https://doi.org/10.1038/ncomms8396

Backer, J. M. (2010). New methods for capturing the mystery lipid, PtdIns5P. *Biochemical Journal*, *428*(3), e1–e2. https://doi.org/10.1042/BJ20100688

Baker-Austin, C., Trinanes, J. A., Taylor, N. G. H., Hartnell, R., Siitonen, A., & Martinez-Urtaza, J. (2012). Emerging *Vibrio* risk at high latitudes in response to ocean warming. *Nature Climate Change*, *3*(1), 73–77. https://doi.org/10.1038/nclimate1628

Behnia, R., Panic, B., Whyte, J. R., & Munro, S. (2004). Targeting of the Arf-like GTPase Arl3p to the Golgi requires N-terminal acetylation and the membrane protein Sys1p. *Nature Cell Biology*, *6*(5), 405–413. https://doi.org/10.1038/ncb1120

Burnaevskiy, N., Fox, T. G., Plymire, D. A., Ertelt, J. M., Weigele, B. A., Selyunin, A. S., … Alto, N. M. (2013). Proteolytic elimination of N-myristoyl modifications by the *Shigella* virulence factor IpaJ. *Nature*, *496*(7443), 106–109. https://doi.org/10.1038/nature12004

CDC. (2014). Cholera and Other Vibrio Illness Surveillance (COVIS) annual summaries 2008–2014. Retrieved from http://www.cdc.gov/nationalsurveillance/cholera-vibrio-surveillance.html

Deeb, R., Tufford, D., Scott, G. I., Moore, J. G., & Dow, K. (2018). Impact of climate change on *Vibrio vulnificus* abundance and exposure risk. *Estuaries and Coasts: Journal of the Estuarine Research Federation*, *41*(8), 2289–2303. https://doi.org/10.1007/s12237-018-0424-5

Gavin, H. E., & Satchell, K. J. (2015). MARTX toxins as effector delivery platforms. *Pathogens and Disease*, *73*(9), ftv092. https://doi.org/10.1093/femspd/ftv092

Gillingham, A. K., & Munro, S. (2007). The small G proteins of the Arf family and their regulators. *Annual Review of Cell and Developmental Biology*, *23*, 579–611. https://doi.org/10.1146/annurev.cellbio.23.090506.123209

Gulig, P. A., Bourdage, K. L., & Starks, A. M. (2005). Molecular pathogenesis of *Vibrio vulnificus*. *Journal of Microbiology*, *43 Spec No*, 118–131.

Hoffman, S. M., Maculloch, B., & Batz, M. (2015). Economic burden of major foodborne illnesses acquired in the United States. *United States Department of Agriculture, Economic Research Service*. Economic Information Bulletin No. (EIB-140). 59 pp.

Hofmann, I., & Munro, S. (2006). An N-terminally acetylated Arf-like GTPase is localised to lysosomes and affects their motility. *Journal of Cell Science*, *119*(Pt 8), 1494–1503. https://doi.org/10.1242/jcs.02958

Hong, J. X., Haun, R. S., Tsai, S. C., Moss, J., & Vaughan, M. (1994). Effect of ADP-ribosylation factor amino-terminal deletions on its GTP-dependent stimulation of cholera toxin activity. *The Journal of Biological Chemistry*, *269*(13), 9743–9745.

Horseman, M. A., & Surani, S. (2011). A comprehensive review of *Vibrio vulnificus*: An important cause of severe sepsis and skin and soft-tissue infection. *International Journal of Infectious Diseases*, *15*(3), e157–e166. https://doi.org/10.1016/j.ijid.2010.11.003

Kahn, R. A., Randazzo, P., Serafini, T., Weiss, O., Rulka, C., Clark, J., … Rothman, J. E. (1992). The amino terminus of ADP-ribosylation factor (ARF) is a critical determinant of ARF activities and is a potent and specific inhibitor of protein transport. *The Journal of Biological Chemistry*, *267*(18), 13039–13046.

Kim, B. S. (2018). The modes of action of MARTX toxin effector domains. *Toxins (Basel)*, *10*(12). https://doi.org/10.3390/toxins10120507

Kim, B. S., & Satchell, K. J. (2016). MARTX effector cross kingdom activation by Golgi-associated ADP-ribosylation factors. *Cellular Microbiology*, *18*(8), 1078–1093. https://doi.org/10.1111/cmi.12568

King, M., Rose, L., Fraimow, H., Nagori, M., Danish, M., & Doktor, K. (2019). *Vibrio vulnificus* infections from a previously nonendemic area. *Annals of Internal Medicine*, *171*, 520. https://doi.org/10.7326/L19-0133

Kwak, J. S., Jeong, H. G., & Satchell, K. J. (2011). *Vibrio vulnificus rtxA1* gene recombination generates toxin variants with altered potency during intestinal infection. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(4), 1645–1650. https://doi.org/10.1073/pnas.1014339108

Lee, B. C., Choi, S. H., & Kim, T. S. (2008). *Vibrio vulnificus* RTX toxin plays an important role in the apoptotic death of human intestinal epithelial cells exposed to *Vibrio vulnificus*. *Microbes and Infection*, *10*(14-15), 1504–1513. https://doi.org/10.1016/j.micinf.2008.09.006

Liu, S., & Storrie, B. (2015). How Rab proteins determine Golgi structure. *International Review of Cell and Molecular Biology*, *315*, 1–22. https://doi.org/10.1016/bs.ircmb.2014.12.002

Machamer, C. E. (2015). The Golgi complex in stress and death. *Frontiers in Neuroscience*, *9*, 421. https://doi.org/10.3389/fnins.2015.00421

Mead, P. S., Slutsker, L., Dietz, V., McCaig, L. F., Bresee, J. S., Shapiro, C., … Tauxe, R. V. (1999). Food-related illness and death in the United States. *Emerging Infectious Diseases*, *5*(5), 607–625. https://doi.org/10.3201/eid0505.990502

Morgan, K. L., Stevens III, T. J., & Degner, R. L. (2010). Economic impacts of alternative regulatory scenarios on the Florida fresh half shell oyster industry. Florida Agricultural Market Research Center, University of Florida, Gainsville, FL. Industry Report 10-1.

Moss, J., & Vaughan, M. (1991). Activation of cholera toxin and *Escherichia coli* heat-labile enterotoxins by ADP-ribosylation factors, a family of 20 kDa guanine nucleotide-binding proteins. *Molecular Microbiology*, *5*(11), 2621–2627. https://doi.org/10.1111/j.1365-2958.1991.tb01971.x

Muthamilarasan, M., Mangu, V. R., Zandkarimi, H., Prasad, M., & Baisakh, N. (2016). Structure, organization and evolution of ADP-ribosylation factors in rice and foxtail millet, and their expression in rice. *Scientific Reports*, *6*, 24008. https://doi.org/10.1038/srep24008

O'Neal, C. J., Jobling, M. G., Holmes, R. K., & Hol, W. G. (2005). Structural basis for the activation of cholera toxin by human ARF6-GTP. *Science*, *309*(5737), 1093–1096. https://doi.org/10.1126/science.1113398

Rawlings, N. D., Waller, M., Barrett, A. J., & Bateman, A. (2014). MEROPS: The database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Research*, *42*(Database issue), D503–D509. https://doi.org/10.1093/nar/gkt953

Roig, F. J., Gonzalez-Candelas, F., & Amaro, C. (2011). Domain organization and evolution of multifunctional autoprocessing repeats-in-toxin (MARTX) toxin in *Vibrio vulnificus*. *Applied and Environmental Microbiology*, *77*(2), 657–668. https://doi.org/10.1128/AEM.01806-10

Satchell, K. J. F. (2015). Multifunctional-autoprocessing repeats-in-toxin (MARTX) toxins of Vibrios. *Microbiology Spectrum*, *3*(3). https://doi.org/10.1128/microbiolspec.VE-0002-2014

Sengupta, R., Poderycki, M. J., & Mattoo, S. (2019). CryoAPEX — An electron tomography tool for subcellular localization of membrane proteins. *Journal of Cell Science*, *132*(6). https://doi.org/10.1242/jcs.222315

Shteyn, E., Pigati, L., & Folsch, H. (2011). Arf6 regulates AP-1B-dependent sorting in polarized epithelial cells. *Journal of Cell Biology*, *194*, 873–887. https//doi.org/10.1083/jcb.201106010

Tsai, S. C., Adamik, R., Haun, R. S., Moss, J., & Vaughan, M. (1993). Effects of brefeldin A and accessory proteins on association of ADP-ribosylation factors 1, 3, and 5 with Golgi. *The Journal of Biological Chemistry*, *268*(15), 10820–10825.

Wei, J. H., & Seemann, J. (2010). Unraveling the Golgi ribbon. *Traffic*, *11*(11), 1391–1400. https://doi.org/10.1111/j.1600-0854.2010.01114.x

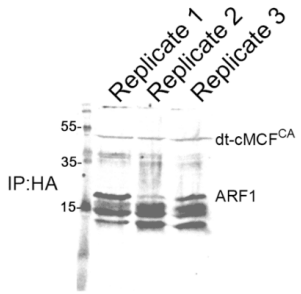
Wu, N., Christendat, D., Dharamsi, A., & Pai, E. F. (2000). Purification, crystallization and preliminary X-ray study of orotidine 5′-monophosphate decarboxylase. *Acta Crystallographica. Section D, Biological Crystallography*, *56*(Pt 7), 912–914. https://doi.org/10.1107/s090744490000576x

Ziolo, K. J., Jeong, H. G., Kwak, J. S., Yang, S., Lavker, R. M., & Satchell, K. J. (2014). *Vibrio vulnificus* biotype 3 multifunctional autoprocessing RTX toxin is an adenylate cyclase toxin essential for virulence in mice. *Infection and Immunity*, *82*(5), 2148–2157. https://doi.org/10.1128/IAI.00017-14

SUPPORTING INFORMATION

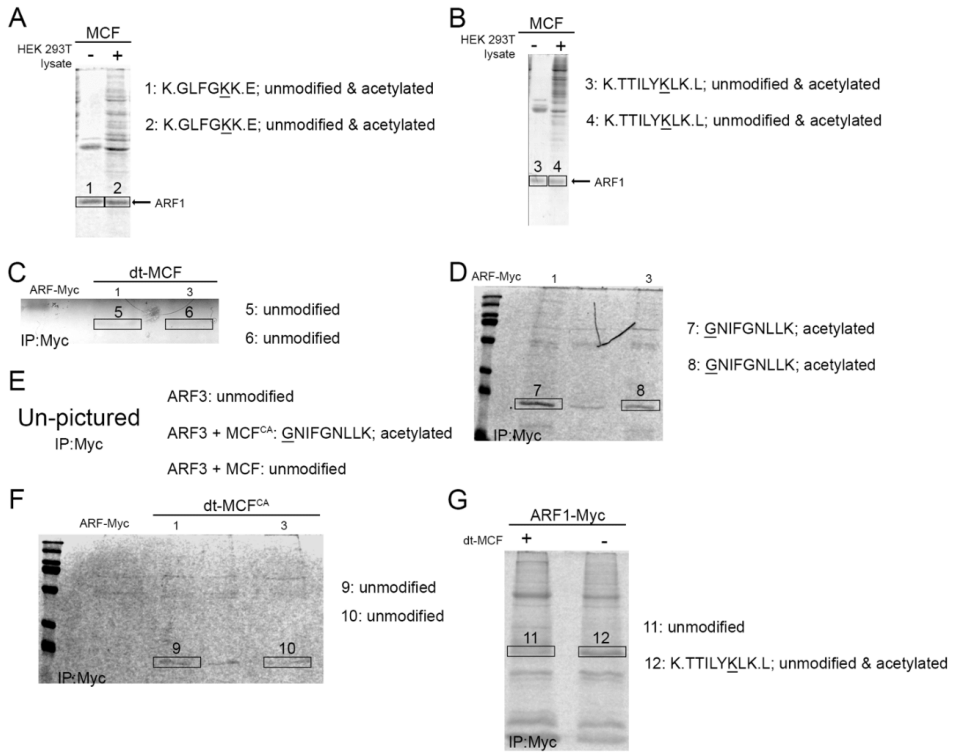
Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Herrera A, Muroski J, Sengupta R, et al. N-terminal autoprocessing and acetylation of multifunctional-autoprocessing repeats-in-toxins (MARTX) Makes Caterpillars Floppy-like effector is stimulated by adenosine diphosphate (ADP)-Ribosylation Factor 1 in advance of Golgi fragmentation. *Cellular Microbiology*. 2020;22:e13133. https://doi.org/10.1111/cmi.13133



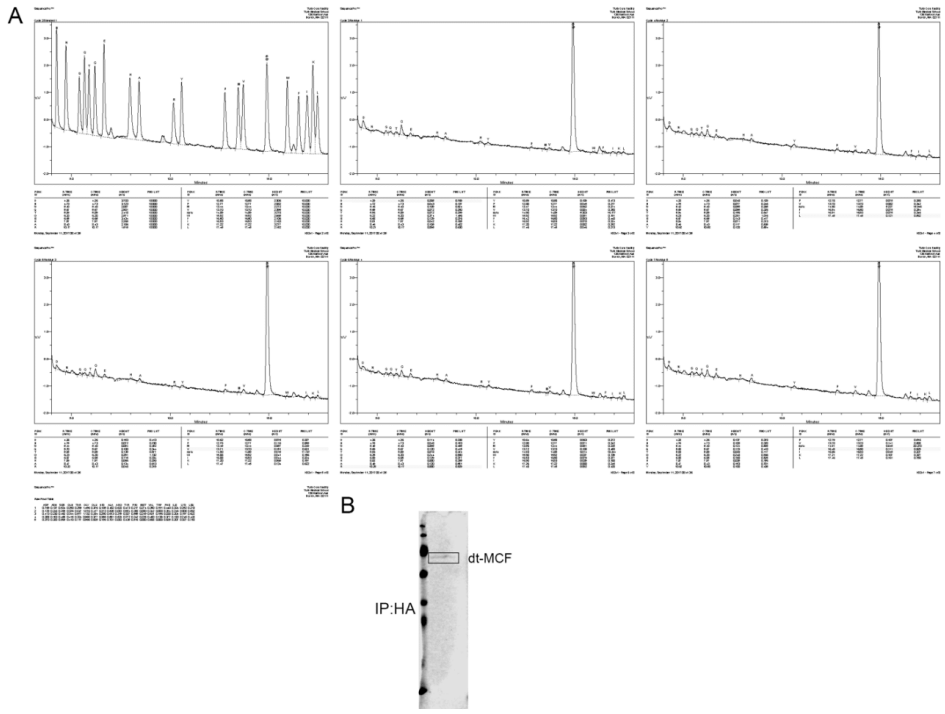
Supplemental Figure 1. MCF reproducibly pulls down native ARF1.

HEK293T cells were transfected with dt-cMCF^{CA} in triplicate. For each replicate, MCF was immunoprecipitated from lysate using anti-HA beads and analyzed by Western blot. dt-cMCF^{CA} was detected by anti-HA and endogenous ARF1 detected by anti-ARF1 antibodies.



Supplemental Figure 2. MCF does not cleave or directly modify ARF1 or ARF3.

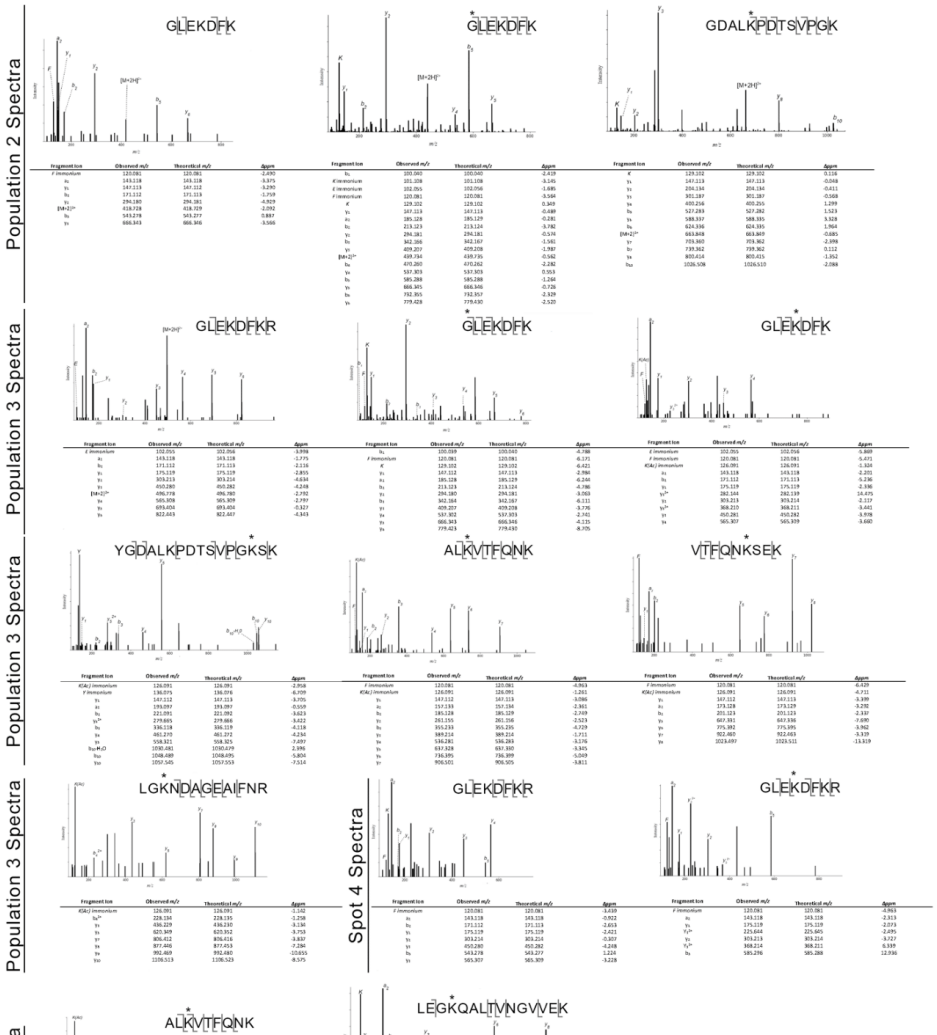
A-G. Bottom-up mass spectrometry was performed on ARF1 and ARF3 samples recovered from recombinant proteins incubated with (A and B), or anti-Myc IPs (C-G) from HEK293T cells transfected with and without MCF or MCF^{CA}. Modifications that were detected are denoted. Any modifications found on either ARF1 or ARF3 were not reproducible across replicates. Furthermore, there were no modifications detected that were found in ARF + MCF samples that were not found in ARF samples alone and thus not attributable to MCF.



Supplemental Figure 3. Edman degradation is blocked in MCF recovered from cells.

A. Reads from Edman degradation analysis completed on dt-flMCF immunoprecipitated from HEK293T cell lysate using anti-HA antibody. No signal and very little background was detected.

B. Gel of sample used for analysis.

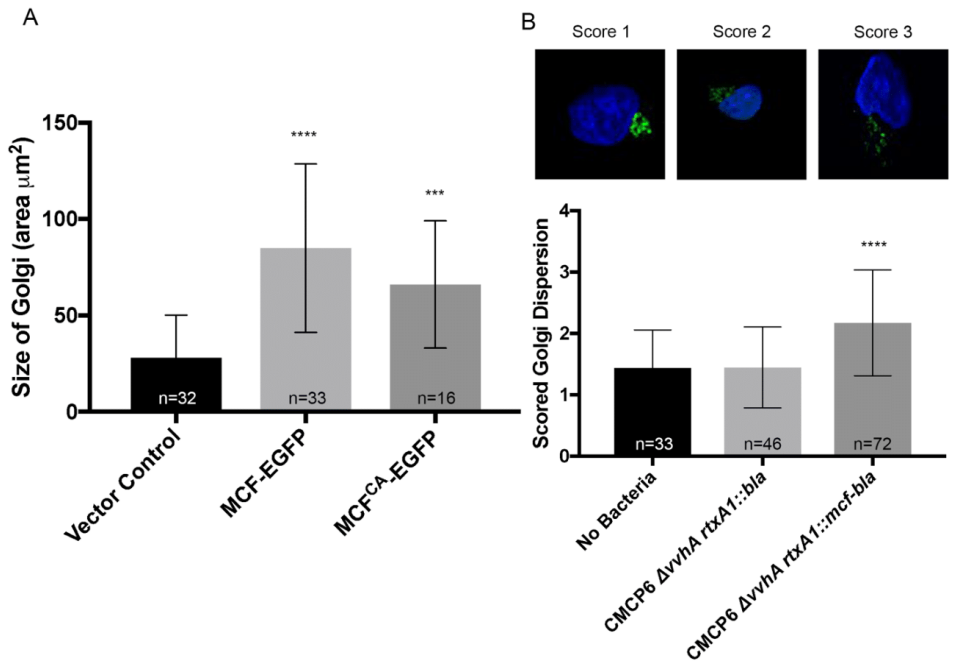


Supplemental Figure 4. Mass Spectrometry analysis shows MCF is acetylated inside cells.

Representative mass spectrometry spectra for each modification detected per population in the replicates of the two-dimensional analysis of dt-flMCF expressed in HEK293T cells recovered by anti-HA IP shown in Supplemental Fig 5. Asterisk above peptide sequence denotes residue acetylated in that particular spectra.

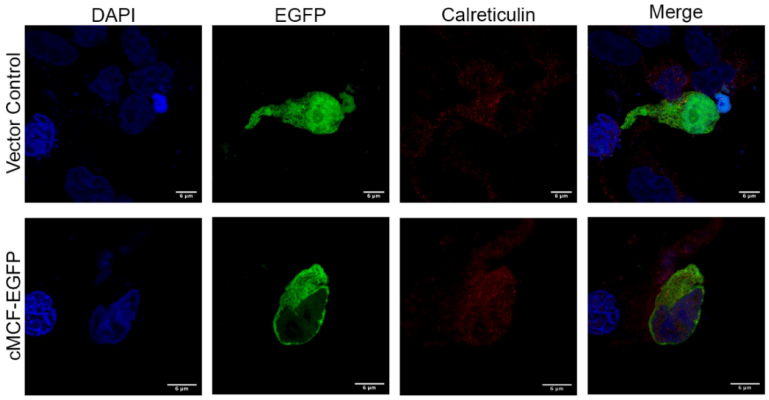
Supplemental Figure 5. MCF is N-terminally acetylated inside host cells.

Replicates of two-dimensional gel analysis of dt-flMCF expressed in HEK293T cells, recovered by anti-HA immunoprecipitation. Amino acids acetylated in each circled population indicated in corresponding color on sequence. Representative mass spectrometry data for the acetylations shown in Supplemental Fig 4.

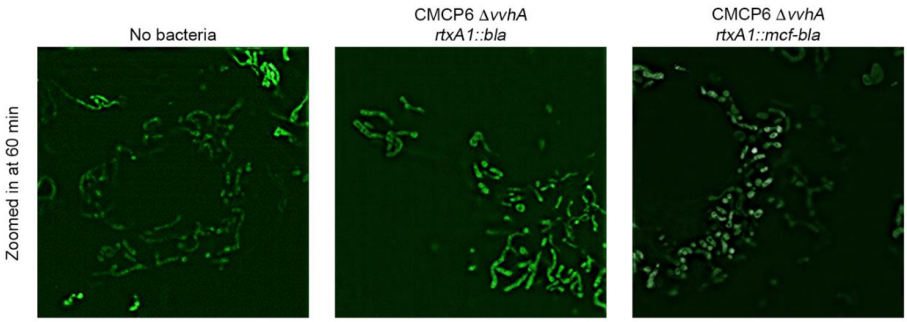


Supplemental Figure 6. Golgi staining is more diffuse in cells expressing MCF.

A. Area in μm^2 of Golgi staining by IF measured for individual Cos7 cells ectopically expressing the specified vector. B.The Golgi of Cos7 cells intoxicated with *V. vulnificus* strains was scored on a scale of 1-3 for extent of dispersal. Images representative of the amount of dispersion each score signifies are shown.

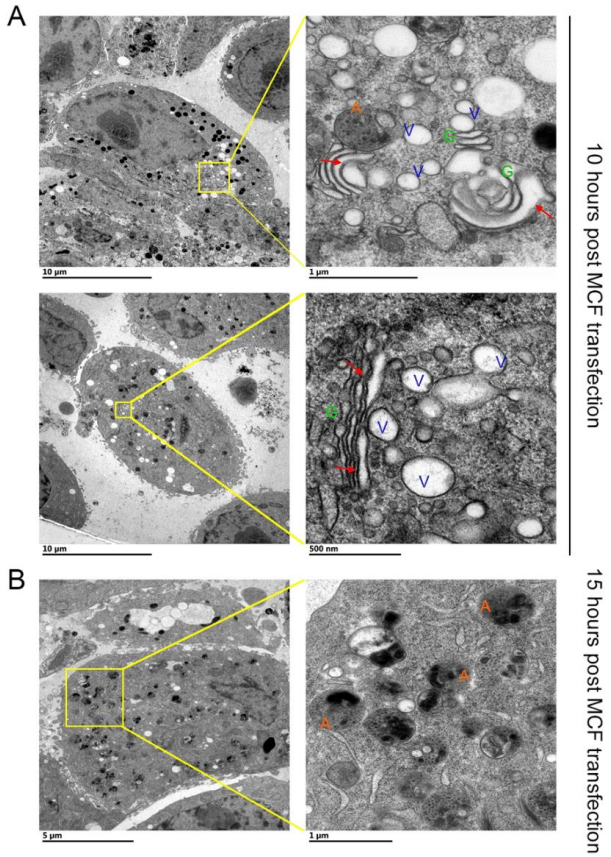


Supplemental Figure 7. MCF does not alter normal endoplasmic reticulum structure. Cos7 cells were transfected with empty vector or MCF-EGFP (green) for 18 hours, fixed, and stained for DAPI (blue), and endoplasmic reticulum marker (calreticulin) (red).



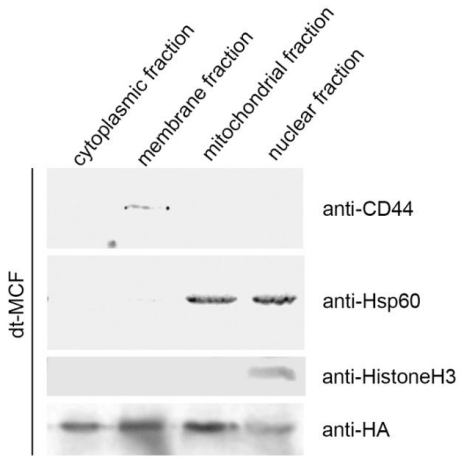
Supplemental Figure 8. MCF induces fragmentation of host mitochondria

Replicates for cells intoxicated, as in Fig 6, with *V. vulnificus* strains and stained with Mitotracker. Images show mitochondria of additional cells across a different experiment at 60 minutes.



Supplemental Figure 9. Transmission electron microscopy shows Golgi vesiculation induced by MCF.

A, B. Electron microscopy tomograms taken of HeLa cells ectopically expressing MCF for 10 (A) or 15 hours (B). A-D. On right, higher magnification of boxed region on left showing Golgi (G), herniated Golgi (red arrows), vesicles (V). and autolysomes (A).



Supplemental Figure 10. Western blot of subcellular fractions from HEK 293T cells ectopically expressing dt-flMCF. Each fraction probed with standards for membrane (CD-44), mitochondria (Hsp-60), and nucleus (HistoneH3), and MCF (anti-HA).

Supplemental Table 1. Sequences of gBlocks used for plasmid construction.	
Primer/Gblock	**Sequence**
Δ17ARF3	CCTGTACTTCCAATCCAATGCTATGCGCATCCTGATGGTGGGCCTGGATGCC GCAGGAAAGACCACCATCCTATACAAGCTGAAACTGGGGGAGATCGTCACCA CCATCCCTACCATTGGGTTCAATGTGGAGACAGTGGAGTATAAGAACATCAG CTTTACAGTGTGGGATGTGGGTGGCCAGGACAAGATTCGACCCCTCTGGAGA CACTACTTCCAGAACACCCAAGGGTTGATATTTGTGGTCGACAGCAATGATCG GGAGCGAGTAAATGAGGCCCGGGAAGAGCTGATGAGAATGCTGGCGGAGGA CGAGCTCCGGGATGCTGTACTCCTTGTCTTTGCAAACAAACAGGATCTGCCT AATGCTATGAACGCTGCTGAGATCACAGACAAGCTGGGCCTGCATTCCCTTC GTCACCGTAACTGGTACATTCAGGCCACCTGTGCCACCAGCGGGGACGGGC TGTACGAAGGCCTGGACTGGCTGGCCAATCAGCTCAAAAACAAGAAGTGATA AATTGGAAGTGGATAACGG
Δ17ARF4	CCTGTACTTCCAATCCAATGCTATGCGCATTTTGATGGTTGGATTGGATGCTG CTGGCAAGACAACCATTCTGTATAAACTGAAGTTAGGGGAGATAGTCACCAC CATTCCTACCATTGGTTTTAATGTGGAAACAGTAGAATATAAGAACATTTGTTT CACAGTATGGGATGTTGGTGGTCAAGATAGAATTAGGCCTCTCTGGAAGCAT TACTTCCAGAATACCCAGGGTCTTATTTTTGTGGTAGATAGCAACGATCGTGA AAGAATTCAGGAAGTAGCAGATGAGCTGCAGAAAATGCTTCTGGTAGATGAAT TGAGAGATGCAGTGCTGCTACTTTTTGCAAACAAACAGGATTTGCCAAATGCT ATGGCCATCAGTGAAATGACAGATAAACTAGGGCTTCAGTCTCTTCGTAACAG AACATGGTATGTTCAAGCCACTTGTGCAACACAAGGAACTGGTCTGTATGAAG GACTTGACTGGCTGTCAAATGAGCTTTCAAAACGTTAATAAATTGGAAGTGGA TAACGG
Δ17ARF6	CCTGTACTTCCAATCCAATGCTATGTTGGGCCTGGACGCGGCCGGCAAGACA ACAATCCTGTACAAGTTGAAGCTGGGCCAGTCGGTGACCACCATTCCCACTG TGGGTTTCAACGTGGAGACGGTGACTTACAAAAATGTCAAGTTCAACGTATGG GATGTGGGCGGCCAGGACAAGATCCGGCCGCTCTGGCGGCATTACTACACT GGGACCCAAGGTCTCATCTTCGTAGTGGACTGCGCCGACCGCGACCGCATC GATGAGGCTCGCCAGGAGCTGCACCGCATTATCAATGACCGGGAGATGAGG GACGCCATAATCCTCATCTTCGCCAACAAGCAGGACCTGCCCGATGCCATGA AACCCCACGAGATCCAGGAGAAACTGGGCCTGACCCGGATTCGGGACAGGA ACTGGTATGTGCAGCCCTCCTGTGCCACCTCAGGGGACGGACTCTATGAGG GGCTCACATGGTTAACCTCTAACTACAAATCTTAATAAATTGGAAGTGGATAA CGG
ARF1	CCTGTACTTCCAATCCAATGCTATGGGGAACATCTTCGCCAACCTCTTCAAGG GCCTTTTTGGCAAAAAAGAAATGCGCATCCTCATGGTGGGCCTGGATGCTGC AGGGAAGACCACGATCCTCTACAAGCTTAAGCTGGGTGAGATCGTGACCACC ATTCCCACCATAGGCTTCAACGTGGAAACCGTGGAGTACAAGAACATCAGCT TCACTGTGTGGGACGTGGGTGGCCAGGACAAGATCCGGCCCCTGTGGCGCC ACTACTTCCAGAACACACAAGGCCTGATCTTCGTGGTGGACAGCAATGACAG AGAGCGTGTGAACGAGGCCCGTGAGGAGCTCATGAGGATGCTGGCCGAGGA CGAGCTCCGGGATGCTGTCCTCCTGGTGTTCGCCAACAAGCAGGACCTCCC CAACGCCATGAATGCGGCCGAGATCACAGACAAGCTGGGGCTGCACTCACT ACGCCACAGGAACTGGTACATTCAGGCCACCTGCGCCACCAGCGGCGACGG GCTCTATGAAGGACTGGACTGGCTGTCCAATCAGCTCCGGAACCAGAAGTAA ATTGGAAGTGGATAACGG
flARF3	CCTGTACTTCCAATCCAATGGGCAATATCTTTGGAAACCTTCTCAAGAGCCTG ATTGGGAAGAAGGAGATGCGCATCCTGATGGTGGGCCTGGATGCCGCAGGA

	AAGACCACCATCCTATACAAGCTGAAACTGGGGGAGATCGTCACCACCATCC CTACCATTGGGTTCAATGTGGAGACAGTGGAGTATAAGAACATCAGCTTTACA GTGTGGGATGTGGGTGGCCAGGACAAGATTCGACCCCTCTGGAGACACTAC TTCCAGAACACCCAAGGGTTGATATTTGTGGTCGACAGCAATGATCGGGAGC GAGTAAATGAGGCCCGGGAAGAGCTGATGAGAATGCTGGCGGAGGACGAGC TCCGGGATGCTGTACTCCTTGTCTTTGCAAACAAACAGGATCTGCCTAATGCT ATGAACGCTGCTGAGATCACAGACAAGCTGGGCCTGCATTCCCTTCGTCACC GTAACTGGTACATTCAGGCCACCTGTGCCACCAGCGGGGACGGGCTGTACG AAGGCCTGGACTGGCTGGCCAATCAGCTCAAAAACAAGAAGTAAATTGGAAG TGGATAACGG
cMCF	CCTGTACTTCCAATCCAATGCTATGGGACTAGAGAAAGACTTTAAACGCTATG GCGACGCGCTGAAACCAGATACGAGCGTGCCGGGTAAATCGAAAGACATTC GCACCACTAAAGATTTCCTAAATGGTTACAAAAATGACCATGCGAAAGAGATC GTTGACGGCTTCCGCTCAGATATGAGTATCAAGCAACTGGTGGATCTGTTTGT TAAAGGTAACTGGAGTGCAGAGCAAAAAGGTGCGCTTGCTTGGGAAATCGAA AGTCGTGCACTGAAAGTGACGTTCCAGAACAAGTCTGAGAAGTACAACCGAT TGTTCCGTGAGATTGCTTCTGCTGGCGTGGTGGATGCGAAAGCGACTGAACA GCTTGCGCCACAGTTAATGCTGCTGAACCTATCGAATGACGGTTTTGGTGGG CGTTGTGATCCACTTTCTAAACTCGTTTTGGTTGCGAAACAGCTTGAAAACGA TGGTCAAGTTGGCGTGGCAAGACAACTGCTAGAAAAGATGTACTCTGCGGCA GCGGTGCTGAGCAATCCAACCCTTTACTCAGACAGTGAAAAAGCCAATGCAA GCAAGTTGCTCAGCAGCTTGGCGGCCATTCATGCGAAGAACCCAATGCATGA TACGTCGATGAAAGTGTGGCAGGAAAAGCTGGAAGGGAAGCAAGCGCTGAC CGTAAACGGTGTGGTTGAGAAAATCACTGATGCATCGGCTAACGGTAAACCT GTGCTGTTGGAACTTGATGCTCCGGGGCATGCGATGGCAGCTTGGGCAAAA GGCTCAGGCGACGATCGTGTTTACGGCTTCTACGATCCAAATGCTGGCATCG TTGAGTTTTCGTCAGCAGAGAAGTTTGGCGACTACCTAACGCGTTTCTTCGGC AAGTCCGATCTGAACATGGCTCAAAGCTATAAGCTGGGTAAAAACGACGCAG GTGAAGCAATCTTCAACCGCGTGGTGGTAATGGATGGCAATACATTAGCAAG CTACAAGCCGACCTTCGGTGACAAGACCACCATGCAGGGGATCCTAGATCTA CCTGTGTTTGACGCTACACCGATTAAAAAGCCTACGGGTGGAGTCGCGAGCG ATCTCGAAGCATTGTAAATTGGAAGTGGATAACGG
flMCF	CCTGTACTTCCAATCCAATGCTATGGCAGCGGTAGAAGCGGGCGAAGTGCTG AAAGGACTAGAGAAAGACTTTAAACGCTATGGCGACGCGCTGAAACCAGATA CGAGCGTGCCGGGTAAATCGAAAGACATTCGCACCACTAAAGATTTCCTAAAT GGTTACAAAAATGACCATGCGAAAGAGATCGTTGACGGCTTCCGCTCAGATA TGAGTATCAAGCAACTGGTGGATCTGTTTGTTAAAGGTAACTGGAGTGCAGA GCAAAAAGGTGCGCTTGCTTGGGAAATCGAAAGTCGTGCACTGAAAGTGACG TTCCAGAACAAGTCTGAGAAGTACAACCGATTGTTCCGTGAGATTGCTTCTGC TGGCGTGGTGGATGCGAAAGCGACTGAACAGCTTGCGCCACAGTTAATGCTG CTGAACCTATCGAATGACGGTTTTGGTGGGCGTTGTGATCCACTTTCTAAACT CGTTTTGGTTGCGAAACAGCTTGAAAACGATGGTCAAGTTGGCGTGGCAAGA CAACTGCTAGAAAAGATGTACTCTGCGGCAGCGGTGCTGAGCAATCCAACCC TTTACTCAGACAGTGAAAAAGCCAATGCAAGCAAGTTGCTCAGCAGCTTGGC GGCCATTCATGCGAAGAACCCAATGCATGATACGTCGATGAAAGTGTGGCAG GAAAAGCTGGAAGGGAAGCAAGCGCTGACCGTAAACGGTGTGGTTGAGAAA ATCACTGATGCATCGGCTAACGGTAAACCTGTGCTGTTGGAACTTGATGCTCC GGGGCATGCGATGGCAGCTTGGGCAAAAGGCTCAGGCGACGATCGTGTTTA CGGCTTCTACGATCCAAATGCTGGCATCGTTGAGTTTTCGTCAGCAGAGAAG TTTGGCGACTACCTAACGCGTTTCTTCGGCAAGTCCGATCTGAACATGGCTCA AAGCTATAAGCTGGGTAAAAACGACGCAGGTGAAGCAATCTTCAACCGCGTG GTGGTAATGGATGGCAATACATTAGCAAGCTACAAGCCGACCTTCGGTGACA AGACCACCATGCAGGGGATCCTAGATCTACCTGTGTTTGACGCTACACCGAT TAAAAAGCCTACGGGTGGAGTCGCGAGCGATCTCGAAGCATTGTAAATTGGA AGTGGATAACGG

flMCF^{CA}	CCTGTACTTCCAATCCAATGCTATGGCAGCGGTAGAAGCGGGCGAAGTGCTG AAAGGACTAGAGAAAGACTTTAAACGCTATGGCGACGCGCTGAAACCAGATA CGAGCGTGCCGGGTAAATCGAAAGACATTCGCACCACTAAAGATTTCCTAAAT GGTTACAAAAATGACCATGCGAAAGAGATCGTTGACGGCTTCCGCTCAGATA TGAGTATCAAGCAACTGGTGGATCTGTTTGTTAAAGGTAACTGGAGTGCAGA GCAAAAAGGTGCGCTTGCTTGGGAAATCGAAAGTCGTGCACTGAAAGTGACG TTCCAGAACAAGTCTGAGAAGTACAACCGATTGTTCCGTGAGATTGCTTCTGC TGGCGTGGTGGATGCGAAAGCGACTGAACAGCTTGCGCCACAGTTAATGCTG CTGAACCTATCGAATGACGGTTTTGGTGGGCGTGCTGATCCACTTTCTAAACT CGTTTTGGTTGCGAAACAGCTTGAAAACGATGGTCAAGTTGGCGTGGCAAGA CAACTGCTAGAAAAGATGTACTCTGCGGCAGCGGTGCTGAGCAATCCAACCC TTTACTCAGACAGTGAAAAAGCCAATGCAAGCAAGTTGCTCAGCAGCTTGGC GGCCATTCATGCGAAGAACCCAATGCATGATACGTCGATGAAAGTGTGGCAG GAAAAGCTGGAAGGGAAGCAAGCGCTGACCGTAAACGGTGTGGTTGAGAAA ATCACTGATGCATCGGCTAACGGTAAACCTGTGCTGTTGGAACTTGATGCTCC GGGGCATGCGATGGCAGCTTGGGCAAAAGGCTCAGGCGACGATCGTGTTTA CGGCTTCTACGATCCAAATGCTGGCATCGTTGAGTTTTCGTCAGCAGAGAAG TTTGGCGACTACCTAACGCGTTTCTTCGGCAAGTCCGATCTGAACATGGCTCA AAGCTATAAGCTGGGTAAAAACGACGCAGGTGAAGCAATCTTCAACCGCGTG GTGGTAATGGATGGCAATACATTAGCAAGCTACAAGCCGACCTTCGGTGACA AGACCACCATGCAGGGGATCCTAGATCTACCTGTGTTTGACGCTACACCGAT TAAAAAGCCTACGGGTGGAGTCGCGAGCGATCTCGAAGCATTGTAAATTGGA AGTGGATAACGG
CMCF^{CA}	CCTGTACTTCCAATCCAATGCTATGGGACTAGAGAAAGACTTTAAACGCTATG GCGACGCGCTGAAACCAGATACGAGCGTGCCGGGTAAATCGAAAGACATTC GCACCACTAAAGATTTCCTAAATGGTTACAAAAATGACCATGCGAAAGAGATC GTTGACGGCTTCCGCTCAGATATGAGTATCAAGCAACTGGTGGATCTGTTTGT TAAAGGTAACTGGAGTGCAGAGCAAAAAGGTGCGCTTGCTTGGGAAATCGAA AGTCGTGCACTGAAAGTGACGTTCCAGAACAAGTCTGAGAAGTACAACCGAT TGTTCCGTGAGATTGCTTCTGCTGGCGTGGTGGATGCGAAAGCGACTGAACA GCTTGCGCCACAGTTAATGCTGCTGAACCTATCGAATGACGGTTTTGGTGGG CGTGCTGATCCACTTTCTAAACTCGTTTTGGTTGCGAAACAGCTTGAAAACGA TGGTCAAGTTGGCGTGGCAAGACAACTGCTAGAAAAGATGTACTCTGCGGCA GCGGTGCTGAGCAATCCAACCCTTTACTCAGACAGTGAAAAAGCCAATGCAA GCAAGTTGCTCAGCAGCTTGGCGGCCATTCATGCGAAGAACCCAATGCATGA TACGTCGATGAAAGTGTGGCAGGAAAAGCTGGAAGGGAAGCAAGCGCTGAC CGTAAACGGTGTGGTTGAGAAAATCACTGATGCATCGGCTAACGGTAAACCT GTGCTGTTGGAACTTGATGCTCCGGGGCATGCGATGGCAGCTTGGGCAAAA GGCTCAGGCGACGATCGTGTTTACGGCTTCTACGATCCAAATGCTGGCATCG TTGAGTTTTCGTCAGCAGAGAAGTTTGGCGACTACCTAACGCGTTTCTTCGGC AAGTCCGATCTGAACATGGCTCAAAGCTATAAGCTGGGTAAAAACGACGCAG GTGAAGCAATCTTCAACCGCGTGGTGGTAATGGATGGCAATACATTAGCAAG CTACAAGCCGACCTTCGGTGACAAGACCACCATGCAGGGGATCCTAGATCTA CCTGTGTTTGACGCTACACCGATTAAAAAGCCTACGGGTGGAGTCGCGAGCG ATCTCGAAGCATTGTAAATTGGAAGTGGATAACGG
CMV7.1.kpnI.APEX2.f	ACGATGTTCCAGATTACGCTGGTACCGGAGGAGGATCATCATCA
CMV7.1.kpnI.APEX2.r	ACCCGGGATCCTCTAGAGTCGACTGGTACCTTAGGCATCAGCAAACCCAAG

APPENDIX VI: *Syntrophomonas wolfei* subsp. *methylbutyratica*, a first look into the proteome

INTRODUCTION

The bacteria *Syntrophomonas wolfei* subsp. *Göttingen* (DSM2245) has been understood to degrade short fatty acid chains of chain lengths of 4-8 carbons long (1). This sub-strain has been the model used to elucidate the properties of the *S. wolfei* species, and in turn has been used as a model for understanding the syntrophic anaerobic degradation of short-chain fatty acids (2–5). In 2007, a novel subspecies of *S. wolfei* was identified in rice field mud in China (6). Uniquely, this organism was able to degrade 2-methylbutyrate (6), a feature not seen in the *Göttingen* subspecies (1). Despite this metabolic difference, the 16S ribosomal sequences of the two subspecies shared 98.9% sequence similarity (6). In order to better understand the enzymes that allow for 2-methylbutyrate degradation, as well as explore the differences between these two species, we have begun to investigate the proteome of this species. Here, we investigated the subspecies grown axenically on crotonate to get a first look at features of the proteome.

METHODS

Samples were obtained from the McInerney group at The University of Oklahoma. Cells were grown axenically on crotonate as *S. wolfei* subsp *Göttingen* had been previously (7).

Cells were lysed using the enhanced filter assisted sample preparation (eFASP) protocol (8, 9). After ethyl acetate extraction, peptides were desalted using C18 spin tips. 200 ng of lysate peptides were separated on an EASY nLC-1200 (Thermo Scientific) and loaded onto the QExactive orbitrap mass spectrometer (Thermo Scientific). Ions were selected for fragmentation using data-dependent acquisition mode where the top 10 most abundant ions were chosen for

fragmentation. Selected ions were fragmented using higher-energy collisional dissociation. Protein identification was perfomed using the Mascot database search algorithm (Matrix Science) whereby *.RAW files were searched using ProteomeDiscoverer (version 1.4) software (Thermo Scientific). The database used for protein identification was the draft genome available at the JGI database under the draft genome for *Syntrophomonas wolfei methylbutyratica* JCM 14075 (10).

RESULTS AND DISCUSSION

Initial reports identified that the subsp. m*ethylbutyratica* was able to sporulate under axenic and coculture conditions, whereas the subsp. *Göttingen* could not (6). The proteomic analysis supported this observation (**Figure 1**). Two homologs involved in sporulation were identified as expressed in equal abundances. The "spore germination and sporulation protein" homologs (Swol_1052 *vs* Ga0126451_105152) and the "stage V sporulation protein homologs" (Swol_1250 *vs* Ga0126451_12349) are present in equal abundances. However, two sporulation related factors, "sporulation transcription factor Spo0A" (Swol_0590 *vs* Ga0126451_12038) and "stage 0 sporulation family protein" (Swol_0043 *vs* Ga0126451_101324) were identified as present in subsp. *methylbutyratica* but not in subsp. *Göttingen*. This could potentially help explain the disparity in the sporulation phenotype between the two subspecies, especially considering the Spo0A gene is essential for spore formation (11).

The gene loci that are upregulated as part of the acyl-degradation pathway are also of significant interest. As the new subspecies can degrade crotonate as well as 2-methylbutyrate, it would be prudent to identify the genes that are expressed in this pathway. As axenic crotonate growth is conserved between the subspecies, it is useful to compare the two. **Figure 2** demonstrates which genes are expressed, allowing for a baseline for the novel syntroph's

degradation enzymes. This is a reasonable comparison for future studies when grown on other carbon sources.

Finally, it was important to catalog the presence of acylation that were found to occur in the subsp. *methylbutyratica* proteome. This would help to identify a conserved mechanism of modification with other sythrophs studied, as well as the subsp. *Göttingen* studies that were performed in Chapter 4 of this dissertation. Indeed, it was found that modifications associated with the β-oxidation pathway were found to be present with acetylation making up the bulk of the acylation identified (**Table 1**). Also similar to previous reports of the *Göttingen* strain in Chapter 4, there is a large amount of modification primarily on enzymes within the β-oxidation pathway. **Figure 3** demonstrates the heterogeneity of the heterogeneity of three enzymes in the pathway that were identified. Not only are the enzyme modified at many different lysine residues in the protein but also many different lysine residues are modified with a heterogeneous mix of different acylations. While this study only looked at one condition, it does appear that acylation of these pathways is a conserved mechanism of the *S. wolfei* species.

CONCLUSION

The *methylbutyratica* subspecies is a very important factor in understanding the physiology of the *S. wolfei* species of syntrophs. It allows for very distinct points of comparison that allow us to tease apart mechanisms underlying the phenotypes of the model syntroph. It also critically helps determine what mechanisms and phenomenon are conserved across the species and potentially beyond. Future studies may also allow for a much better understanding of the unique properties this subspecies has within the *S. wolfei* species to degrade branched chain fatty acids.

ACKNOWLEDGEMENTS

I would like to thank Derek Lam and the Gunsalus Lab at UCLA for helping to overlay the putative enzymes in the sbsp. *methylbutyratica* degradation pathway with their gene loci in **Figure 2**.

CITATIONS

1. McInerney, M. J., Bryant, M. P., and Pfennig, N. (1979) Anaerobic bacterium that degrades fatty acids in syntrophic association with methanogens. *Arch. Microbiol.* **122**, 129–135

2. Sieber, J. R., Sims, D. R., Han, C., Kim, E., Lykidis, A., Lapidus, A. L., McDonnald, E., Rohlin, L., Culley, D. E., Gunsalus, R., and McInerney, M. J. (2010) The genome of Syntrophomonas wolfei: new insights into syntrophic metabolism and biohydrogen production. *Environ. Microbiol.* **12**, no-no

3. Müller, N., Worm, P., Schink, B., Stams, A. J. M., and Plugge, C. M. (2010) Syntrophic butyrate and propionate oxidation processes: From genomes to reaction mechanisms. *Environ. Microbiol. Rep.* **2**, 489–499

4. Schmidt, A., Müller, N., Schink, B., and Schleheck, D. (2013) A Proteomic View at the Biochemistry of Syntrophic Butyrate Oxidation in Syntrophomonas wolfei. *PLoS One*. 10.1371/journal.pone.0056905

5. McInerney, M. J., and Wofford, N. Q. (1992) Enzymes involved in crotonate metabolism in Syntrophomonas wolfei. *Arch. Microbiol.* **158**, 344–349

6. Wu, C., Dong, X., and Liu, X. (2007) Syntrophomonas wolfei subsp. methylbutyratica subsp. nov., and assignment of Syntrophomonas wolfei subsp. saponavida to Syntrophomonas saponavida sp. nov. comb. nov. *Syst. Appl. Microbiol.* **30**, 376–380

7. Beaty, P. S., and McInerney, M. J. (1987) Growth of Syntrophomonas wolfei in pure culture on crotonate. *Arch. Microbiol.* **147**, 389–393

8. Erde, J., Loo, R. R. O., and Loo, J. A. (2014) Enhanced FASP (eFASP) to increase proteome coverage and sample recovery for quantitative proteomic experiments. *J. Proteome Res.* **13**, 1885–1895

9. Erde, J., Loo, R. R. O., and Loo, J. A. (2017) Improving proteome coverage and sample recovery with enhanced FASP (eFASP) for quantitative proteomic experiments. in *Methods in Molecular Biology*, pp. 11–18, Humana Press Inc., **1550**, 11–18

10. Narihiro, T., Nobu, M. K., Tamaki, H., Kamagata, Y., and Liu, W. T. (2016) Draft genome sequence of Syntrophomonas wolfei subsp. methylbutyratica strain 4J5T (JCM 14075), a mesophilic butyrate- and 2-methylbutyrate-degrading syntroph. *Genome Announc.* 10.1128/genomeA.00047-16

11. Brown, D. P., Ganova-Raeva, L., Green, B. D., Wilkinson, S. R., Young, M., and Youngman, P. (1994) Characterization of spo0A homologues in diverse Bacillus and Clostridium species identifies a probable DNA-binding domain. *Mol. Microbiol.* **14**, 411–426

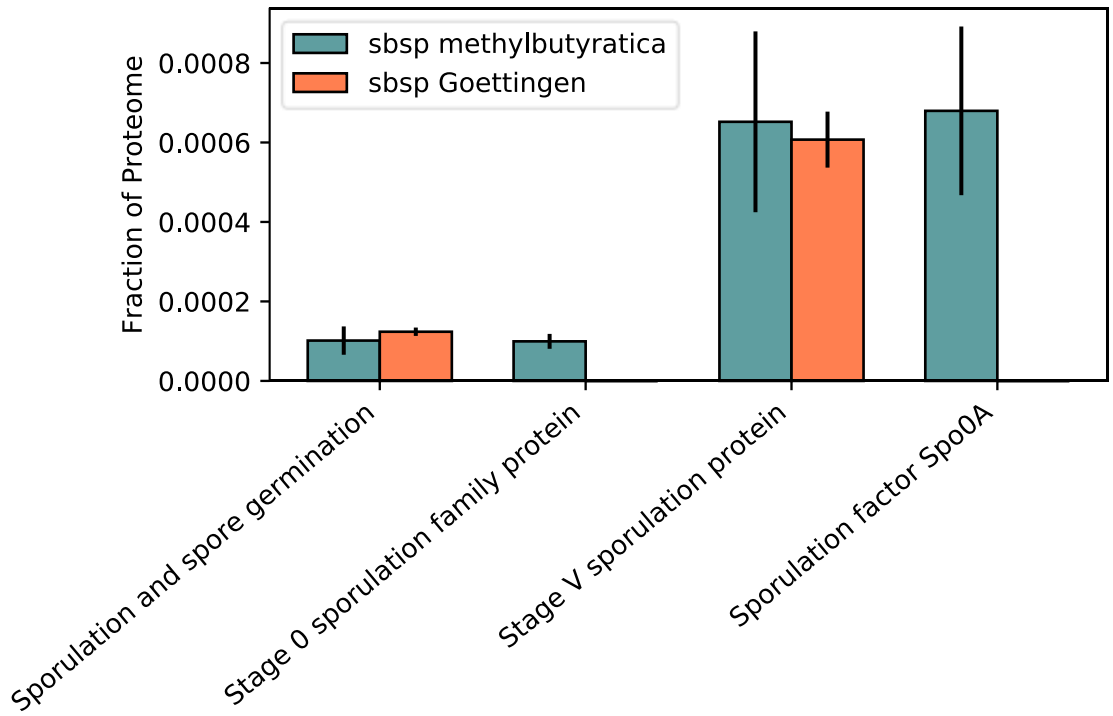
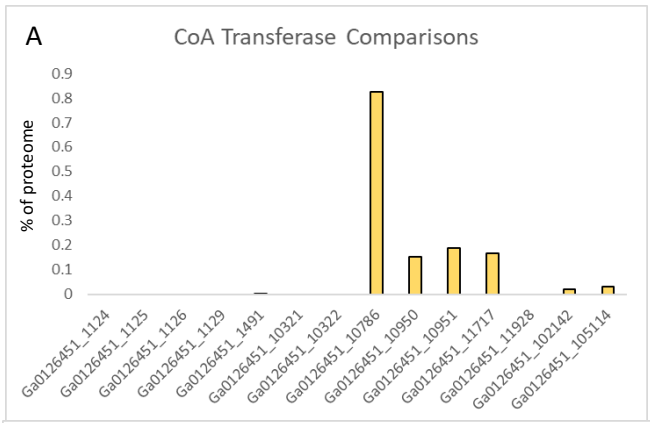
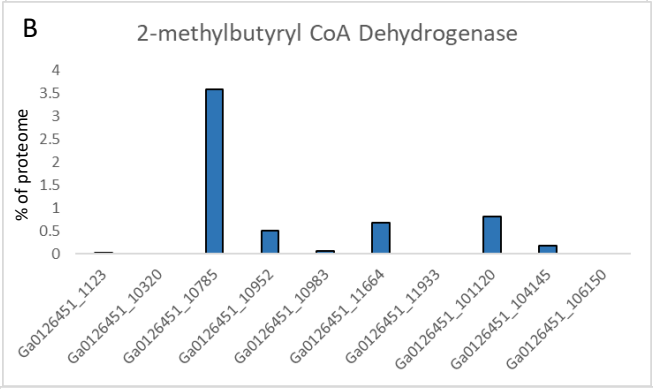
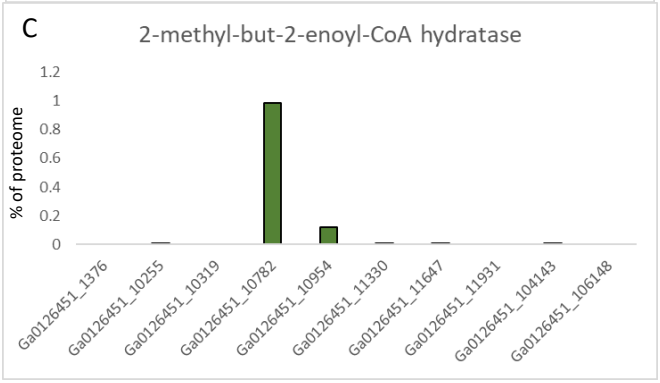
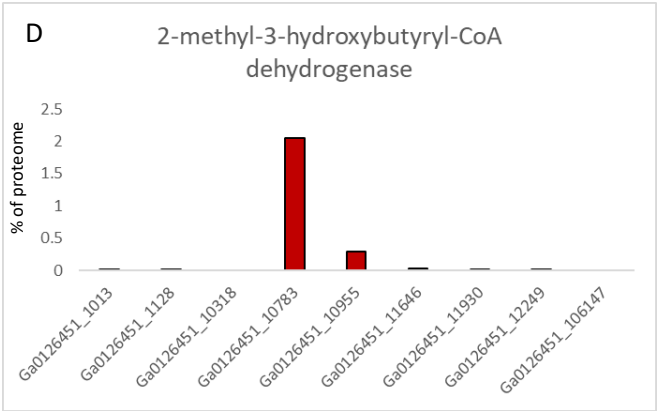


Figure 1. Sporulation related protein identified in sbsp *methylbutyratica*. Some proteins related to sporulation that were identified are shown in terms of fraction of the total proteome in both subspecies of *S. wolfei*.


A
CoA Transferase Comparisons
% of proteome
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Ga0126451_1124
Ga0126451_1125
Ga0126451_1126
Ga0126451_1129
Ga0126451_1491
Ga0126451_10321
Ga0126451_10322
Ga0126451_10786
Ga0126451_10950
Ga0126451_10951
Ga0126451_11717
Ga0126451_11928
Ga0126451_102142
Ga0126451_105114


B
2-methylbutyryl CoA Dehydrogenase
% of proteome
4
3.5
3
2.5
2
1.5
1
0.5
0
Ga0126451_1123
Ga0126451_10320
Ga0126451_10785
Ga0126451_10952
Ga0126451_10983
Ga0126451_11664
Ga0126451_11933
Ga0126451_101120
Ga0126451_104145
Ga0126451_106150


C
2-methyl-but-2-enoyl-CoA hydratase
% of proteome
1.2
1
0.8
0.6
0.4
0.2
0
Ga0126451_1376
Ga0126451_10255
Ga0126451_10319
Ga0126451_10782
Ga0126451_10954
Ga0126451_11330
Ga0126451_11647
Ga0126451_11931
Ga0126451_104143
Ga0126451_106148


D
2-methyl-3-hydroxybutyryl-CoA dehydrogenase
% of proteome
2.5
2
1.5
1
0.5
0
Ga0126451_1013
Ga0126451_1128
Ga0126451_10318
Ga0126451_10783
Ga0126451_10955
Ga0126451_11646
Ga0126451_11930
Ga0126451_12249
Ga0126451_106147

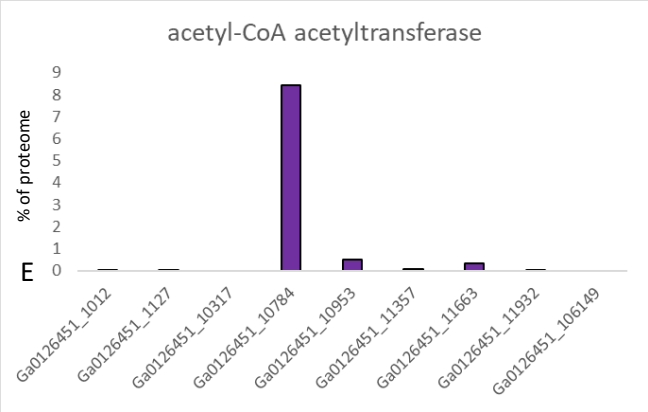
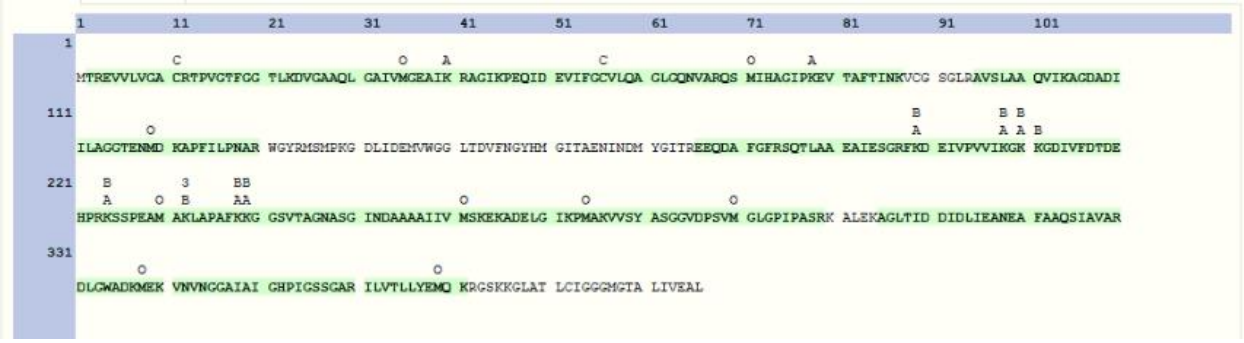


Figure 2. Abundant members of the 2-methylbutyrate degradation pathway in *S. wolfei* sbsp. *methylbutyratica*. The putative members of the 2-methylbutyrate degradation pathway were identified by gene loci. The relative abundances as a percentage of the proteome are displayed for each.

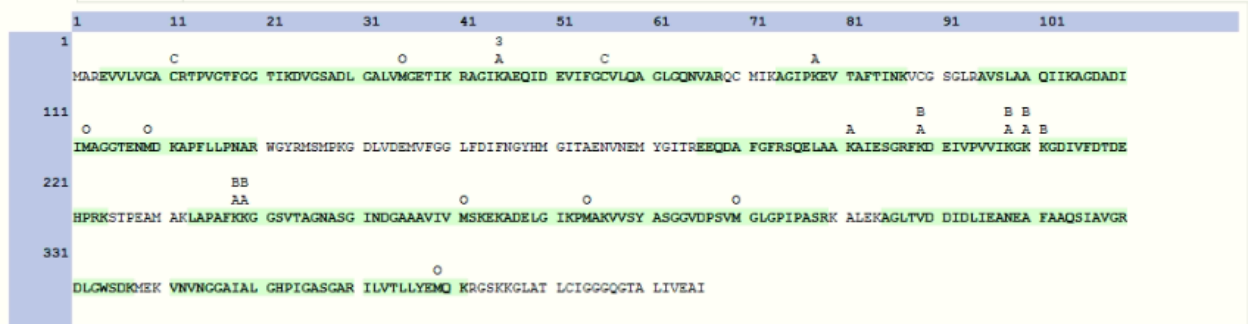
Modification	Count
Butyryl	18
Crotonyl	1
3-hydroxybutyryl	5
Acetyl	70

Table 1. Acylation count in sbsp. *methylbutyratica*. The number of modified proteins total for each acylation are presented.

Ga0126451_10784 acetyl-CoA acetyltransferase



Ga0126451_11663 acetyl-CoA acetyltransferase



Ga0126451_10783 2-methyl-3-hydroxyacyl-CoA dehydrogenase

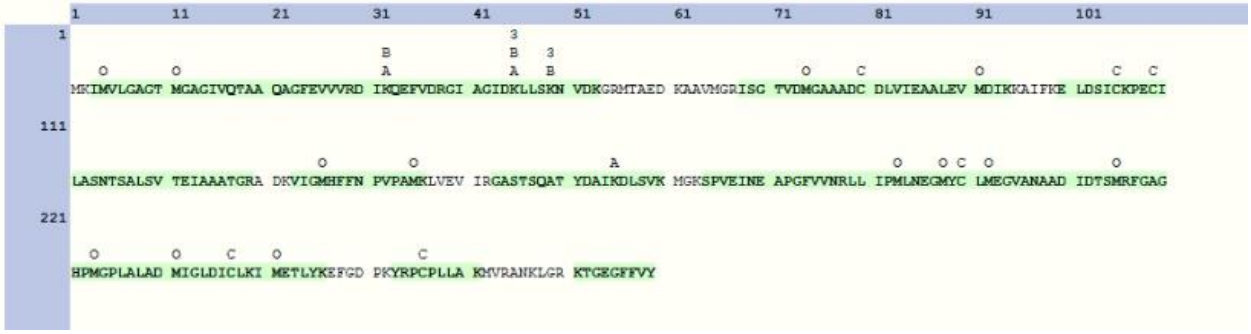


Figure 3. 2-methylbutyrate degradation pathway enzymes tend to be highly modified. A selection of the most abundant enzymes in the 2-methylbutyrate degradation pathway are displayed. Areas shaded in green have high-confidence peptides identified containing that region. The symbol above each site signifies which modification was found. A = acetyl, B = Butyryl, 3 = 3-hydroxybutyrylation, C = carbamidomethylation, O = oxidation. \

APPENDIX VII: *Syntrophomonas wolfei* tricultures demonstrate the acetylation of enzymes critical for *Methanosaeta concilii* (*Methanothrix soehngenii GP6*) acetoclastic methanogenesis

Note: For the purposes of this work, the methanogen of interest will be referred to by the name *Methanosaeta concilii*. There has been a push to revert the name back, *Methanothrix soehngenii*, though it is unclear if and when such a nomenclature change will take place.

Syntrophic bacteria require methanogenic partners to grow due to thermodynamic constraints on some of the core pathways in their metabolism (1). *Syntrophomonas wolfei* is typically coupled with the hydrogen scavenger *Methanospirillum hungatei* in order to overcome these constrains (2, 3). However, the degradation of short-chain fatty acids by *S. wolfei* produces not only an excess of hydrogen, but acetate as well. Acetoclastic methanogens, such as *Methanosaeta concilii*, may consume the acetate produced by the syntrophic bacteria to in turn produce methane (4). The first step of this process is analogous to the steps taken in syntrophs whereby acetate is activated on a Coenzyme A (CoA) scaffold. As discussed in depth in Chapters 1-4, these intermediates are reactive and may cause acetylation of lysine residue.

In an attempt to better understand the population dynamics of microbial consortia, *S. wolfei* cells were grown in the presence of their typical methanogenic partners, *M. hungatei*, as well as the acetoclastic methanogens *M. concilii*. Cells from this triculture were harvested and prepared via enhanced filter-assisted sample preparation method (eFASP) (5, 6) and run on a QExactive mass spectrometer (Thermo Scientific) using the same method discussed in Appendix VI. Spectra obtained were searched using the Mascot algorithm against a database of all three

microbial proteomes and included in the dynamic modification of acetylation (+42.01056) on lysine in the search. The results of what were found in the *M. concilii* proteome are displayed in **Table 1**. Notable acetate-CoA ligase is the enzyme that catalyzes first step of the acetoclastic methanogenesis pathway (4). The resulting acetyl-CoA is then converted by the carbon monoxide dehydrogenase (cdh)/acetyl-CoA decarbonylase complex, tetrahydromethanopterin S-methyltransferase, and methyl coenzyme M reductase to ultimately produce methane (7). All four of these enzymes are shown to have acetylation sites on in the *M. concilii* proteome (**Table 1**). When coupled with data presented in Chapter 4, it demonstrates that every enzyme involved in the pathway that degrades butyrate in syntrophs, to the methane released by methanogens, contain the presence of acyl-lysine modifications (**Figure 1**). This opens up the possibility of a regulatory mechanism and paves the way for future functional studies of the modifications on these enzymes.

Citations

1. Schink, B. (1997) Energetics of syntrophic cooperation in methanogenic degradation. *Microbiol. Mol. Biol. Rev.* **61**, 262–80

2. McInerney, M. J., Bryant, M. P., and Pfennig, N. (1979) Anaerobic bacterium that degrades fatty acids in syntrophic association with methanogens. *Arch. Microbiol.* **122**, 129–135

3. McInerney, M. J., Bryant, M. P., Hespell, R. B., and Costerton, J. W. (1981) Syntrophomonas wolfei gen. nov. sp. nov., an Anaerobic, Syntrophic, Fatty Acid-Oxidizing Bacterium. *Appl. Environ. Microbiol.* **41**, 1029–39

4. Stams, A. J. M., Teusink, B., and Sousa, D. Z. (2019) Ecophysiology of Acetoclastic Methanogens. in *Biogenesis of Hydrocarbons*, pp. 109–121, Springer International Publishing, 10.1007/978-3-319-78108-2_21

5. Erde, J., Loo, R. R. O., and Loo, J. A. (2017) Improving proteome coverage and sample recovery with enhanced FASP (eFASP) for quantitative proteomic experiments. in *Methods in Molecular Biology*, pp. 11–18, Humana Press Inc., **1550**, 11–18

6. Erde, J., Loo, R. R. O., and Loo, J. A. (2014) Enhanced FASP (eFASP) to increase proteome coverage and sample recovery for quantitative proteomic experiments. *J. Proteome Res.* **13**, 1885–1895

7. Carr, S. A., Schubotz, F., Dunbar, R. B., Mills, C. T., Dias, R., Summons, R. E., and Mandernack, K. W. (2018) Acetoclastic Methanosaeta are dominant methanogens in organic-rich Antarctic marine sediments. *ISME J.* **12**, 330–342

Peptide Sequence	Gene Tag	Gene description	Site (Modification)
LANALKSVGVEK	MCON_2868	acetate--CoA ligase	K6(Acetyl)
SVGVEKGDR	MCON_2868	acetate--CoA ligase	K6(Acetyl)
KVIVVKR	MCON_2868	acetate--CoA ligase	K6(Acetyl)
TAPSVQKVVVVK	MCON_0558	acetate--CoA ligase	K7(Acetyl)
GAGVVKGDR	MCON_0558	acetate--CoA ligase	K6(Acetyl)
KQMISAIK	MCON_2932	carboxymuconolactone decarboxylase family protein	K1(Acetyl); M3(Oxidation)
AIAAIVNKGK	MCON_3321	class I fructose-bisphosphate aldolase family protein	K8(Acetyl)
VAKGGANAVLQQK	MCON_3321	class I fructose-bisphosphate aldolase family protein	K3(Acetyl)
GKNADEAIK	MCON_3321	class I fructose-bisphosphate aldolase family protein	K2(Acetyl)
GHVVKLGGEK	MCON_1326	CO dehydrogenase/acetyl-CoA synthase subunit delta	K5(Acetyl)
ADKITLSDLNK	MCON_1326	CO dehydrogenase/acetyl-CoA synthase subunit delta	K3(Acetyl)
IATEDDAKDIAALK	MCON_1330	CO dehydrogenase/CO-methylating acetyl-CoA synthase complex subunit beta	K8(Acetyl)
SDKIDLYSDR	MCON_0762	coenzyme-B sulfoethylthiotransferase subunit beta	K3(Acetyl)
TVAVNLAGVEGALKTGK	MCON_0762	coenzyme-B sulfoethylthiotransferase subunit beta	K14(Acetyl)
DQVGIPLDKK	MCON_0760	coenzyme-B sulfoethylthiotransferase subunit gamma	K9(Acetyl)
GSKDEFDTAR	MCON_0940	hypothetical protein	K3(Acetyl)
KIGLVGIR	MCON_0394	methanogenesis marker 3 protein	K1(Acetyl)
SDIGKIILVGGPTR	MCON_1037	molecular chaperone DnaK	K5(Acetyl)
ALEALQVVDKNK	MCON_2613	peroxiredoxin	K10(Acetyl)
FKEAGITPEDIK	MCON_0819	phenylacetate--CoA ligase	K2(Acetyl)
SELEELQLKR	MCON_2332	phenylacetate--CoA ligase	K9(Acetyl)
KATSILK	MCON_2883	phosphoenolpyruvate synthase	K1(Acetyl)
QKPVPIATASK	MCON_0501	pyridoxamine 5'-phosphate oxidase family protein	K2(Acetyl)
KQEVFDFGK	MCON_1068	tetrahydromethanopterin S-methyltransferase subunit H	K1(Acetyl)
LLKMVGYK	MCON_1649	translation elongation factor EF-1 subunit alpha	K3(Acetyl); M4(Oxidation)
ILIATDGSEKSK	MCON_2642	universal stress protein	K10(Acetyl)
VLSGSGKPR	MCON_2515	V-type ATP synthase subunit B	K7(Acetyl)

AVLIALKK	MCON_2514	V-type ATP synthase subunit D	K7(Acetyl)
VVPDLKLTIK	MCON_2514	V-type ATP synthase subunit D	K6(Acetyl)

Table 1. Acetylated peptides in *M. concilii*. Peptides identified containing acetylation in *M. concilii* when grown in syntrophic tricultures.

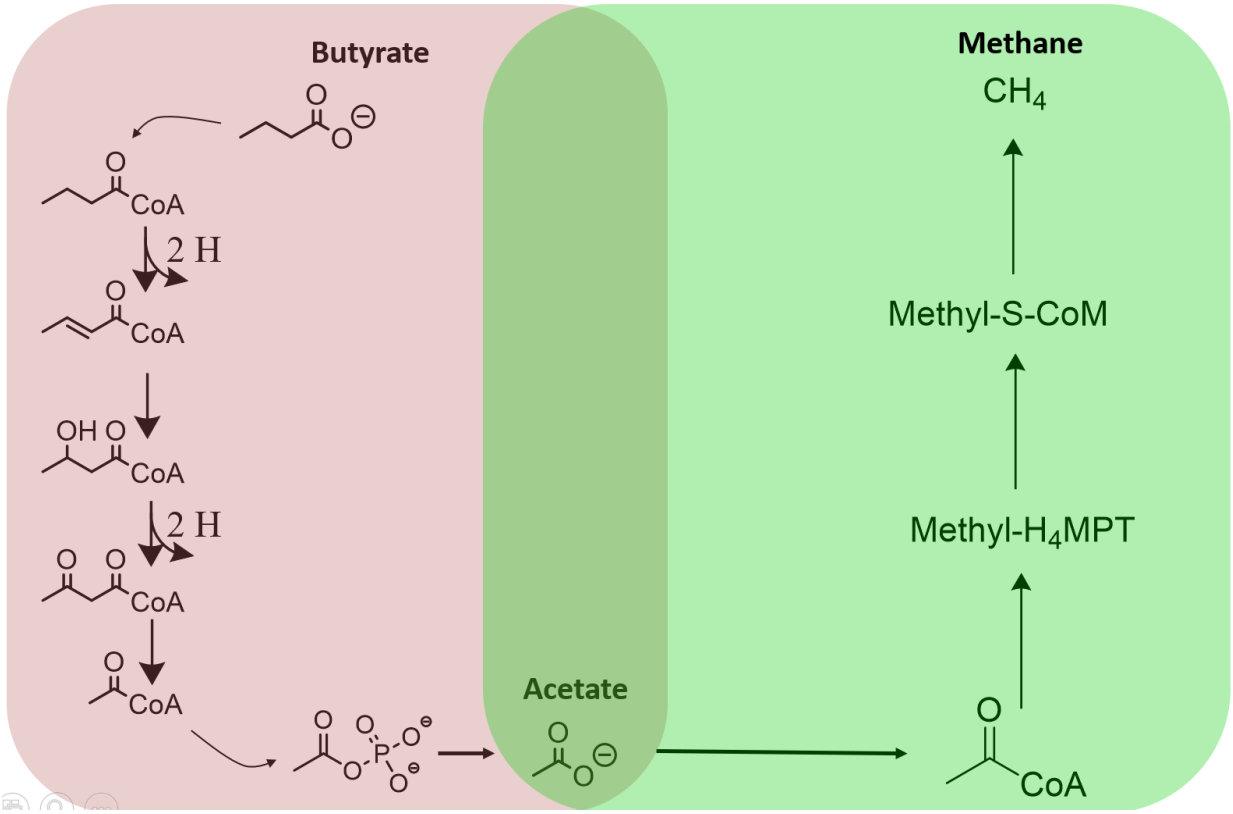


Figure 1. Pathway from butyrate to methane. The pathway that shows the degradation of butyrate to acetate in *S. wolfei* (red) followed by the acetoclastic methanogenesis that occurs in the *M. concilii* methanogen (green). Each arrow signifies an enzyme mediated reaction in the pathway. Each enzyme reaction in the pathway displays at least one acylation on a lysine residue.

APPENDIX VIII: *Methanosaeta concilii* (*Methanothrix soehngenii GP6*) has a remarkably resilient proteinaceous sheath

Note: For the purposes of this work, the methanogen of interest will be referred to by the name *Methanosaeta concilii*. There has been a push to revert the name back, *Methanothrix soehngenii*, though it is unclear if and when such a nomenclature change will take place.

INTRODUCTION

The cell walls and membranes of archaea show a diverse array of biochemical properties (1), unsurprising given the diverse and often extreme environments in which these microbes inhabit. Several species of methanogenic archaea have demonstrated an outer proteinaceous layer surrounding the cell membrane forming filamentous cells, termed a sheath (2, 3). Despite previous attempts to biochemically assess these superstructures, the first protein identity was not identified until 2015 when the *Methanosaeta thermophila* sheath was disassembled, analyzed, and shown to possibly be a functional amyloid speculated to have multiple inter-protein disulfide bridges that result in increased stability (4). Subsequent genomic and biochemical analysis of *Methanospirillum hungatei* demonstrated that this species contained a homolog of major sheath protein A (MspA) and that, indeed, it's properties reflected that of an amyloid protein (5, 6). Previous work with the *M. hungatei* sheath protein demonstrated that they could not be dissolved with standard harsh denaturants, including 6M urea, 6M guanidinium chloride, 10M lithium thiocyanate hydrate, and 1% sodium dodecyl sulfate (7). The extreme resilience of these proteins makes their analysis uniquely challenging. Interestingly, while the *M. hungatei* genome had a homolog of the *M. thermophila* protein MspA, the genome of *Methanosaeta concilii* does not

despite the presence of another sheath superstructure around the *M. concilii* cells (2). It has previously been determined that the *M. concilii* sheath is more stable than that of *M. hungatei,* and that there are many more glycans present in the *M. concilii* sheath (8). Here, we sought the identity of this protein as well as the possibility of modifications that may be present on the proteins.

METHODS

Empty sheaths, often referred to as "ghosts", were obtained from the Gunsalus lab at UCLA. Briefly, the cells were grown on minimal media supplemented with acetate. Cells then mechanically sheared in a bead beater and washed with a solution of sodium dodecyl sulfate to remove soluble protein that are present in the system.

Samples were prepared using a variety of techniques. Similar to previous reports for *M. hungatei* (7), the *M. concilii* sheath samples were not solubilized using 8M urea, 6M guanidinium chloride or hexafluoroisopropanol (**Figure 1**). A successful solubilization and digestion of sheath proteins was eventually identified.

Sheath proteins were reduced in 66mM dithiothreitol (DTT) for 1 hour at 60°C. Samples were then dried down and resuspended in 100% formic acid of 30 minutes at 25°C. Samples were once again dried down and resuspended in 100mM ammonium bicarbonate. An ethyl acetate extraction was then performed and repeated three times to further purify the protein components of the sheath. Once purified, the proteins were reduced with 10mM DTT and alkylated with 50mM of iodoacetamide. 200ng of chymotrypsin was added to the sheath to digest the proteinaceous components.

Samples were separated on an EASY nLC-1200 (Thermo Scientific) before being loaded onto the QExactive orbitrap mass spectrometer (Thermo Scientific). For tandem mass spectrometry,

ions were selected for fragmentation using data-dependent acquisition mode where the top 10 most abundant ions were selected. Precursor ions were fragmented using higher-energy collisional dissociation. For protein identification, *.RAW files were searched using ProteomeDiscoverer (version 1.4) with the Mascot database search algorithm (Matrix Science).

RESULTS AND DISCUSSION

When the successful solubilization method was used on the sheath obtained for *M. hungatei*, the bulk of the peptide signal identified (~90%) came from the MspA protein (**Figure 2A**). Based on the same analysis several similar proteins were identified as predominant in the *M concilii* sheath protein (**Figure 2B**). The two most abundant components identified were Mcon_1139 and Mcon_1134 which combined accounted for ~20% of the proteins identified, but were the only proteins identified in greater abundance than the S-layer protein. These proteins share 78.39% identity, perhaps indicating that they are part of a family. Using BLAST, another possible homolog was identified that would be a part of the putative family; Mcon_1141. Notably, experiments using elastase instead of chymotrypsin also found that the three most abundant proteins were Mcon_1141, Mcon_1139, and Mcon_1134. To better identify if these were likely to be exported proteins, the sequences were searched for signal peptides that indicate the proteins are exported across the membranes. All three of the homologs shared a signaling peptide at the N-terminal (**Figure 3A-C**), based on analysis by SignalP-5.0 (9). Worth noting, all members of this family also share a similar trait of a cysteine present at the C-terminal or C-penultimate residue. The C-terminal peptide was not identified in the LC-MS/MS experiment, leaving open the possibility of C-terminal modification. Also of note is that the sequences contain several N-X-S/T/C motifs, which are motifs that are indicative of glycosylation occurring at the sites (10, 11). These regions are unidentified in our samples, leaving open the

possibility of glycan modifications. Indeed, previous biochemical characterization of the sheath protein demonstrate a very high abundance of glycans relative to *M. hungatei*, particularly mannose (8). While peptides containing this type of modification were not identified, that showed evidence of chains containing hexoses present in the LC-MS/MS runs (**Figure 4**). The *M. concilii* genome also contains a putative oligosaccharide transferase in the protein Mcon_1133, giving more evidence of the likelihood of protein glycosylation.

CONCLUSION

The extreme and diverse structures that compose the archaea surfaces are worthy of study are necessary given the extreme and diverse environments in which these microbes inhabit. Understanding the means by which these microbes' interface with their environment would help us understand their physiology in a much deeper level. Proteinaceous sheaths appear to be a somewhat rare phenomenon, and that of *M. concilii* appear to be especially unique. Having a full characterization of the proteins are involved in this structure would not only contribute to the understanding of proteinaceous biomaterials, but also how it can interface with the environment they inhabit, which commonly involves keeping out unwanted environmental factors, and in the case of *M. concilii* may have a functional role in interspecies electron transportation (12).

CITATIONS

1. Albers, S. V., and Meyer, B. H. (2011) The archaeal cell envelope. *Nat. Rev. Microbiol.* **9**, 414–426
2. Beveridge, T. J., Patel, G. B., Harris, B. J., and Sprott, G. D. (1986) The ultrastructure of Methanothrix concilii, a mesophilic aceticlastic methanogen. *Can. J. Microbiol.* **32**, 703–710
3. Zeikus, J. G., and Bowen, V. G. (1975) Fine structure of Methnospirillum hungatii. *J. Bacteriol.* **121**, 373–380
4. Dueholm, M. S., Larsen, P., Finster, K., Stenvang, M. R., Christiansen, G., Vad, B. S., Bøggild, A., Otzen, D. E., and Nielsen, P. H. (2015) The tubular sheaths encasing Methanosaeta thermophila filaments are functional amyloids. *J. Biol. Chem.* **290**, 20590–20600
5. Gunsalus, R. P., Cook, L. E., Crable, B., Rohlin, L., McDonald, E., Mouttaki, H., Sieber, J. R., Poweleit, N., Zhou, H., Lapidus, A. L., Daligault, H. E., Land, M., Gilna, P., Ivanova, N., Kyrpides, N., Culley, D. E., and McInerney, M. J. (2016) Complete genome sequence of Methanospirillum hungatei type strain JF1. *Stand. Genomic Sci.* **11**, 2
6. Christensen, L. F. B., Hansen, L. M., Finster, K., Christiansen, G., Nielsen, P. H., Otzen, D. E., and Dueholm, M. S. (2018) The Sheaths of Methanospirillum Are Made of a New Type of Amyloid Protein. *Front. Microbiol.* **9**, 2729
7. Beveridge, T. J., Stewart, M., Doyle, R. J., and Sprott, G. D. (1985) Unusual stability of the Methanospirillum hungatei sheath. *J. Bacteriol.*
8. Patel, G. B., Sprott, G. D., Humphrey, R. W., and Beveridge, T. J. (1986) Comparative analyses of the sheath structures of Methanothrix concilii GP6 and Methanospirillum

hungatei strains GP1 and JF1. *Can. J. Microbiol.* **32**, 623–631

9. Almagro Armenteros, J. J., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., von Heijne, G., and Nielsen, H. (2019) SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* **37**, 420–423
10. Welply, J. K., Shenbagamurthi, P., Lennarz, W. J., and Naider, F. (1983) Substrate recognition by oligosaccharyltransferase. Studies on glycosylation of modified Asn-X-Thr/Ser tripeptides. *J. Biol. Chem.* **258**, 11856–11863
11. Lowenthal, M. S., Davis, K. S., Formolo, T., Kilpatrick, L. E., and Phinney, K. W. (2016) Identification of Novel N-Glycosylation Sites at Noncanonical Protein Consensus Motifs. *J. Proteome Res.* **15**, 2087–2101
12. Morita, M., Malvankar, N. S., Franks, A. E., Summers, Z. M., Giloteaux, L., Rotaru, A. E., Rotaru, C., and Lovleya, D. R. (2011) Potential for direct interspecies electron transfer in methanogenic wastewater digester aggregates. *MBio*. 10.1128/mBio.00159-11

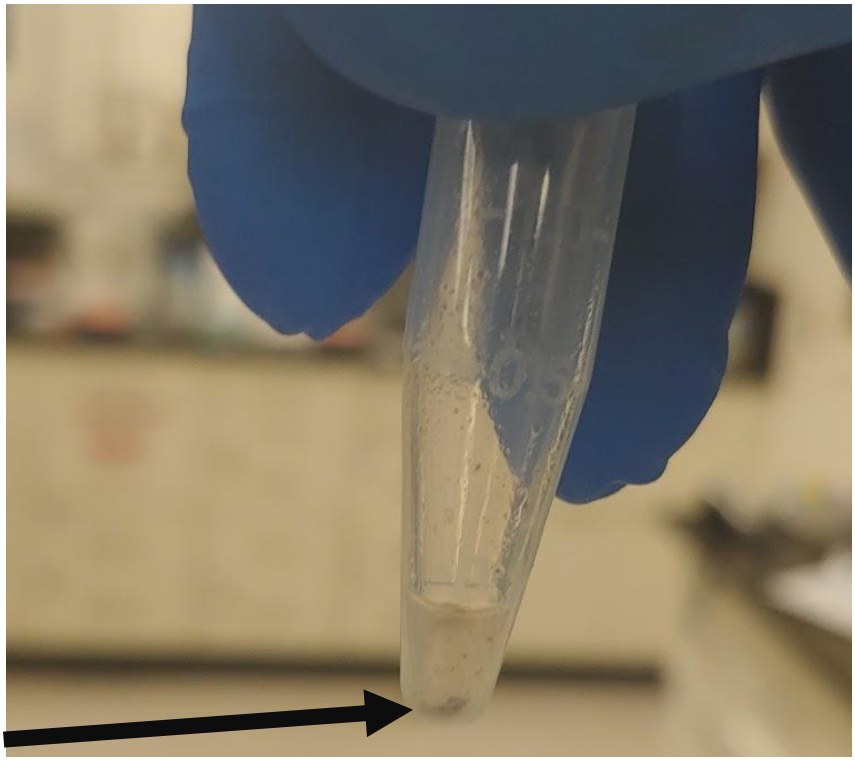


Figure 1. A demonstration of the insolubility of sheath protein. Many solutions commonly used in dissolving and denaturing proteins have proven to be unable to dissolve the sheath protein. The black arrow points to the sheath protein precipitate floating in a solution of HFIP.

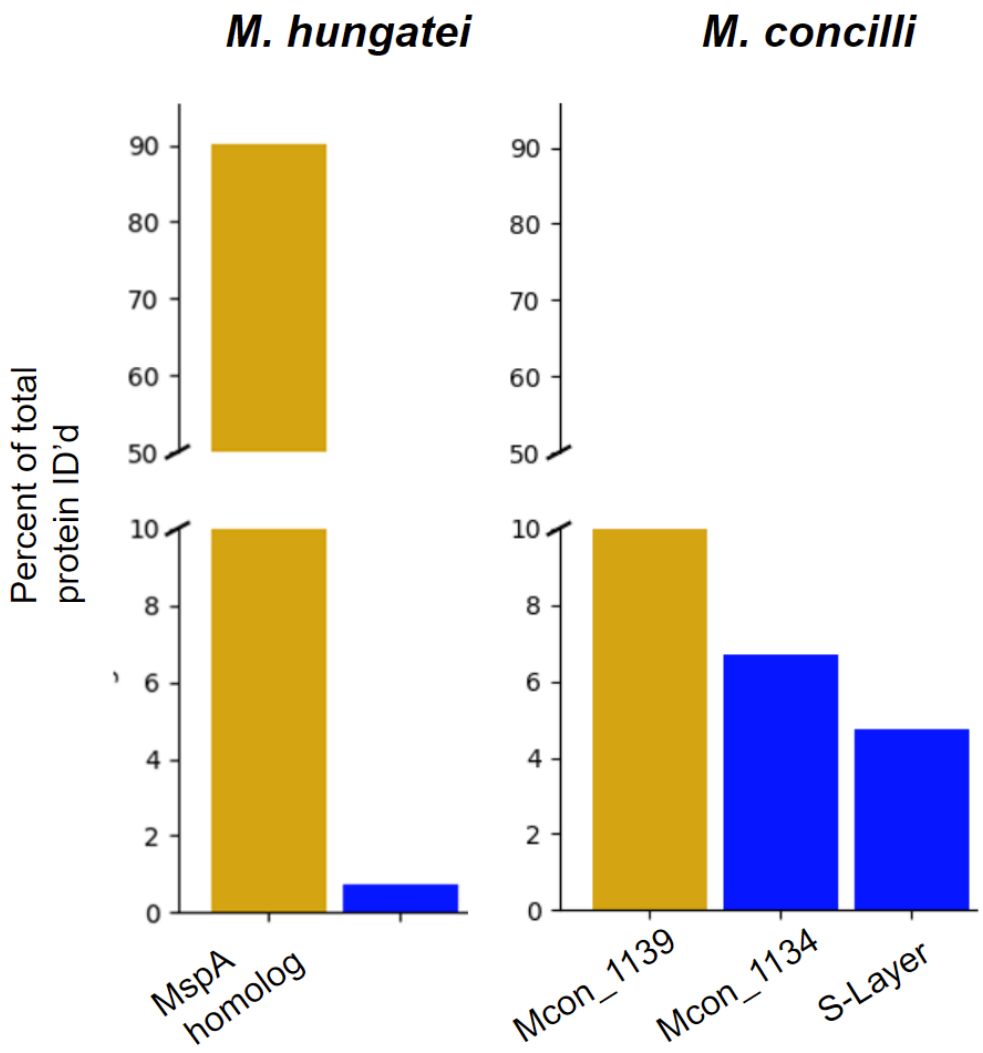


Figure 2. Ghost proteins by abundance. Relative abundance of proteins identified upon solubilization and digestion by chymotrypsin. The top protein identified is shown in gold, all other are depicted in blue.

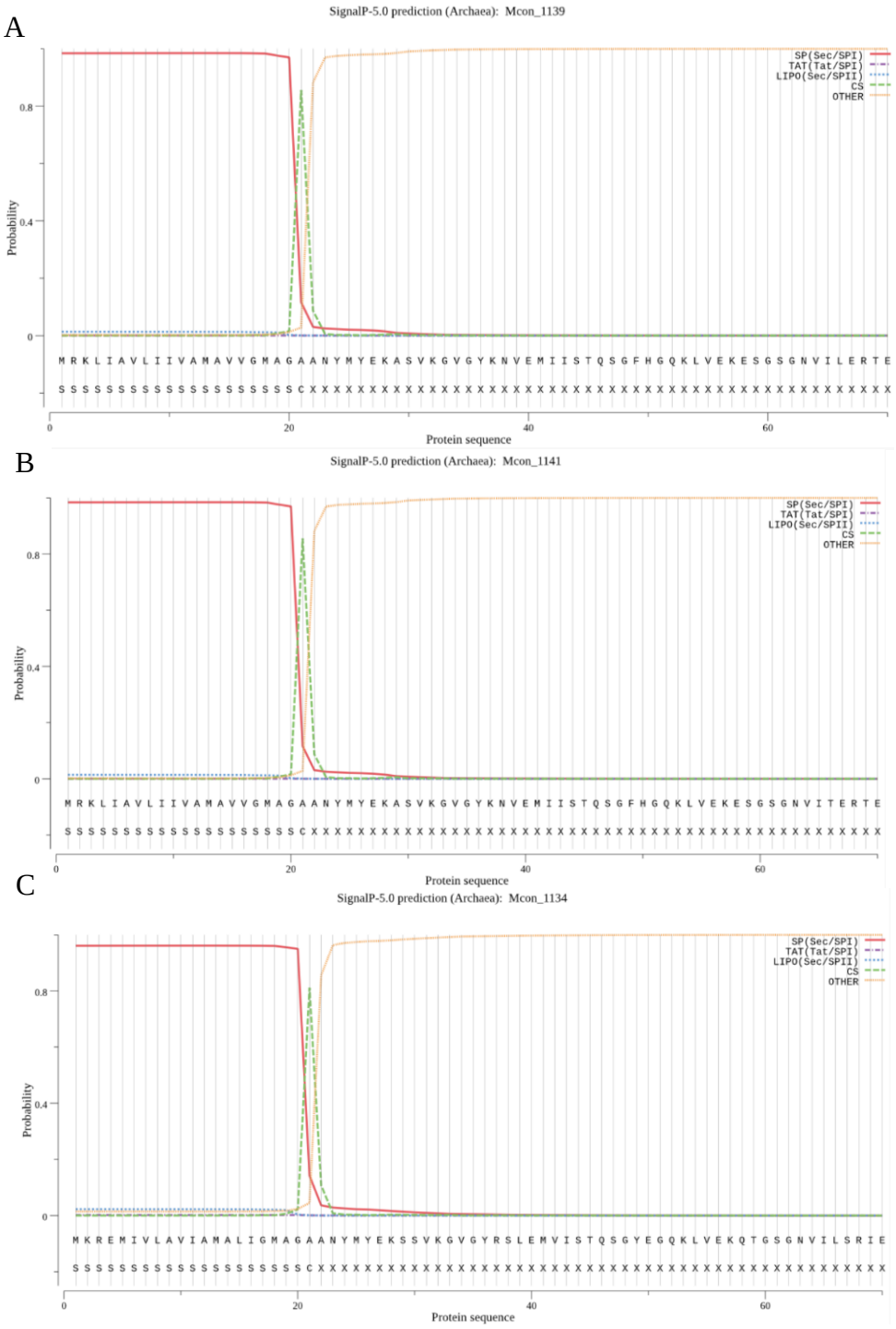
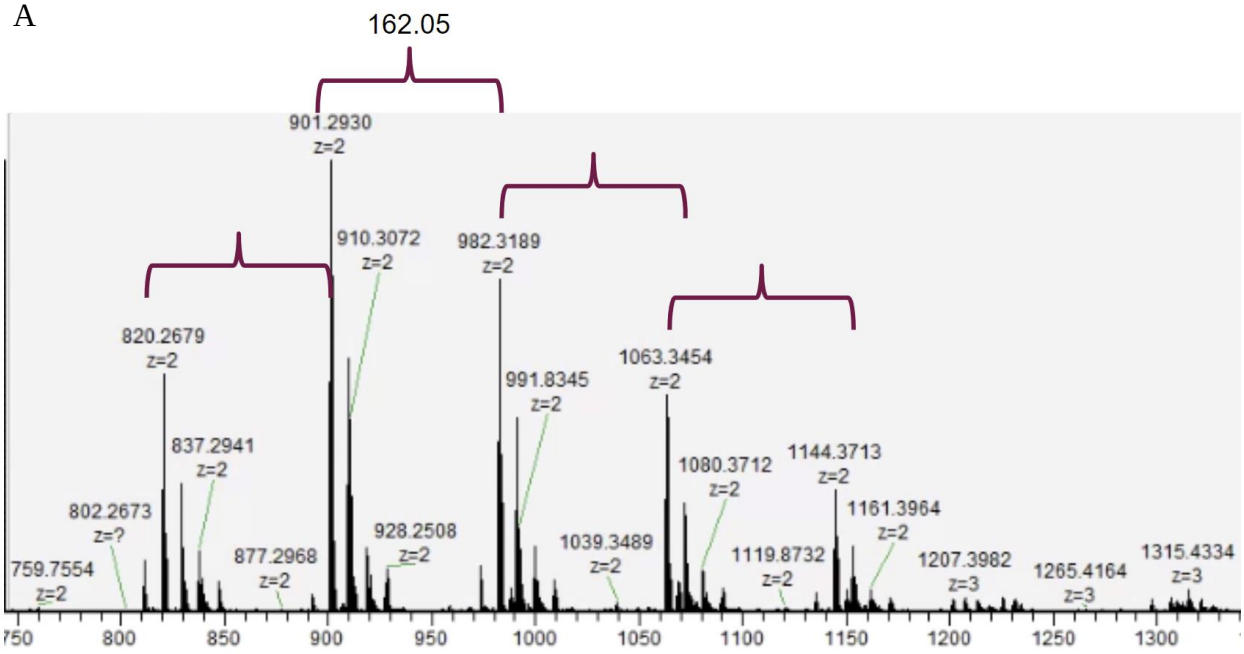


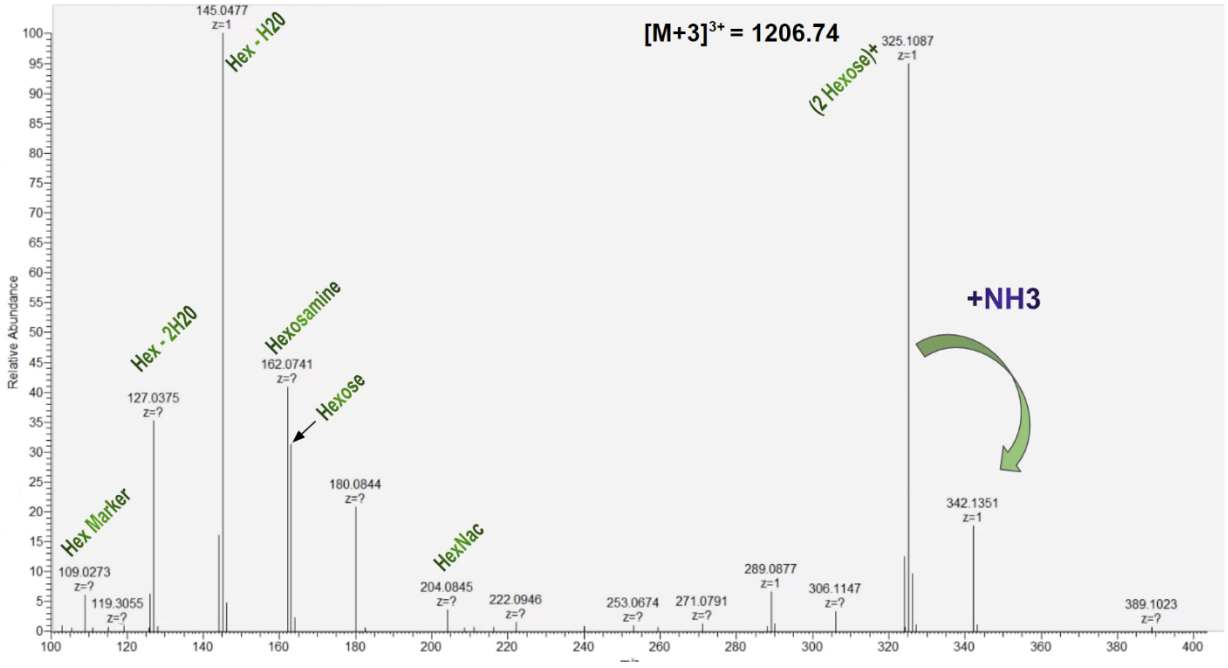
Figure 3. N-terminal signaling sequence. Proteins identified as a high abundance family of protein, (A) Mcon _1139, (B) Mcon_1141, and (C) Mcon_1134 are subject to predictive algorithm SignalP5.0 to identify possible N-terminal signal peptide. The red line corresponds to

the probability of a sequence as a part of the signal peptide and the green line indicates the probability of residue being part of a cleavage site.

A


162.05
901.2930
z=2
910.3072
z=2
982.3189
z=2
820.2679
z=2
991.8345
z=2
1063.3454
z=2
837.2941
z=2
1080.3712
z=2
1144.3713
z=2
802.2673
z=?
928.2508
z=2
1161.3964
z=2
759.7554
z=2
877.2968
z=2
1039.3489
z=2
1119.8732
z=2
1207.3982
z=3
1265.4164
z=3
1315.4334
z=3
750
800
850
900
950
1000
1050
1100
1150
1200
1250
1300

B


[M+3]3+ = 1206.74
145.0477
z=1
Hex - H20
325.1087
z=1
(2 Hexose)+
+NH3
Hex - 2H20
Hexosamine
162.0741
z=?
Hexose
127.0375
z=?
180.0844
z=?
Hex Marker
HexNac
109.0273
z=?
119.3055
z=?
204.0845
z=?
222.0946
z=?
253.0674
z=?
271.0791
z=?
289.0877
z=1
306.1147
z=?
342.1351
z=1
389.1023
z=?
Relative Abundance

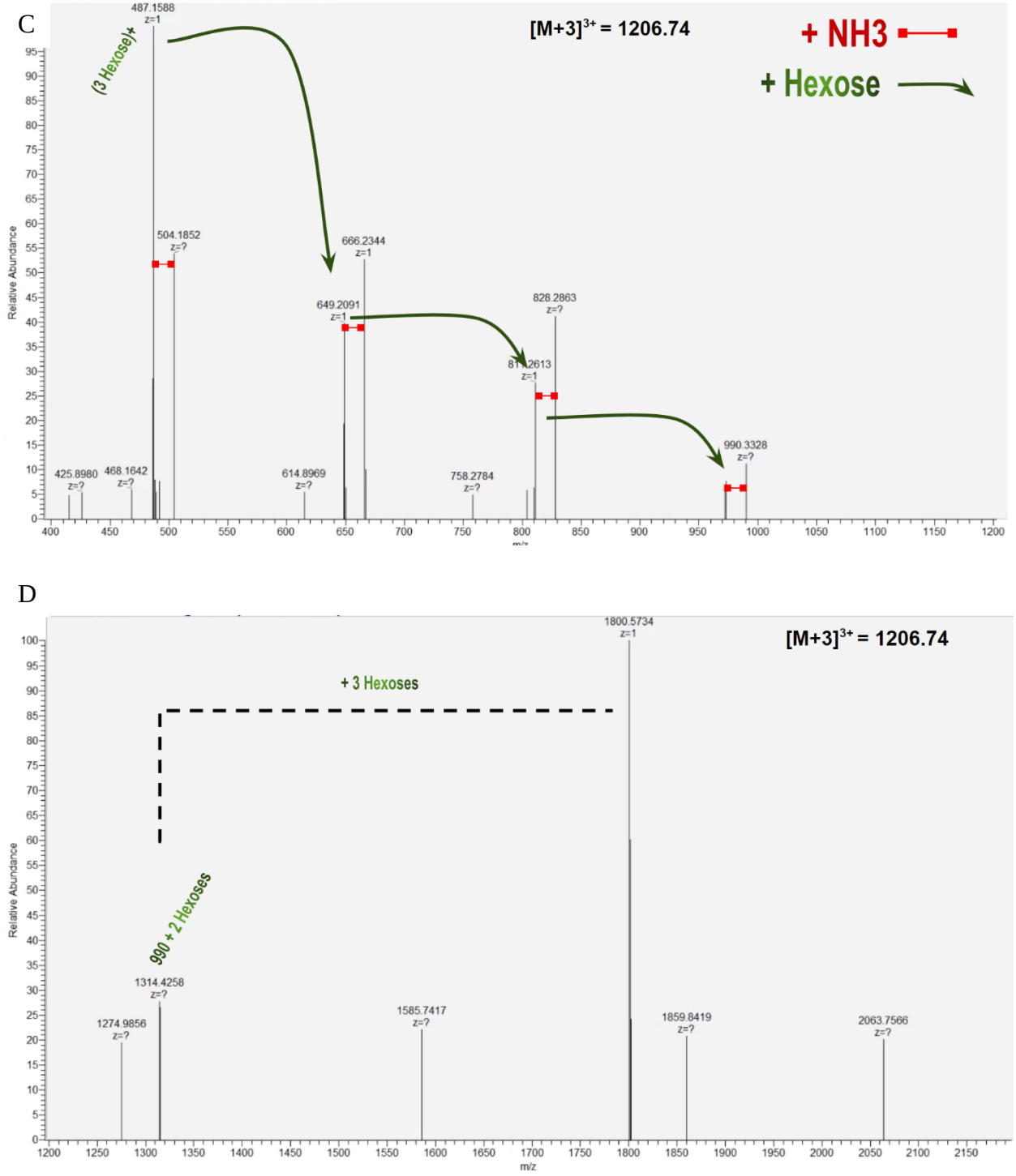


Figure 4. Evidence of glycans in *M. concilii* sheath samples. (**A**) MS^1 spectra of species found when running the sheath on LC-MS/MS. A monoisotopic mass shit of 162.05 is typically associated with hexoses. (**B**-**C**) Fragment spectra from one of the putative glycan precursors

selected at *m/z* 1206.74. Annotated spectra of the *m/z* 100 - 400 range (**B**), *m/z* 400 – 1200 (**C**), and *m/z* 1200 – 2200 (**D**).