**Title**
Elucidating Syntrophic Metabolism through Noncanonical Protein Modifications and Protein Sequences



**Author**
Nguyen, Hanh Thi Hong

**Publication Date**
2015

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

# Elucidating Syntrophic Metabolism through Noncanonical Protein Modifications and Protein Sequences

A dissertation submitted in partial satisfaction

of the requirements for the degree Doctor of Philosophy

in Biochemistry, Molecular and Structural Biology

by

Hanh Thi Hong Nguyen

2015

# ABSTRACT OF THE DISSERTATION

## Elucidating Syntrophic Metabolism through Noncanonical Protein Modifications and Protein Sequences

by

Hanh Thi Hong Nguyen

Doctor of Philosophy in Biochemistry, Molecular and Structural Biology

University of California, Los Angeles, 2015

Professor Joseph Ambrose Loo, Chair

Carbon is recycled in the anaerobic environment by fatty acid and aromatic compound-degrading syntrophic bacteria. These linchpins to sustaining life on earth survive at the limits of thermodynamic feasibility. They use two-thirds of the generated ATP in substrate degradation, leaving one-third (1/3 mole ATP per mole substrate, or a minimum energy quantum) to sustain microbial life. Despite their importance to biogeochemical cycles and alternative energy, syntrophic bacteria remain poorly understood.

With energy so limited, how do they adjust their metabolism to environmental changes or altered nutrient availability? Complex “proteoforms” generated by integrating protein sequence variants and/or protein modifications could hold the key to their metabolic regulation. Studying the proteomes of *Syntrophus aciditrophicus* and *Syntrophomonas wolfei*, two model syntrophic bacteria, we discovered many protein sequence variants, including fusion forms of adjacent proteins on the same operons. Many protein modifications were also observed. Many are known,

many are first recorded, and many belong to the rising family of lysine acylations, which has emerged as an important regulatory mechanism coordinating metabolism in response to changes in nutrient availability and metabolic fluxes.

The dissertation of Hanh Thi Hong Nguyen is approved.

Robert T. Clubb

Robert P. Gunsalus

Joseph Ambrose Loo, Committee Chair

University of California, Los Angeles

2015

# TABLE OF CONTENTS

# ACKNOWLEDGMENTS

First, I thank my graduate advisors, Drs. Joseph A. Loo and Rachel R. Ogorzalek Loo, for always being kind, patient, available and supportive. Transitioning from a molecular biology undergraduate background to mass spectrometry seasoned with bioinformatics, I met challenging yet rewarding learning experiences. Rachel and previous lab members - Eric Pang, Jonathan Erde, Raji Lakshmanan and others - greatly helped in bridging the gaps! Joe provided invaluable mass spectrometry training opportunities through attending American Society for Mass Spectrometry (ASMS) conferences and with companies making mass spectrometers such as Waters and Thermo. Six years have come by fast with fun lab members! I always look forward to lab (re)unions with past and current lab members at the annual ASMS conference. My mind still remembers the days of me laughing it off when the HPLC instrument required a thorough systematic troubleshooting plan or with my happy days when sharing cool pieces of data with others. Thanks to my committee members for helpful advice and critical evaluation of my work. I will miss the friendly people, helpful staff and happy night workers here!

Last, I want to thank my parents, Dien Nguyen and Mui Hoang, for always cheering, caring and supporting from across the North Pacific Ocean. Despite being far apart, my siblings Vinh and Hien are always there to give me inspiration and support whenever needed. Thanks to my boyfriend Chris for all the love, help, support and valuable advice. To my cat Sharpie, thank you for making the place home.

# PREFACE

The work started with 2D-gel proteomic data on the Gram-negative syntrophic bacterium *Syntrophus aciditrophicus*. Dr. Rachel R. Ogorzalek Loo detected many examples of sequence variations and protein modifications in the proteomic data of this organism. I followed it up with a focus on the lysine acylation family, which we discovered to be widely abundant in syntrophic bacteria.

Chapter 2 shows collective evidence that *S. aciditrophicus*, a model syntrophic bacterium digesting aromatic compounds and fatty acids, bears many protein sequence variances and modifications. The data include analysis of 2D-gel proteomic data generated by Dr. Yanan Yang in 2005 and of RNA-seq data for *S. aciditrophicus* provided by Professor Robert P. Gunsalus. Cell pellets were shipped frozen from the lab of Professor Michael J. McInerney at the University of Oklahoma, Norman. Recombinant Swol_2052, SYN_42 and SYN_1020 were expressed by the UCLA DOE Protein Expression Facility.

Chapters 3 and 4 applied mass spectrometric methods (with and without immune-enrichment for acylated peptides) to study lysine acylations in *Syntrophomonas wolfei* and *Syntrophus aciditrophicus.* Placing lysine acylation information into the big picture of the cellular proteome, we gain more insights into cellular metabolism regulation and energy distribution in organisms producing only a small quantum amount of energy to support life. Professor Yingming Zhao from the University of Chicago provided us with anti-acyllysine antibodies and good advice during our first phase of the acyllysine discoveries. Jon Warren, an undergraduate at UCLA, and Phuong Nguyen, a visiting graduate student from Vietnam through the UCLA Center Global Mentoring Program, provided ample help in setting up and configuring the experimental conditions.

Chapter 5 was done with Dr. Huilin Li from our group to find a better method for peak matching and modification localization in top-down proteomics data. One of its usages could be to characterize syntrophic bacteria protein complexes bearing fused sequences and modifications.

# BIOGRAPHICAL SKETCH

## EDUCATION

- **The University of Texas at Austin**, Biochemistry (B.Sc.), *08/05 – 08/09*

  *Thesis title*: Aptamer selection and application in cancer diagnostics and therapeutics, *Supervisor:* Dr. Andrew Ellington

- **University of Technology, Ho Chi Minh**, Vietnam, Food Chemistry Engineering (B.E.), *07/03 – 05/04*; awarded a full-ride fellowship and transferred to the University of Texas at Austin

## AWARDS AND HONORS

- Dissertation Year Fellowship, *UCLA*, 2014-2015
- Samson H. Cheng Biochemistry Teaching Award, *UCLA Chemistry and Biochemistry Department,* 2014
- Excellence in 2nd Year Academics and Research, *UCLA Chemistry and Biochemistry Department,* 2011
- Hanson-Dow Teaching Assistant Award, *UCLA Chemistry and Biochemistry Department,* 2010

## PUBLICATIONS AND ORAL PRESENTATION

- "Novel Acyl-Lysine Modifications in a Bacterial Proteome Elucidate Substrate Metabolism." Nguyen, H.H., Warren, J., McInerney, M., Gunsalus, R., Loo, J.A., and Ogorzalek Loo, R.R. Oral presentation at the Annual American Society for Mass Spectrometry (ASMS) Conference, 2014
- "Inhibition of Cell Proliferation by an Anti-EGFR Aptamer." Li, N., Nguyen, H.H., and Ellington, A.D. *PLoS One,* 2011
- "Directed Evolution of Gold Nanoparticle Delivery to Cells." Li, N., Larson, T., Nguyen, H.H., Sokolov, K.V., and Ellington, A.D. *Chem. Comm.,* 2010, 46(3), 392-4.
- "Technical and Biological Issues Relevant to Cell Typing by Aptamers." Li, N., Ebright, J.N., Stovall, G., Chen, X., Nguyen, H.H., Singh, A., Syrett, A., and Ellington, A.D. *J. Proteome Res.*, 2009, 8(5), 2438–48.

# CHAPTER 1: OVERVIEW

## 1.1. Motivation for studying syntrophic bacteria

Methanogenesis is the degradation of organic matter into methane by microorganisms and this process releases the majority of methane into the atmosphere (Kirschke et al., 2013). Despite being a potent greenhouse gas (Yvon-Durocher et al., 2014), methane can also be harvested from naturally occurring and anthropogenic sources to be used as a renewable source of energy. In methanogenic environments, methane production and carbon cycling occurs through sequential steps: fermentative bacteria first degrade complex organic matter into acetate and longer-chain fatty acids, propionate, alcohols, carbon dioxide, formate, and hydrogen; syntrophic bacteria then convert these products plus some aromatic compounds to hydrogen, formate, and acetate; and hydrogenotrophic and acetotrophic methanogens finish the last step by converting acetate, formate, and hydrogen into carbon dioxide and methane (McInerney, Sieber, & Gunsalus, 2009).

Syntrophic bacteria play an indispensable and rate-limiting step in this process (Sieber, Le, & Mcinerney, 2014). Many of the reactions to activate and metabolize substrates are thermodynamically unfavorable; the bacteria depend on methanogens to maintain end products at a very low concentration to drive the reactions. Moreover, critical oxidation-reduction reactions remain unfavorable even at the equilibrium hydrogen partial pressure (~$10^{-9}$ Torr) achieved by hydrogenotrophic methanogenesis, necessitating reverse electron transport or other exotic means to drive the chemical conversions essential to life (Schink, 1997). As a result, metabolizing a mole of substrate only yields one-third mole of ATP, redefining the smallest quantum of energy a living cell can make use (Jackson & McInerney, 2002). Living at such a “thermodynamic limit of microbial growth”, how do syntrophic bacteria distribute and prioritize their energy usage? Using

a proteomic approach, the work from this dissertation has revealed traits of the organisms' proteome to elucidate their cellular activities and choices. Understanding syntrophic bacteria and the syntrophy lifestyle is important to understand and engineer organic matter degradation for global carbon cycling, waste remediation and biofuel production.

## 1.2. Syntrophic bacteria have complex "proteoforms"

Since proteins are the "workhorses" of the cell and take part in essentially every structure and activity of life, studying the proteomes of syntrophic bacteria can elucidate the dynamic life of the cells. Studying the proteome of *Syntrophus aciditrophicus*, a model syntrophic bacterium, we observed the abundance of metabolic enzymes, translation-related proteins, signal transduction proteins, and general or unknown functional proteins. Protein expression alone, however, does not fully differentiate cells grown under different conditions or reveal energy conservation strategies. Thus we looked deeper into the proteomes of *S. aciditrophicus* and revealed in **Chapter 2** the signature of the proteome: complex "proteoforms". The term was coined by Lloyd Smith and Neil Kelleher to describe all of the different molecular forms of a single gene product, including changes due to genetic variations, alternatively splicing and post-translational modifications (Smith & Kelleher, 2013). We discovered many amino acid mutations and protein sequence variants, including fusion forms of adjacent proteins on the same operons. Many protein modifications were also observed. Many are known, many are first recorded, and many belong to the rising family of lysine acylations. Perhaps such complex proteoforms generated by integrating protein sequence variants and/or protein modifications could hold the key to metabolic regulation in syntrophic bacteria to help it to adjust to environmental changes or altered nutrient availability at low energy cost.

## 1.3. Studying lysine acylations elucidates cellular metabolism

Although acetylation of lysines in histone proteins has long been known (Yang, 2004)(Glozak, Sengupta, Zhang, & Seto, 2005), the catalog of lysine acetylations outside eukaryotic cell nuclei was only greatly expanded in the mid-2000s, when mass spectrometry and acetyl-lysine specific antibodies were applied (S. C. Kim et al., 2006)(Choudhary et al., 2009)(Lundby et al., 2012). In prokaryotes, acetylation has emerged as an important regulatory mechanism coordinating metabolism in response to changes in nutrient availability, and the extent of acetylation has been found to be influenced by changes in carbon source, growth phase, and availability of acetylphosphate (Q. Wang et al., 2010a)(B. Weinert et al., 2013). In mitochondria, lysine acetylation has similarly been shown to be tightly coupled to nutrient availability (Wagner & Payne, 2013)(Papanicolaou, O'Rourke, & Foster, 2014). Work from the last few years has revealed a new class of lysine modifications, primarily in eukaryotes, that stem from the reaction of proteins with thioester-coenzymeA (CoA) byproducts of metabolism. Acyl groups can be transferred by CoA esters either enzymatically or non-enzymatically (Papanicolaou et al., 2014). Thus we have become fascinated by the potential role of acylations in enabling organisms to survive at the thermodynamic limit of life.

**Chapters 3 and 4** describe our lysine acylation studies in *Syntrophomonas wolfei* and *Syntrophic aciditrophicus,* which produce a wide range of acyl-CoA metabolites during substrate activation and degradation. This further confirms the correlation between lysine acylation modifications and their relevant acyl-CoA metabolites and leads to the discovery of many lysine acylations first seen in any organism. The prevalence, dynamics and relation to metabolic fluxes of these acylations suggest potential roles for metabolic regulation to optimize ATP production in such organisms living at thermodynamic limit of life.

## 1.4. Intact protein study

To further characterize fused protein complexes and acylated protein stoichiometry and conformation changes, we looked into the possibility of studying them in their intact forms. **Chapter 5** describes an additional approach for data analysis to maximize peak matching capabilities and protein modification detection. This will also be beneficial to most top-down proteomic applications.

# CHAPTER 2: ENERGETICALLY STRESSED SYNTROPHIC BACTERIA HAVE COMPLEX PROTEOFORMS

## 2.1. Abstract

Syntrophic bacteria are essential for the complete conversion of natural polymers to carbon dioxide and methane and are responsible for the rate-limiting step in global carbon recycling. They spend two-thirds of the generated ATP in substrate degradation, leaving one-third or an energy quantum to sustain microbial life. Thus they are well adapted to an energetically stressed lifestyle but how the energy is conserved and distributed is poorly understood. Protein expression data alone is enough to fully explain their syntrophic lifestyles so we looked deeper into the proteomic data of *Syntrophus aciditrophicus*, a model syntrophic bacterium. Many of its proteins appear to be highly heterogeneous when migrating to various locations on two-dimensional polyacrylamide gels; many especially have observed molecular weights much higher than expected. These “heavier” isoforms appear to arise from either insertion/deletions inside the genome or programmed translational recoding. Besides having diverse sequences, *S. aciditrophicus* proteins also contain many modifications. Many are known, and many have not been reported previously; some are based on carbon addition, and many are novel lysine acylations. Such protein diversity suggests a mechanism for stress adaptation and energy conservation.

## 2.2. Introduction

Methanogenic microbiota is responsible for most methane production and for cycling organic matter back into the environment. Fermentative microbes, anaerobic syntrophic bacteria and methanogens form an “assembly-line” to digest products of preceding organisms to feed

succeeding ones, forming discrete metabolic steps (McInerney et al., 2009). Syntrophic bacteria are essential for the complete conversion of natural polymers such as polysaccharides, proteins, nucleic acids, and lipids to $CO_2$ and $CH_4$ and are responsible for the rate-limiting step in global carbon recycling (Sieber et al., 2014). Making barely enough energy for survival (Jackson & McInerney, 2002), how do syntrophic bacteria conserve and distribute energy? Understanding biological processes of syntrophic bacteria is important to better understand methane production and to improve anaerobic digesters for fuel production.

Syntrophic bacteria studies are still limited due to isolation and cultivation difficulties (McInerney et al., 2009). Only a small fraction of the total microbial community present in natural environments can be successfully isolated from the complex microbiota while still having their interactions with syntrophic metabolizers preserved. Another limiting factor is genome annotation. Different from model organisms, many of their genes are only shared among certain microbes and most are understudied and remain unannotated.

*Syntrophus aciditrophicus*, a Gram-negative strictly anaerobic bacterium, is one of the more frequently studied syntrophs and can be cultured without a syntrophic partner (in crotonate or benzoate/crotonate substrate mixture). Its genome annotation has been revealed (McInerney et al., 2007) and many genes have been well-studied (Kung, Seifert, von Bergen, & Bolla, 2013)(M. Kim, Le, McInerney, & Buckel, 2013)(Löffler et al., 2011). Our collaborator (Professor Robert Gunsalus, UCLA) also has transcriptomic data for this system to aid our proteomic study.

Preliminary proteomic data from our lab compared the proteomes of *S. aciditrophicus* under different growth conditions but could not reveal large distinctions or energy conservation strategies. We hypothesized that information hidden behind protein sequences predicted from the genome could help elucidate features of the syntrophic lifestyles. Various data analysis approaches were applied to two different proteomic datasets and we began to unveil sets of

complex “proteoforms” (all of the different molecular forms of a single gene product, including changes due to genetic variations, noncanonical translation or post-translational modifications) for this organism, and probably other syntrophic organisms as well.

## 2.3. Materials and Methods

**2D-PAGE proteomic dataset** *S. aciditrophicus* proteins from whole cell lysates were separated by two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) spanning pH 3-10 isoelectric point (pI). Hundreds of gel spots were excised. Gel-embedded proteins were digested with trypsin and analyzed by LC-MS/MS. LC-MS/MS employed a quadrupole time-of-flight mass spectrometry system (Applied Biosystems/Sciex QSTAR Pulsar XL mass spectrometer) equipped with a Dionex/LC Packings nano-LC system (Jupiter Proteo C12 resin, 75 µm x 150 mm). This dataset was also used in **Chapter 4.**

**High throughput, high mass accuracy proteomic dataset** *S. aciditrophicus* was cultured axenically on either crotonate or benzoate/crotonate mixture (Elshahed & McInerney, 2001). Cells were lysed and digested by the eFASP method (Erde, Loo, & Loo, 2014). Tryptic peptides were pre-fractionated using strong cation exchange. Fractions were analyzed by liquid chromatography-tandem mass spectrometry (LC-MS/MS) on an EASY nLC1000 (Thermo Scientific) coupled to a hybrid mass spectrometer (Q-Exactive, Thermo Scientific). Peptides were desalted on an Acclaim PepMap100 C18 nano-trap column (Dionex, 75 um x 2 cm, Product # 164535) and separated on an Acclaim PepMap RSLC analytical column (Dionex, 75 um x 25 cm, Product # ES802). This dataset was also used in **Chapter 4.**

**Transcriptomic data analysis** RNA-seq data on *S. aciditrophicus* grown axenically on 10 mM crotonate were provided by Professor Robert Gunsalus (UCLA). Sequence alignment was done with NovoAlign (Novocraft Technologies Sdn Bhd, Selangor, Malaysia) using default

parameters. SAMtools (Sequence Alignment/Map) was used for consensus/indel calling (http://samtools.sourceforge.net/).

**Protein sequence databases** The genome annotation of *Syntrophus aciditrophicus* was downloaded from ftp://ftp.ncbi.nih.gov/genomes/. Another protein sequence database containing transcriptomics-evidenced insertions, deletions or high frequency SNPs (single-nucleotide polymorphism) was built by in-house software.

**"Blind" protein modification search** Besides known modifications curated in the Unimod database (www.unimod.org), we looked for protein modifications unique to *S. aciditrophicus* by performing mass spectrometry data searching *without specifying in advance* which modifications are present in the dataset. MODa v1.3 (Na, Bandeira, & Paek, 2012) was obtained from Dr. Eunok Paek (Hanyang University, Seoul, Korea) (http://prix.hanyang.ac.kr/download/moda.jsp) and run on a local computer.

**"Flexible" database search** Mass spectral data were exported into mgf format and searched with Mascot version 2.5 (Matrix Science, Boston, MA) against protein sequence databases with the appropriate mass error tolerance (±0.3 Da for precursor and fragment ions for 2D-PAGE dataset and ±5 ppm for precursor and ±0.01 Da for fragment ions from the high mass accuracy dataset). Carbamidomethylation on cysteine and oxidation on methionine were specified as initial variable modifications. Mascot "error tolerant search mode" allows at most an additional type of protein modification from the modification databases, an additional missed cleavage and an additional non-specific protease-digested site (http://www.matrixscience.com/help/error_tolerant_help.html). Searched data were confirmed semi-automatically by in-house software.

**Selenoprotein prediction** Sequences were submitted to the web server from the Gladyshev lab (http://gladyshevlab.org/SelenoproteinPredictionServer/about.html) (Harvard Medical School, Boston, MA) (Kryukov & Gladyshev, 2004).

**Phylogenetic tree** Protein sequences were submitted to the Clustal Omega website (http://www.ebi.ac.uk/Tools/msa/clustalo/) for sequence alignment and phylogenetic tree presentation.

## 2.4. Results

### 2.4.1. Atypical protein migration on 2D-PAGE gels might indicate protein fusion events

Two-dimensional gel electrophoresis coupled with mass spectrometry not only allowed us to thoroughly separate proteins but also to observe the migration patterns of complex protein samples. Evidence was found for many proteins migrating to multiple positions, with higher and lower molecular weights (MWs) and pIs than predicted, on the 2D-gels. **Table 2.1** includes a list of proteins observed to migrate to positions with molecular weights higher than predicted, especially for proteins co-migrating with products from the adjacent genes. The first 4 groups of proteins listed in the Table, using homology and transcriptomic data, can be explained by insertion/deletion in the mRNA/DNA leading to frameshifting.

Many other proteins, however, were observed to migrate at predicted MWs and to randomly higher MW spots. **Figure 2.1** highlights the migration patterns of 2 representative proteins. Because the RNA-seq data do not show any sequence variants to explain this behavior, these proteins might have been fused by programmed translational recoding (Sund, Andér, & Aqvist, 2010)(Stahl, McCarty, & Farabaugh, 2002).

### 2.4.2. Some of the detected *S. aciditrophicus* proteins may contain selenocysteine

Besides detecting predicted peptides of SYN_00603, a formate dehydrogenase major subunit, we also found two peptides not predicted from the protein sequence (HSPTVPALGESFGR and GAMTNHWIDIR) with good spectrum matching qualities. These reside in the connecting region between the stop codon ahead of SYN_00603 and its predicted initiation site (**Figure 2.2**). Since translation initiation has only been observed for AUG, GUG, UUG and sometimes CUG, and the first amino acid residue should be methionine (Lobanov, Turanov, Hatfield, & Gladyshev, 2010)(Touriol et al., 2003), the peptide HSPTVPALGESFGR must belong to the translation of the region ahead of SYN_00603 as well. Since the transcriptomics data do not show any mRNA sequence variants in this region, some translational recoding event could have caused it (Baranov et al., 2001). In fact, the selenoprotein prediction webserver predicts the leading stop codon to be recoded for selenocysteine (Kryukov & Gladyshev, 2004).

The list of proteins in **Table 2.1** was checked with the selenoprotein prediction algorithm. The stop codon between SYN_00630 and SYN_00629 (formate dehydrogenase proteins) could probably be used alternatively for selenocysteine. SYN_01644 (heterodisulfide reductase subunit A) and SYN_01645 could also be fused by similar recoding.

### 2.4.3. Protein fusions were also observed in another syntrophic bacterium, *Syntrophomonas wolfei*

To investigate if such protein fusion events are common in syntrophic bacteria that share an energetically stressed lifestyle, we performed electrophoresis of proteins from *Syntrophomonas wolfei*. Protein samples were boiled in SDS before 1D-SDS-PAGE separation and mass spectrometry analysis of the protein bands also indicated many proteins migrating to higher molecular weight positions (not shown).

Swol_2051 (acetyl-CoA C-acetyltransferase) and Swol_2052 (acyl-CoA dehydrogenase) are among the detected *S. wolfei* proteins migrating to higher MW on the SDS-PAGE gel. Since acetyl-CoA C-acetyltransferases and acyl-CoA dehydrogenases are next to each other on three different operons (**Figures 2.3** and **2.4**), we had the UCLA DOE Protein Expression Facility clone the region containing Swol_2051 and Swol_2052 with a His-tag on Swol_2052.

Recombinant protein expression showed that both Swol_2051 and Swol_2052 are expressed at the expected molecular weights and Swol_2052 is retained on the Ni-NTA column because of the His-tag (**Figure 2.5**). However an additional protein band at high molecular weight was also observed (**Figure 2.5**). It contains peptide sequences from both Swol_2051 and Swol_2052. This fused isoform could not be separated from the 60 kDa Swol_2052 product during size-exclusion chromatography separation (data not shown), suggesting a strong physical association between them.

### 2.4.4. *Syntrophus aciditrophicus* contains many protein modifications, including carbon addition and lysine acylations

The majority of *S. aciditrophicus* peptide mass spectra were not matched to any predicted peptide sequence; some of those spectra could be explained by unknown protein modifications. We used MODa (Na et al., 2012), a “blind protein modification” search software, to unveil such potential modifications. All of the potential novel protein modifications suggested by MODa were confirmed manually. These modifications were then concatenated with the Unimod protein modification database so that Mascot can perform an error-tolerant search for peptides containing any of these modifications. Many novel protein modifications were found, especially modifications on methionine at the protein N-termini. Many modifications target the same residues on a protein.

We also observed a unique and prevalent protein modification that appeared to be due to addition of a single carbon (+12.000 Da). This modification can be localized to various amino acid

residues such as P, H, Y, I, L, V, S, A, or K. A number of mass spectra representing the same peptide sequence with different residues bearing the carbon addition were detected (**Figure 2.6**). This could arise from either the lability of the modification or from peptides eluting at the same LC retention time. This modification is frequently seen near the N-terminus of digested peptides and in peptides bearing proline. Carbon addition is also observed frequently in *Syntrophomonas wolfei* proteins but not in *Escherichia coli* proteins.

Lysine crotonylation, glutarylation and many other novel lysine acylations were also detected. Those are described in more detail in **Chapters 3** and **4**.

## 2.5. Discussion

### 2.5.1. Protein fusion from programmed translational recoding events may function to increase enzyme efficiency

Various types of experimental data suggest gene fusion events in *S. aciditrophicus* and *S. wolfei* arise from programmed translational recoding: (i) the *S. aciditrophicus* transcriptomics data did not show insertion or deletion in genes of most fused proteins; (ii) the cloned *S. wolfei* DNA made in *E. coli* produced fused Swol_2051_2052; (iii) the two-dimensional gel data show only a proportion of proteins are fused, similar to recoding characteristics (Gesteland & Atkins, 1996)(Baranov et al., 2001); and (iv) a few *S. aciditrophicus* proteins are predicted to have selenocysteine. Programmed translational recoding has been suggested to increase the diversity of gene expression and allows the organism to adapt to a new ecological niche without losing its ability to survive in the original one (Namy, Rousset, Napthine, & Brierley, 2004).

Fatty acid synthase is a classic example of gene/protein fusion to prevent dissociation at low protein concentration (McCarthy and Hardie, 1984). Nature contains many other gene/protein fusion events to ensure coregulation or stoichiometry production or to give rise to new functions

(Long, 2000). Syntrophic bacteria probably take advantage of this strategy to deal with its energetically stressed lifestyle.

*S. wolfei* enzymes participating in Steps 3 and 4 of the β-oxidation pathway are duplicated together on 3 operons (**Figures 2.3** and **2.4**) and each pair has mirrored expression levels. Besides coregulation and stoichiometry maintenance, protein fusion could probably help substrate channeling from one step to the other to enhance productivity. The physical contact advantage for *S. wolfei* enzymes participating in Steps 2 and 5 of β-oxidation (**Figures 2.3, 2.4** and **2.5**) is not as clear but we know that (i) in Step 2, the enzyme catalyzes the most endergonic step of the pathway and two-thirds of the produced ATP have to be used to regenerate oxidizing equivalents (Schink & Friedrich, 1994), and (ii) in Step 5, the enzyme catalyzes the most exergonic step of the pathway and the reaction could happen spontaneously.

### 2.5.2. Syntrophic bacteria might produce formaldehyde

Proteins from *S. aciditrophicus* and *S. wolfei,* but not from *E. coli*, have the +12-Da carbon addition modification. The modification can be found on various amino acid residues, especially the peptide N-term terminus. Formaldehyde has been shown to react with C, R, W, K, H, and N-terminus *in vitro* with a mass addition of 12 or 30 Da (Metz et al., 2004). This suggests that our observed carbon addition could be caused by formaldehyde or by other similar active compounds.

Our protein samples were prepared either by the eFASP method (i.e., lyse cells and then perform protein clean up and digestion on filter-aided columns) (Erde et al., 2014) or the normal method used for 2D-PAGE separation (i.e., lyse cells in rehydration buffer, reduce and alkylate proteins before running them on a gel, and then cutting out the gel spots and trypsin digestion). Both methods used did not introduce formaldehyde to the sample, and formaldehyde is not likely to be present during peptide digestion. The tendency for carbon addition to be found on the

peptide N-terminus might suggest the modification to be labile and re-localize to the N-terminus after proteolytic digestion.

More work will be needed to identify the origin of the +12-Da modification. Syntrophic bacteria do not produce formaldehyde but their syntrophic partners, methanogens, do. It would be interesting to see how the level of carbon addition could vary among cultures of syntrophic bacteria growing by themselves or with the methanogens and to see if syntrophic bacteria actually produce formaldehyde as a mechanism to dispose of some of the reducing equivalents.

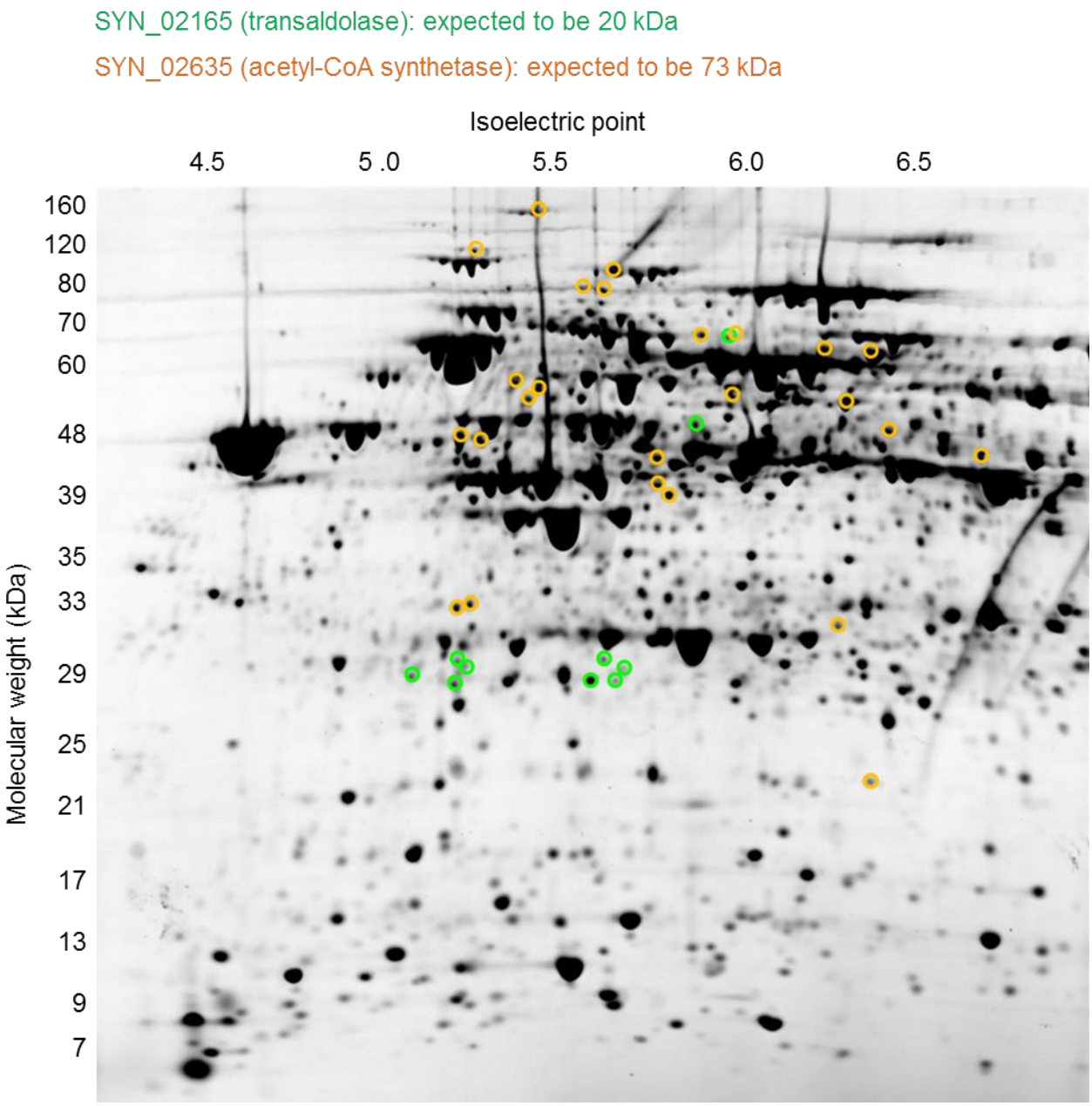



**Figure 2.1:** Representative examples of proteins migrating to many locations, especially at molecular weights higher than predicted, on a two-dimensional gel. Peptide sequences from SYN_02165 were found in the green circled spots, and peptide sequences from SYN_02635 were found in the orange circled spots on the gel.

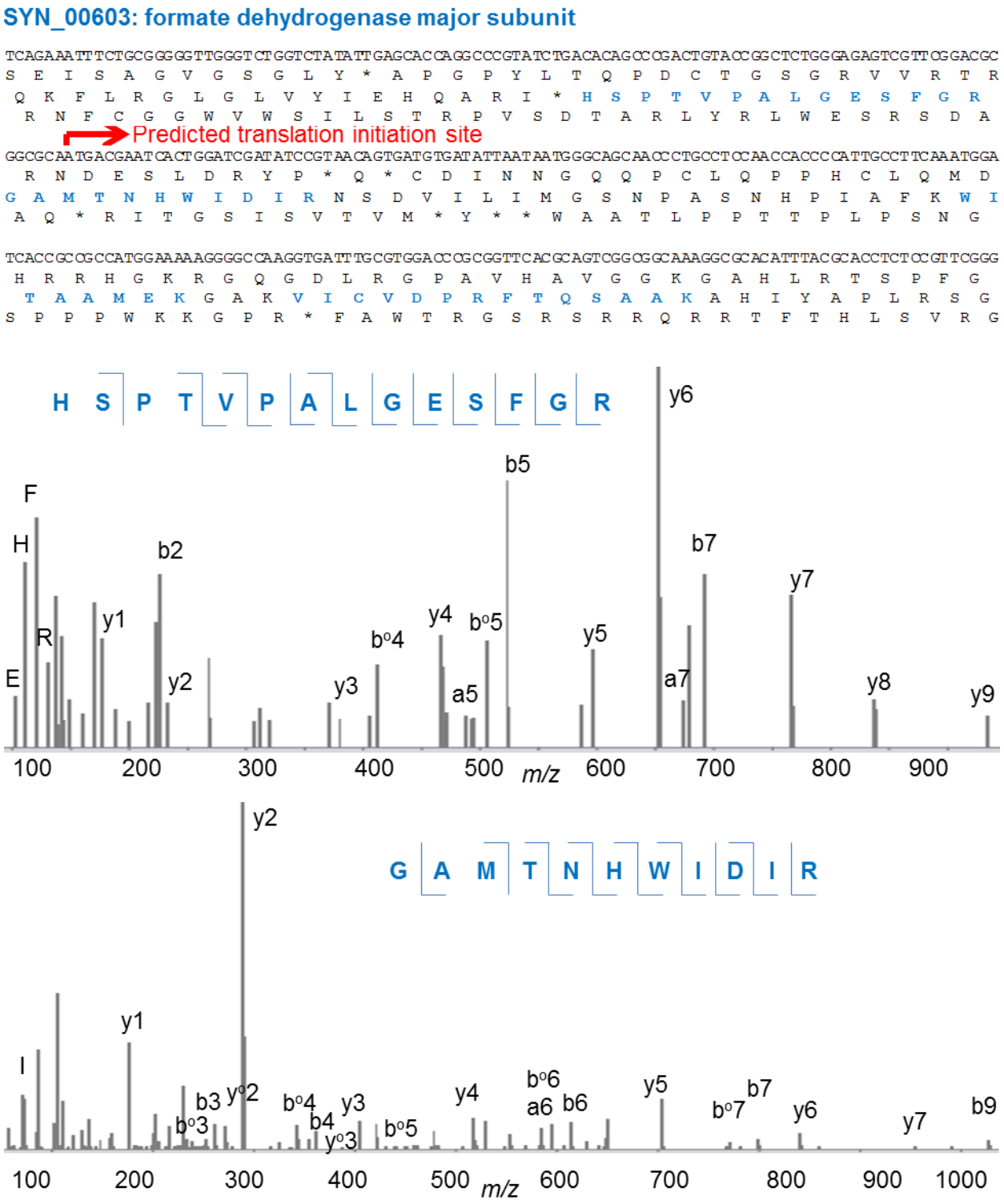



**Figure 2.2:** SYN_00603 with 2 (supposedly tryptic) peptides in the connecting region between the stop codon ahead of SYN_00603 and its predicted initiation site. Peptides colored in blue were detected by mass spectrometry. Product ion mass spectra for two of the peptides are shown.

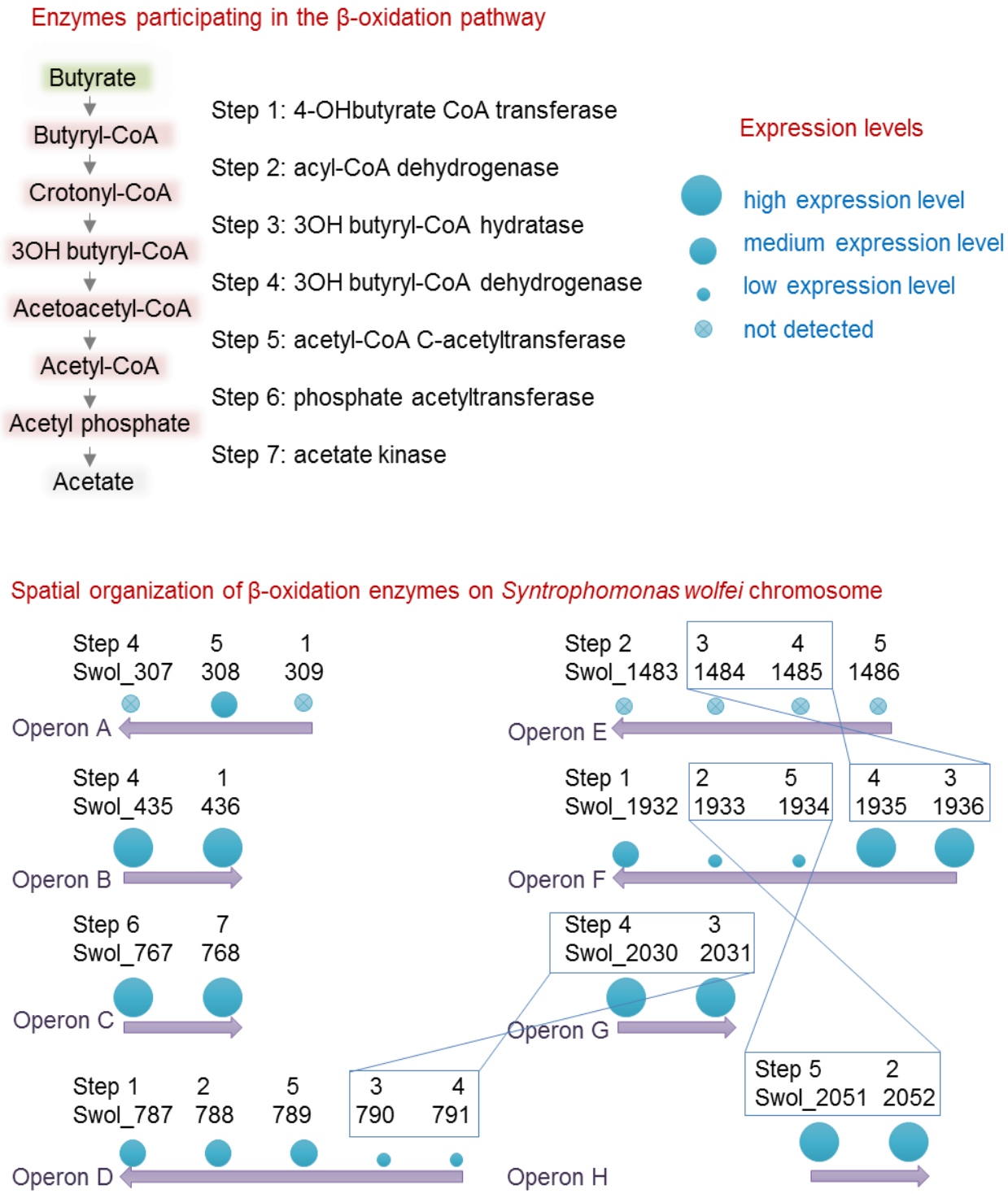



**Figure 2.3:** Enzymes participating in the β-oxidation pathway in *Syntrophomonas wolfei*. Enzymes involved in Steps 3 and 4 and in Steps 2 and 5 tend to pair with each other on both spatial organization and protein expression levels. The lines connecting the boxes show enzyme pairs having high sequence homology (refer to Figure 2.4).

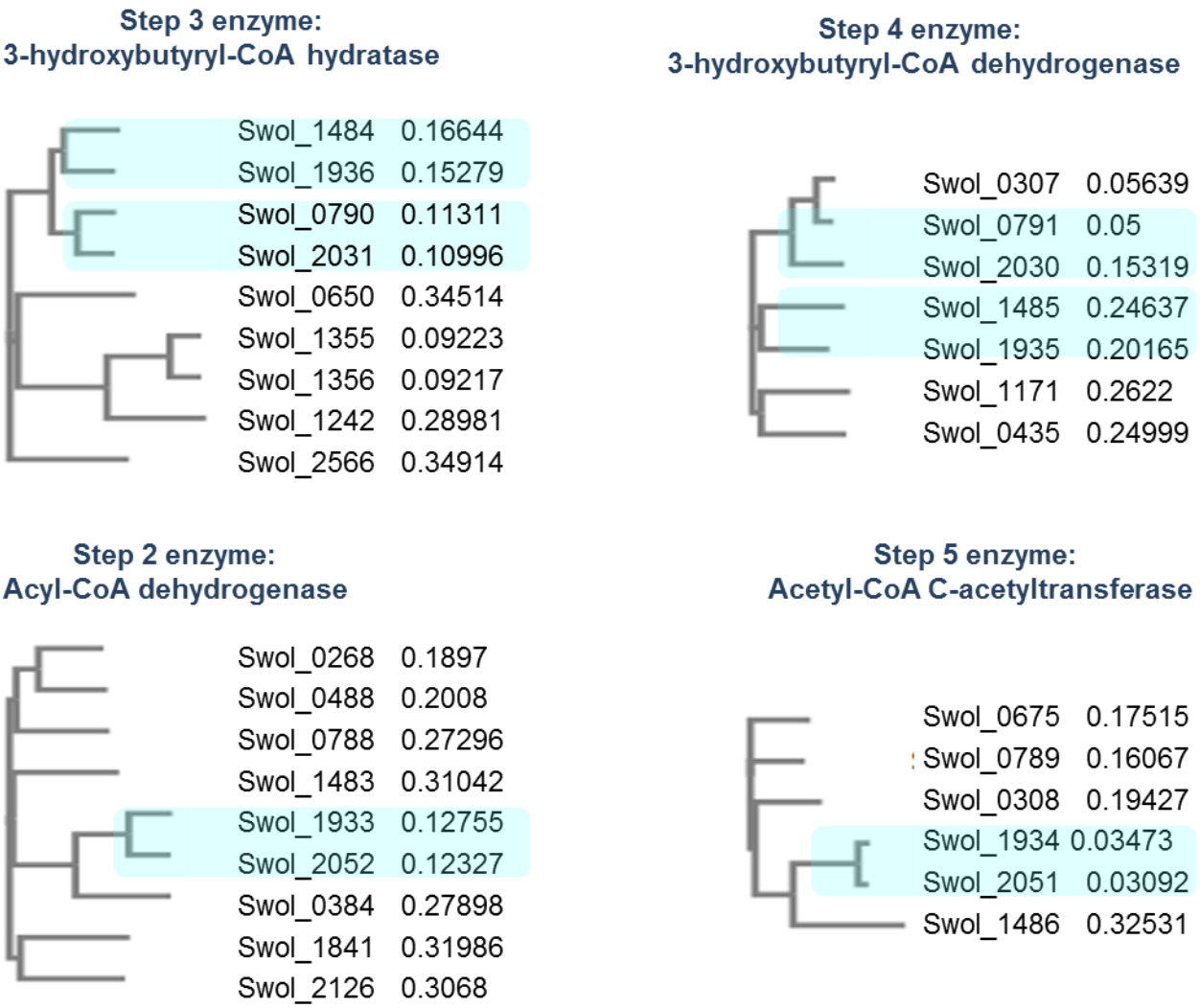



**Figure 2.4:** Phylogenetic trees of enzymes participating in different steps of the β-oxidation pathway. Colored enzyme pairs have spatial organization association (refer to Figure 2.3).

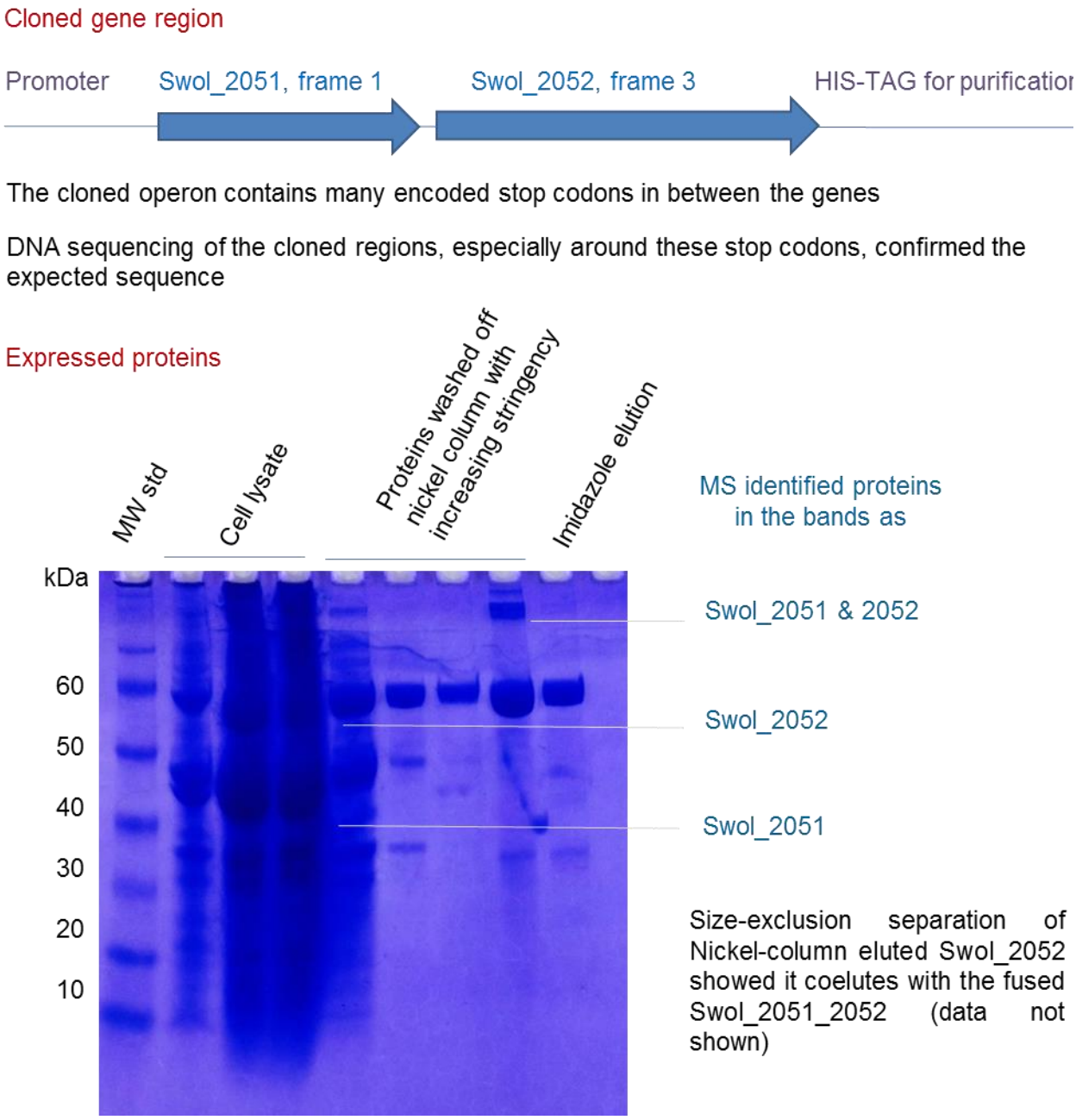



**Figure 2.5:** The gene region containing Swol_2051 (acetyl-CoA C-acetyltransferase) and Swol_2052 (acyl-CoA dehydrogenase) were cloned, with a His-tag on the C-terminal end of Swol_2052, and expressed. Both proteins were expressed and Swol_2052 was retained on the nickel column. The band on the SDS gel at high molecular weight was identified with peptides from both Swol_2051 and Swol_2052.

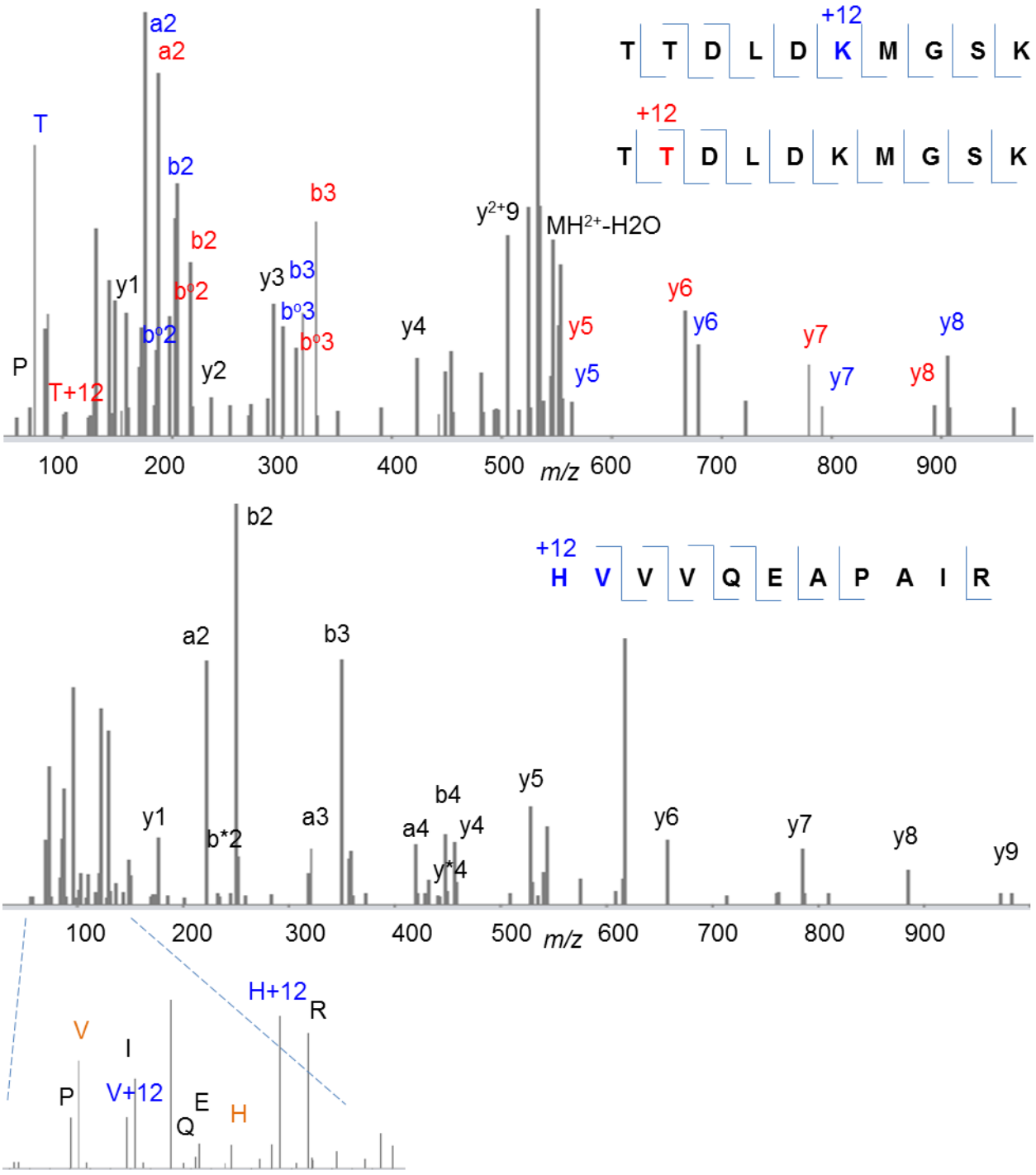



**Figure 2.6:** (Top) A product ion mass spectrum showing a single peptide with two possible location assignments for carbon addition. The peptide belongs to SYN_00480, an acyl-CoA dehydrogenase. (Bottom) A product ion mass spectrum showing carbon addition at the peptide N-terminus. Based on the detection of the appropriate immonium ions and the relative intensities among these ions, both H and V can be modified and H is probably modified at a higher ratio. The peptide belongs to SYN_01370, an iron-only hydrogenase (large subunit).

**Table 2.1:** Proteins migrating to many positions on the gel, especially locations with molecular weights higher than expected

| Protein | Description | Peptides detected in frame # | Cal. MW (kDa) | Observed MW on 2D gel (kDa) | Homologous protein MW (kDa) | RNA-seq mutations (% frequency) |
|---|---|---|---|---|---|---|
| - | - | 2 | | | 28 | **C+1G (73%)** |
| SYN_02983 | cytoplasmic prot | 1 | 4 | 29 | | **G+1A (79%)** |
| SYN_03790 | triosephosphate isomerase | 3 | 14 | 29 | | |
| SYN_02165 | transaldolase | 1 | 20 | 64, 48, 30 | 24 | G+1T (4%), **T+1C (92%)** |
| SYN_03220 | hypothetical prot | 3 | 5 | 28 | | |
| SYN_03154 | branched-chain aa aminotransfrs. | 1 | 26 | 40 | 40 | **T-1C (86%), A-1T (93%)** |
| - | - | 3 | | | | |
| SYN_03155 | branched-chain aa aminotransfrs. | 1 | 14 | 40 | | **A-1G (81%)** |
| - | - | 1 | | | 42 | |
| SYN_00124 | phosphoserine aminotransfrs. | 2 | 39 | 41 | | **C+1G (91%), G+1C (91%)** |
| SYN_00630 | formate dehydro. Fe-S sub | 1 | 38 | 97, 66, 38 | 39 **& 99** | |
| SYN_00629 | formate dehydro. cat sub | 1 | 59 | 97, 66 | 59 **& 99** | |
| SYN_01653 | enoyl-CoA hydratase | 3 | 28 | 40, 30 | 29 | |
| SYN_01654 | enoyl-CoA hydratase | 3 | 43 | 97, 46, 40, 30 | 42 | C+1A (6%) |
| SYN_01655 | Zn-dpdnt. alcohol dehydro. | 1 | 40 | 97, 40, 30 | 36 - 45 | |
| SYN_02586 | acyl-CoA dehydro. | 1 | 42 | 73, 42, 30 | 42 | |
| SYN_02587 | acyl-CoA dehydro. | 2 | 45 | 45, 40 | 42 | G+1A (2%) |
| SYN_02635 | acetyl-CoA synthetase | 1 | 73 | 100, 90, 80, 70, 60, 50, 40, 30 | 72 | |
| SYN_02636 | ETF, b subunit | 1 | 27 | 70, 60, 40, 30 | 28 | |
| SYN_02637 | ETF, a subunit | 2 | 34 | 100, 90, 70, 60, 40, 35, 30 | 36 - 44 | |
| SYN_02641 | acetyl-CoA acetyltransfrs. | 2 | 44 | 57, 46, 35 | 44 | C+1T (2%) |
| SYN_02642 | cofactor-indpdnt. phosphoglycerate n | 1 | 45 | 97, 50, 45, 40, 30 | 44 | |
| SYN_01369 | NADH-quinone ox-re chain F | 3 | 66 | 70, 66 | 24 - 53 | |
| SYN_01370 | Fe only hydrogenase large sub | 1 | 66 | 97, 77, 70, 66 | 64 | |
| SYN_00480 | acyl-CoA dehydro. | not co-localize | 42 | 104, 55, 40, 30 | 43 | |
| SYN_00481 | glutaconyl-CoA decarb. sub A | not co-localize | 65 | 97, 73, 66, 50 | 62 | C+1A (4%), C+1A (4%) |
| SYN_03861 | hypothetical prot | not co-localize | 6 | 104 | NA | |
| SYN_01644 | heterodisulfide reductase subunit A | not co-localize | 7 | 104 | 7 - 165 | |
| SYN_00084 | multimeric flavodoxin | not co-localize | 17 | 53 | 20 | |
| SYN_00419 | aspartate 1-decarboxylase | not co-localize | 13 | 40 | 13 | |
| SYN_00007 | transcriptional regulator | not co-localize | 59 | 97 | 51 - 66 | |
| SYN_00027 | molecular chaperone GroEL | not co-localize | 59 | 97, 82, 70, 60, 50, 40, 35 | 61 | |
| SYN_00544 | ATP synthase F0F1 subunit beta | not co-localize | 51 | 97, 50, 40, 30 | 52 | C+1A (2%) |
| SYN_01681 | acetyl-CoA acetyltransferase | not co-localize | 41 | 73, 50, 40, 30 | 43 | |
| SYN_01909 | molecular chaperone GroEL | not co-localize | 58 | 97, 70, 60, 50, 40, 35 | 61 | G+1A (3%) |
| SYN_01945 | anthranilate phosphoribosyltransferas | not co-localize | 36 | 68, 41, 30 | 37 | A+1T (6%), A+1T (10%) |
| SYN_02896 | 4-hydroxybenzoate--CoA ligase / ben | not co-localize | 59 | 97, 73, 60, 50, 40, 30 | 58 | |
| SYN_03223 | molecular chaperone GroEL | not co-localize | 58 | 104, 97, 82, 60, 40, 30 | 61 | G+1A (1%) |

# CHAPTER 3: METABOLITE-DRIVEN MODIFICATIONS: LYSINE ACYLATIONS ELUCIDATE SUBSTRATE METABOLISM IN *SYNTROPHOMONAS WOLFEI*

## 3.1. Abstract

Lysine acetylation knowledge has extended greatly beyond its roles in chromatin maintenance and transcription regulation to an emerging function as an important regulatory mechanism coordinating metabolism in response to changes in nutrient availability. Research during the past few years has revealed a new class of lysine modifications, primarily in eukaryotes, that stem from the reaction of proteins with thioester-coenzymeA (CoA) byproducts of metabolism. Acyl groups can be transferred by CoA esters either enzymatically or non-enzymatically. These discoveries of new acyl-modifications decorating eukaryotic and, especially, mitochondrial proteins motivated us to explore their potential in syntrophic microbes that participate in organic matter degradation in methanogenic microbiota to cycle carbon back to the environment.

These environmental and biofuel important organisms metabolize substrates by activating them with coenzyme A and convert them stepwise to other acyl-CoA compounds and then to acetate, formate, $CO_2$ and $H_2$ with a feeble ATP (energy) harvest that barely sustains life. The proteome of *Syntrophomonas wolfei* reveals lysine acylation levels distinctive for different growth substrates: cells grown on butyrate have more lysine butyrylation and crotonylation whereas cells grown on crotonate result in more acetylation. Precursor ion intensities of peptides detected by mass spectrometry show the level of each acylation in this order: acetylation > butyrylation > 3-hydroxy butyrylation > crotonylation ≈ acetoacetylation. The acylations and their variation with cultivation conditions recapitulate the metabolic pathway for energy production in

*S. wolfei,* suggesting that analyses of lysine acylation may help elucidate substrate metabolism in an organism living at the thermodynamic limit of feasibility.

## 3.2. Introduction

Reversible lysine N-ε acetylation, distinct from generally irreversible N-α acetylation, is regulated by many acetyltransferases, deacetylases, and sirtuins – deacetylases that require cosubstrate $NAD^+$. In the 40 years following the 1964 discovery of lysine N-ε acetylation, less than one hundred modified proteins were identified, mostly eukaryotic, non-mitochondrial, and DNA-binding (Yang, 2004). At the time, acetylation was known to exert opposing effects on modified proteins, such as increasing or decreasing DNA binding affinity, transcriptional activation, protein-protein interactions, or protein stability (Glozak et al., 2005).

In the last decade, the successful combination of immunoprecipitation (IP) with mass spectrometry (MS) (S. C. Kim et al., 2006) led to the discovery of thousands of more acylated proteins in both eu- and prokaryotes. Newly found acetylated proteins participate in a wide range of activities, spanning from transcription and translation to metabolism, cell cycle, and subcellular localization (Choudhary et al., 2009)(Lundby et al., 2012). The modification has been shown to strongly couple to nutrition availability and/or to regulate metabolism (Q. Wang et al., 2010b) (Papanicolaou et al., 2014).

Mass spectrometry also revealed heretofore unknown acylations: propionylation (Chen et al., 2007), butyrylation (Chen et al., 2007), succinylation (Zhang et al., 2011), malonylation (Peng et al., 2011), crotonylation (Tan et al., 2011), glutarylation (Tan et al., 2014), and 2-hydroxy isobutyrylation (Dai et al., 2014). The first four have been found in both eu- and prokaryotes, whereas the last three have so far been observed only in eukaryotes. Except for crotonylation and 2-hydroxyisobutyrylation (found only in histones to date) (Tan et al., 2011)(Dai et al., 2014), the modified proteins participate in a broad range of processes. No acyl transferases have been

found that act solely on these novel modifications. Only a few lysine acetyl transferases, such as p300 and CBP, transfer propionyl, butyryl and crotonyl groups onto lysine (Chen et al., 2007)(Cheng et al., 2009)(B. Liu et al., 2009)(Sabari et al., 2015). Some metabolic enzymes can moonlight by transferring acyl groups onto lysines: acetyl-CoA acetyltransferase can catalyze acetylation (Montgomery, Sorum, & Meier, 2015) and alpha-ketoglutarate dehydrogenase complex succinylation (G. E. Gibson et al., 2015). Wagner et al showed another route to lysine acylation (Wagner & Payne, 2013), namely a non-enzymatic one. Their *in vitro* experiments revealed acylations occurring at high pH and high acyl-CoA ester concentrations - conditions similar to the mitochondrial matrix. These newly arisen acylations also respond to metabolic flux and all are reversible by sirtuins, which uses $NAD^+$ as co-substrate, responds to cellular redox status ($NAD^+$ to NADH ratio) and does not distinguish between acylations arisen enzymatically or spontaneously (Papanicolaou et al., 2014)(Shi & Tu, 2015)(AbouElfetouh et al., 2014). Lysine acylations have emerged as a new mechanism for regulating cellular metabolism in physiological and pathophysiological states at both the protein activity and protein expression levels (Sabari et al., 2015)(Kosono et al., 2015)(M. Hirschey & Zhao, 2015).

Our preliminary proteomic data have shown syntrophic bacteria to contain many novel protein modifications. Those organisms degrade short chain fatty acids or certain aromatic compounds, a process important for carbon cycling back to the environment and for biofuel production, and yield very little energy to support growth. *Syntrophomonas wolfei* generates acetyl phosphate and many acyl-CoA esters: acetyl-CoA, acetoacetyl-CoA, 3-hydroxy (3-OH) butyryl-CoA, crotonyl-CoA and butyryl-CoA while metabolizing butyrate in co-culture with a methanogen (Sieber et al., 2010). Lysine acylation primarily occurs in the presence of corresponding acyl-CoA esters, whether proceeding enzymatically (Chen et al., 2007) or spontaneously (Wagner & Payne, 2013); other than acyl-CoA, only acetyl phosphate and acetyl adenylate have been shown capable of acylation (Ramponi, Manao, & Camici, 1975)(Verdin & Ott, 2013). Thus we

hypothesized acyl-CoAs produced during *S. wolfei* substrate degradation could lead to acyl-lysine modifications, which could explain some novel observed modifications and provide an insight into a syntrophic lifestyle.

## 3.3. Materials and Methods

**Protein extraction and digestion** *Syntrophomonas wolfei was* cultivated axenically (crotonate substrate), or in co-culture with *Methanospirillum hungatii* (crotonate or butyrate substrate) (Sieber et al., 2015). Co-cultured organisms were separated by Percoll-gradient centrifugation. *S. wolfei* cells were lysed and digested by the eFASP method (Erde et al., 2014). Recovered peptides were dried in a vacuum concentrator for subsequent immunoprecipitations and mass spectrometry.

**UPLC-MS/MS Method** Tryptic peptides were pre-fractionated using strong cation exchange. Fractions were analyzed by liquid chromatography-tandem mass spectrometry (LC-MS/MS) with an EASY nLC1000 (Thermo Scientific) coupled to a hybrid mass spectrometer (Q-Exactive, Thermo Scientific). Peptides from the above section were desalted on an Acclaim PepMap100 C18 nano-trap column (Dionex, 75 um x 2 cm, Product # 164535) and separated on an Acclaim PepMap RSLC analytical column (Dionex, 75 um x 25 cm, Product # ES802).

**Immunoblotting** 40 mg of *Syntrophomonas wolfei* grown on crotonate (either monoculture or coculture with *M. hungatei* and then Percoll-separated) or butyrate (coculture and then separated) was resuspended in 300 µl of NETN buffer (100 mM NaCl, 20 mM Tris, 0.5 mM EDTA, 0.4% NP-40, pH 8.0) in the presence of protease inhibitors (Sigma, Product #P8340, St. Louis, MO). Cells were sonicated on ice for three 10 s intervals (Sonic Dismembrator model 100, Thermo Scientific Scientific, Rockford, IL), and centrifuged at 16 000 g for 10 min. Sonication and centrifugation were repeated once. Total proteins were quantified with a Pierce Micro BCA Protein Assay Kit (Thermo Scientific, Product # 23235).

50 µg of cell lysates were loaded onto NuPAGE gels (Life technologies, Product #NP0335, Grand Island, NY) and transferred to Immobilon-FL membranes (Millipore, Product # IPFL00010, Billerica, MA). The membranes were dried, re-activated, and incubated with blocking buffer (PBS, 0.5% BSA, 0.1% Tween-20, 0.02% SDS and 0.36 % sodium azide) before incubation with primary and secondary antibodies. Primary antibodies (PTM Biolabs, Product # PTM-101, PTM-502 and PTM-1151, Chicago, IL) were diluted 1:1,000 in blocking buffer and secondary antibodies (LI-COR, Product # 827-08364 and 827-08365, Lincoln, NE) were diluted 1:10,000 in blocking buffer. Membranes were washed in between antibody incubations with washing buffer (PBS, 0.1% Tween-20) and rinsed with PBS before imaging with a LI-COR Odyssey Infrared Scanner (LI-COR, Odyssey CLx). After imaging, total proteins were stained with Pierce Reversible Protein Stain Kit for PVDF Membranes (Thermo Scientific, Product # 24585).

**Lysine acylation prediction** *S. wolfei* generates acetyl phosphate and many acyl-CoA esters: acetyl-CoA, acetoacetyl-CoA, 3-hydroxy (3-OH) butyryl-CoA, crotonyl-CoA and butyryl-CoA while metabolizing butyrate in co-culture with a methanogen (Sieber et al., 2010). We used this acyl-CoA ester profile to predict for the presence of corresponding lysine acylations in cells grown on buryrate co-culture and crotonate co-culture and pure culture.

**Database search** Mass spectral data were searched with Mascot version 2.4.1 (Matrix Science, Boston, MA) through Thermo Proteome Discoverer version 1.4.0.288 (Thermo Scientific). All of the sample fractions were combined in a single search against the *Syntrophomonas wolfei* protein database downloaded from ftp://ftp.ncbi.nih.gov/genomes/ (2504 sequences) concatenated with a reverse decoy database and protein sequences of common contaminants. Carbamidomethylation on cysteine was specified as a fixed modification. Oxidation on methionine and acetylation, crotonylation, butyrylation, 3-hydroxybutyrylation and acetoacetylation on lysine were specified as variable modifications. Trypsin/P was specified as the cleavage enzyme, and the search allowed for up to two missed cleavages. The precursor

mass error was set to ±10 ppm and the product mass error to ±0.02 Da. Target false discovery rate (FDR) was specified to be 0.01 (strict) and 0.05 (relaxed).

**Stoichiometry calculation** Stoichiometry of an acylated lysine at position n was quantified based on the precursor ion peak areas of peptides with the modified lysines divided by the sum areas of all peptide peaks covering the same lysines as in this formula:

$$Acylation\ stoichiometry\ of\ lysine\ at\ position\ n = \sum Area_{peptides\ containing\ acylated\ lysine\ n} \times \left(\sum Area_{peptides\ containing\ acylated\ lysine\ n} + \sum Area_{peptides\ containing\ NON-acylated\ lysine\ n}\right)^{-1}$$

**Highly and weakly acylated proteins** Highly acylated proteins are defined as proteins with *high number of acylated peptides*; we picked an arbitrary cut-off number of at least 10 acylation events (a combination of acylation sites and acylation types). Weakly acylated proteins are defined as *strongly and moderately expressed proteins* with very *low number of acylated peptides;* we picked an arbitrary cut-off number of at most one acylation event.

**Pathway analysis** Protein sequences were submitted to WebMGA (web server for metagenomic analysis, http://weizhong-lab.ucsd.edu/metagenomic-analysis) hosted by the Weizhong lab (UCSD) and a BLAST search (Basic Local Assignment Search Tool) was performed against KEGG protein database with an e-value cutoff for prediction as 0.001. For each growth condition, if a protein has at least 3 peptides in at least 1 replicate, it is considered for pathway analysis.

**Acylation sequence motifs** Sequence motifs around acylated lysines were analyzed with WebLogo at http://weblogo.berkeley.edu/ (UC-Berkeley).

## 3.4. Results

### 3.4.1. Lysine acetylation, crotonylation and butyrylation are detected at high levels in *S. wolfei* by immunoblotting

*Syntrophomonas wolfei* can grow on crotonate in pure culture because it can bypass the most endergonic step of oxidizing butyryl-CoA to crotonyl-CoA and dispose reducing equivalents by converting crotonyl-CoA to butyryl-CoA and thus butyrate (Beaty & McInerney, 1987). It requires a syntrophic partner when growing on butyrate. *S. wolfei* under all growth conditions and *Syntrophus aciditrophicus,* another syntrophic bacterium, have high levels of acetylation, crotonylation and butyrylation (**Figure 3.1**). These levels appear to be higher than those in other model organisms (immunoblots not shown). Besides validating the presence of lysine acetylation, crotonylation and butyrylation with pan-specific antibodies, immunoblot data suggest the relative level of acylation as: crotonate co-culture ≈ crotonate pure culture > butyryate co-culture; butyrylation in butyrate co-culture > crotonate co-culture ≈ crotonate pure culture; and crotonylation in butyrate co-culture > crotonate pure culture ≈ crotonate co-culture.

### 3.4.2. Lysine acylation detection and semi-quantitation with mass spectrometry

Using acyl-CoA esters generated during crotonate and benzoate metabolism to predict for the presence of corresponding lysine acylations, we searched shotgun proteomic data acquired without any peptide enrichment and detected an average of 700, 300, 150, 13, and 4 sites of, respectively, acetylated, butyrylated, 3-hydroxy (OH) butyrylated, crotonylated and acetoacetylated lysine for each growth condition. Based on the number of modified peptides detected and from the areas of their precursor ions, the relative levels of acylation are: acetylation > butyrylation > 3OH butyrylation > crotonylation ≈ acetoacetylation for all three cultures. More than half of the modified peptides are shared across the three different cultivations. Using the same calculation as above, the relative acylation levels are estimated as: acetylation in crotonate

co-culture ≈ crotonate pure culture > butyrate co-culture; butyrylation in butyrate co-culture > crotonate co-culture > crotonate pure culture; and 3OH-butyrylation in crotonate pure culture > crotonate co-culture > butyrate co-culture. Thus all of the predicted modifications were detected at different levels and corresponded to the different growth conditions (**Figure 3.2**).

Characteristic fragment ions can help identify mass spectra containing acylations unambiguously and improve peptide spectrum matching. Characteristic immonium ions for acetylation are *m/z* 126 (strong intensity and specific) and 143 (medium intensity and non-specific) (Trelle & Jensen, 2008). From structural analogs to these immonium ions, we predicted immonium ions for other acylations and observed them all (**Table 3.1**). Analogs to *m/z* 126 have higher intensity. Ion intensity usually decreases if acylated lysines are closer to the C-terminus.

### 3.4.3. Acylated proteins participate in diverse processes

Pathway analysis of all detected proteins and of acylated proteins showed the same relative abundance trend (**Figure 3.3**). Highly abundant metabolic enzymes have the highest level of acylation, followed by proteins involved in translation.

**Table 3.2** shows further analysis of the data for proteins with high and low tendency for lysine acylation. Highly acylated proteins are the ones with many acylated peptides. Weakly acylated proteins are *strongly and moderately expressed* proteins with few numbers of acylated peptides. It seems like membrane-associated proteins have more acylations and membrane-protected proteins (i.e. integral membrane proteins) or cytoplasmic proteins have less acylations

### 3.4.4. Acylation site analysis

Sequence motifs for lysine acylation in *S. wolfei* show glutamate enrichment (**Figure 3.4**). Sequence motifs for low stoichiometry acylation in this organism also show glutamate enrichment, explainable by the fact that most acylated sites have low stoichiometry (63% sites have less than

10% stoichiometry). Highly acylated sites in *S. wolfei* also tend to have positively charged lysine and sometimes arginine as their neighbors, as displayed in the example shown in **Figure 3.5**.

## 3.5. Discussion

### 3.5.1. Lysine acylation as a new means to elucidate metabolism in syntrophic bacteria

Our proteomic data show that all the acyl-CoAs from the central carbon metabolism seem to yield the corresponding set of lysine acylations. Previous studies, as described in **Table 3.3**, showed that the extent of lysine acylation correlates with the amount of acylating reagent available, strengthening the hypothesis that we could use acyl-CoA profiles to simultaneously predict lysine acylations and confirm MS-detected peptide modifications on lysine acylations as predicted and not as another type of modification.

Most lysine acylation studies to date enrich acylated peptides first with pan-specific acyllysine antibodies before mass spectrometric analysis (Kosono et al., 2015)(Hiroki Okanishi, Kwang Kim, Ryoji Masui, 2014)(Kosono et al., 2015). Those pan-specific antibodies are supposed to recognize acyllysines regardless of the surrounding amino acids. The antibodies are usually developed against a single chemically acylated protein such as BSA (bovine serum albumin) or KLH (keyhole limpet hemocyanin) (http://ptm-biolab.com/). Although the acylated proteins are large, the number of acylated sites are still limited and the antibodies can still have sequence bias. Our proteomic approach to study lysine acylations in *S. wolfei* bypasses such sequence bias and allows simultaneous detection and semi-quantitation for all (still predicted) lysine acylations and for modification stoichiometry. Thus it should be recommended in proteomes containing complex and high levels of lysine acylations.

*S. wolfei* from butyrate co-culture, crotonate co-culture and crotonate pure culture all have high level of butyrylation, indicating the presence of butyryl-CoA. *S. wolfei* grown in crotonate pure

culture produces butyrate to dispose reducing equivalents (Beaty & McInerney, 1987). The organism probably converts a part of the activated crotonyl-CoA to butyryl-CoA and then butyrate in this condition. *S. wolfei* grown in crotonate co-culture does not have to dispose of electrons this way because it has the syntrophic partner, *Methanospirillum hungatei*, to remove end products. Butyrate is barely detected in this condition (Mcinerney & Wofford, 1992). However the presence of butyryl-CoA in *S. wolfei* crotonate co-culture suggests that the organism can flexibly choose between oxidizing crotonate directly to acetate and reducing crotonyl-CoA to butyryl-CoA before the oxidation process, probably depending on the electron removal rate of the syntrophic partner and the abundance of crotonate substrate. Thus studying lysine acylations might suggest a new means to elucidate substrate metabolism in syntrophic bacteria.

Syntrophic bacteria, the linchpins to methanogenesis and carbon cycling, survive on mostly short chain fatty acids and aromatic compounds, which need to be activated with coenzyme A in anaerobic environments (Fuchs, Boll, & Heider, 2011). Thus they should produce a wide range of acyl-CoA metabolites at high level during substrate degradation. (Some amino acid metabolism pathways produce acyl-CoA metabolites too, but those are probably produced at much lower level). The work in this dissertation has successfully revealed lysine acylations in two model syntrophic bacteria, *Syntrophomonas wolfei* and *Syntrophus aciditrophicus* (**Chapter 4**), and demonstrated how the modifications can help understand substrate degradation routes. The analytical mass spectrometric approach should be applicable to syntrophic bacteria having more complex substrate degradation pathways.

### 3.5.2. The tight link between lysine acylation level and acyl-CoA level

Direct measurement of *Syntrophomonas wolfei* primary metabolites have not been done to date, but similar measurements for another syntrophic bacterium, *Syntrophus aciditrophicus,* have been reported (Mouttaki, Nanny, & McInerney, 2007)(Elshahed, Bhupathiraju, Wofford,

Nanny, & Mcinerney, 2001). Measurements of acyl-CoA esters produced during substrate degradation were done by derivatizing the molecules prior to gas chromatography-mass spectrometry analysis. However the derivatizing reagent also reacted with the substrate salts, for example, crotonate and crotonyl-CoA are both labeled. Since obtaining metabolite quantitation was not easily performed for *S. aciditrophicus,* we did not obtain such data for *S. wolfei* and relied on published enzymatic activities to infer the level of acyl-CoA metabolites.

The level of acetylation > butyrylation > 3-OH butyrylation > crotonylation ≈ acetoacetylation in all three cultivates seem to be correlated to the level of the acylating reagents predicted from specific activities of β-oxidation enzymes (**Figure 3.6**). First, the specific activity of acetyl-CoA acetyltransferase converting acetyl-CoA to acetyl phosphate is high while that of acetate kinase, which consumes acetyl phosphate, is low. This likely leads to a high accumulation of acetyl phosphate, which influences the level of acetylation more than acetyl-CoA (B. Weinert et al., 2013), and thus yields very high level of acetylation. Second, the rate of converting butyryl-CoA to crotonyl-CoA in butyrate culture is very low, probably leading to the accumulation of butyryl-CoA and thus high levels of butyrylation. Third, 3OH-butyryl-CoA dehydratase efficiently converts 3OH-butyryl-CoA to crotonate while the consuming rate of 3OH-butyryl-CoA is much lower, possibly leading to an adequate amount of 3OH-butyryl-CoA and thus 3OH-butyrylation levels. Fourth, the level of crotonyl-CoA is low because of the very efficient 3OH-butyryl-CoA dehydratase and the low efficiency of acyl-CoA dehydrogenase and crotonyl-CoA transferase.

The higher acetylation level in *S. wolfei* crotonate-grown cultures seems to be the result of the higher ratio of phosphoacetyltransferase to acetate kinase activity: 47 to 1 in crotonate pure culture and 9 to 1 in butyrate co-culture (Wofford, Beaty, & McInerney, 1986)(Mcinerney & Wofford, 1992). The higher level of 3-OH-butyrylation in crotonate pure culture is probably because *S. wolfei* produces more D-3OH-butyryl-CoA for more polyhydroxyalkanoates (PHA) synthesis (Amos & McInerney, 1991). The higher level of butyrylation in butyrate culture is

probably because of the efficient butyrate activation and the unfavorable butyryl-CoA oxidation leading to butyryl-CoA accumulation. In other words, the levels of lysine acylation in *S. wolfei* seem to correlate to cultivation conditions and to the levels of the corresponding acyl-CoA esters.

### 3.5.3. A hypothesis on the nature of lysine acylations

Lysine acylation is an ancient post-translational modification (PTM) family. Acyl-CoAs, reactants for lysine acylations, have their coenzyme A synthesized from prebiotic pantothenate and pantetheine (Genschel, 2004). The central role of CoA in metabolism is presented at a very early stage of the evolution of life (Genschel, 2004). The evolutionary conservation of acetylated lysines is higher than that of non-acetylated lysines and acetylated residues are more tightly conserved than phosphorylated ones (Papanicolaou et al., 2014). Lysine acylation regulators, N-ε lysine acetyltransferases, $NAD^+$-independent deacetylases, and sirtuins ($NAD^+$-dependent deacetylases), are members of ancient enzyme families found in plants, animals, and fungi as well as archaebacteria and eubacteria (Yuan & Marmorstein, 2013)(Gregoretti, Lee, & Goodson, 2004). *In the ancient world, lysine acylation probably happened spontaneously at first and then lysine acylation regulators were evolved to fine-tune the acylation process and effect.*

Analyzing lysine acylation motifs seems to validate such evolution trend. First, we will review the acylation motifs found in *S. wolfei* and other studies. *S. wolfei* acylation sites enriched for glutamate are abundant and at low stoichiometry; high stoichiometry sites are enriched for the positively charged lysine neighbors. Independent work by Baeza *et al* measured the reactivity of lysines within some well-known metabolic enzymes with acetyl phosphate and acetyl-CoA *in vitro* (Baeza, Smallegan, & Denu, 2015). Their work showed that lysines engaged in electrostatic interactions with acidic residues have lower reactivity and lysines close to positively charged lysine and arginine are highly reactive. They also showed that highly reactive lysines measured *in vitro* are more likely to be detected in an *in vivo* setting. Thus there is a correlation among

glutamate enrichment, low stoichiometry, and low reactivity and another correlation among lysine enrichment, high stoichiometry and high reactivity.

The glutamate enrichment trend detected in *S. wolfei* is similar to acetylation and propionylation sequence motifs observed in the primitive prokaryotes and proteobacteria-originated mitochondria (Hiroki Okanishi, Kwang Kim, Ryoji Masui, 2014)(Okanishi, Kim, Masui, & Kuramitsu, 2013)(Jeffers & Sullivan, 2012). This is different from eukaryotic, especially non-mitochondrial, acetylation sites whose neighbors are enriched for patches of lysines (B. T. Weinert et al., 2011)(Miao et al., 2013). The eukaryotic p300/CBP lysine acetyltransferase family, besides recognizing histones with many positive charge lysines, also acetylases cytosolic proteins and recognizes lysines having positive charge residues around them (L. Wang, Tang, Cole, & Marmorstein, 2008). Thus it is likely that eukaryotes evolved to have lysine acetyltransferases to efficiently acetylase lysines around positive charge amino acids in the cytosol and nucleus, where non-enzymatic acetylation is not probable. This probably only happens when cytosolic and nuclear acetyl-CoA reaches a certain level sensed by lysine acetyltransferases, as most of those are only active after acetylasing themselves (L. Wang et al., 2008)(Friedmann & Marmorstein, 2013).

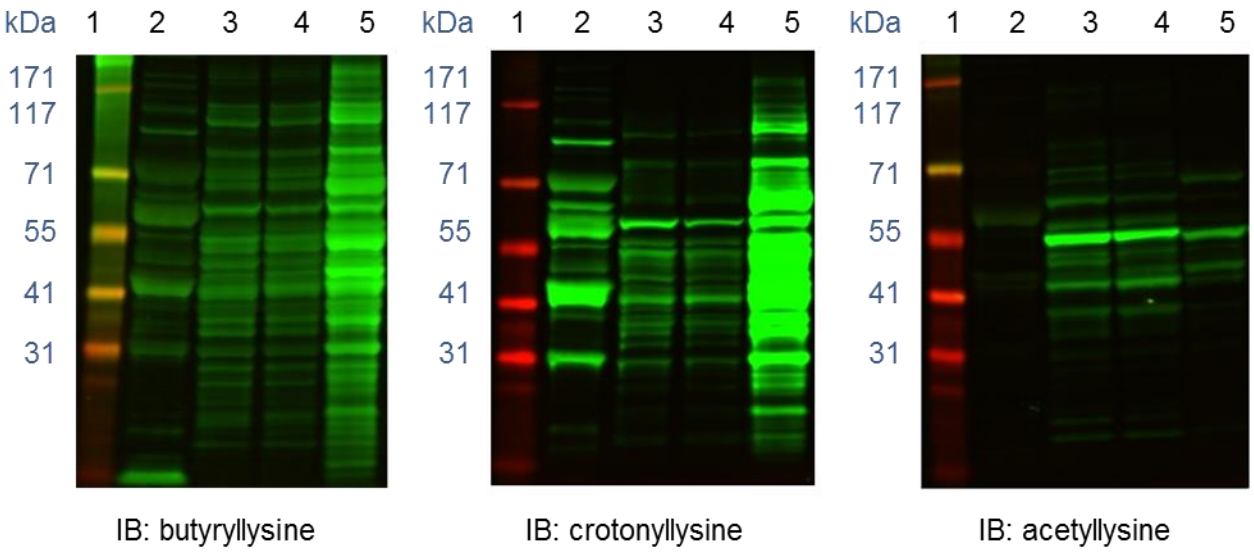
The discovery of secondary functions for metabolic enzymes might also give more insights into the nature of lysine acylation. Acetyl-CoA acetyltransferase, which catalyzes the reversible breakdown of acetoacetyl-CoA into two molecules of acetyl-CoA, can acetylase pyruvate dehydrogenase and pyruvate dehydrogenase phosphatase complexes (Montgomery et al., 2015). Alpha-ketoglutarate dehydrogenase complex, which reversibly converts α-ketoglutarate to succinyl-CoA, can succinylase many enzymes including other enzymes of the tricarboxylic acid cycle (G. E. Gibson et al., 2015). Both of these recognize specific acyl-CoA substrates and form S-acyl intermediates before transferring the acyl groups onto N-ε on lysine (Kursula, Ojala,

Lambeir, & Wierenga, 2002)(G. E. Gibson et al., 2015). These two proteins might suggest a raw and simple way to acylase proteins in mitochondria and prokaryotes.

### 3.5.4. Lysine acylation may help regulate central metabolism in *S. wolfei*

Non-acetyltransferases have not been identified in prokaryotes. *S. wolfei* acylation regulators (based on genome annotation), lysine acetyltransferase Swol_1812, lysine deacetylases Swol_0976 and 2171 and sirtuin Swol_1033, are weak-to-moderately detected in our proteomics data. Having attributes of non-enzymatic modifications (abundant, low stoichiometry and correlate to the modifying reagent level), some *S. wolfei* acylations could happen spontaneously as in mitochondria. An emerging class of metabolic enzymes functioning as acyltransferases (Montgomery et al., 2015)(G. E. Gibson et al., 2015) could also be responsible for lysine acylation.

Lysine acylations in *S. wolfei* originate directly and proportionately from substrate-degradation primary metabolites. Removing lysine acylations can be done by sirtuins (Desuccinylation et al., 2013)(Wagner & Hirschey, 2014)(AbouElfetouh et al., 2014), whose cosubstrate NAD+ also functions as an indicator of cellular redox status. Thus lysine acylations in *S. wolfei* might represent a low cost method for sensing and responding to the metabolic flux and nutrient availability.




1. Molecular weight standards
2. *S. aciditrophicus* from benzoate/crotonate pure culture
3. *S. wolfei* from crotonate pure culture
4. *S. wolfei* from crotonate co-culture with *M. hungatei*
5. *S. wolfei* from butyrate co-culture with *M. hungatei*

**Figure 3.1:** Western blotting data show that lysine crotonylation and butyrylation levels are higher in *S. wolfei* growing on butyrate, whereas lysine acetylation level are higher in cells growing on crotonate.

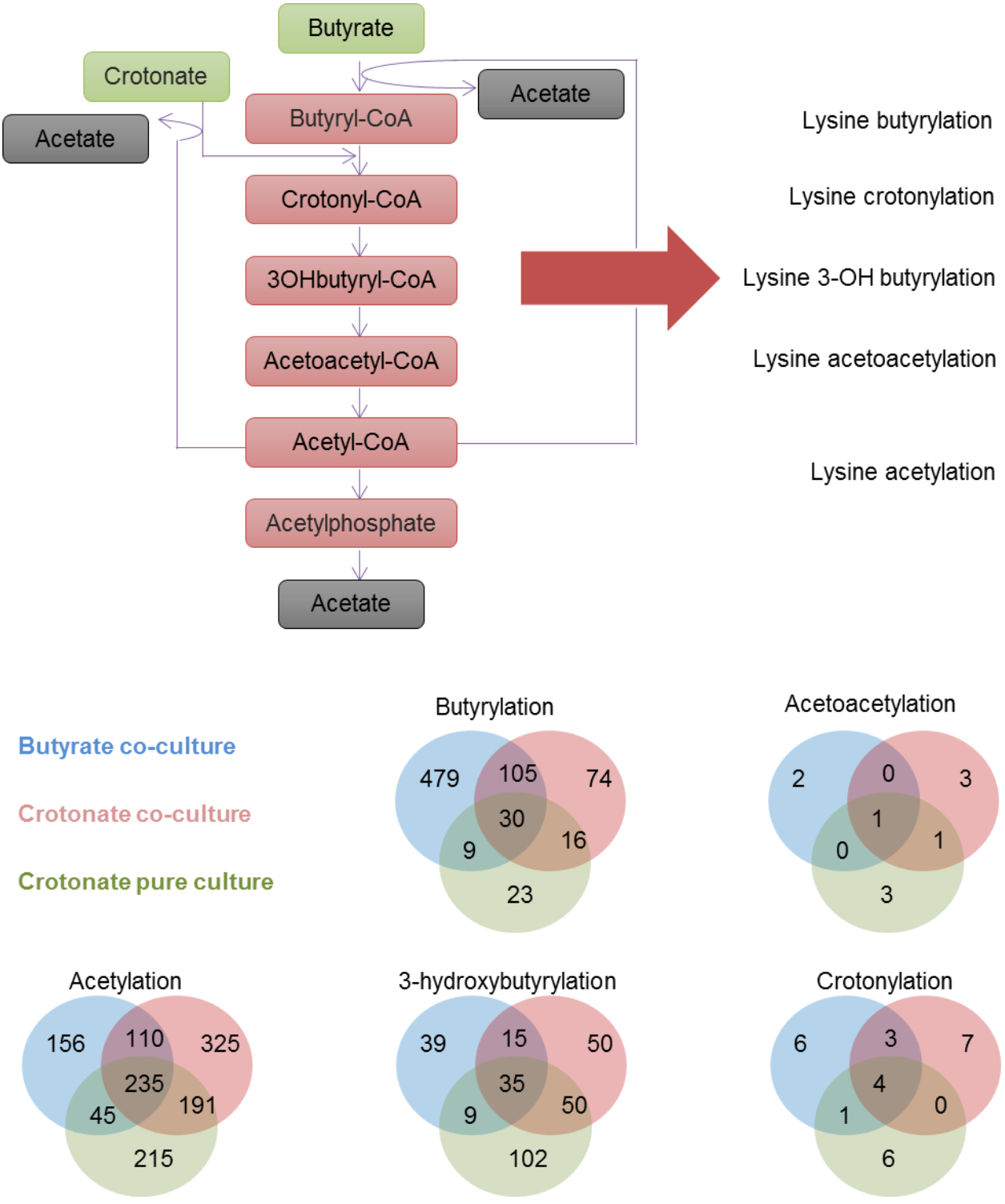



**Figure 3.2:** (Top) Butyrate and crotonate degradation pathway in *S. wolfei*, and generated acyl-CoA metabolites were used to predict for lysine acylations in this organism. (Bottom) Number of unique peptides containing different lysine acylations from proteomes of cells grown in butyrate co-culture, crotonate co-culture or crotonate pure culture.

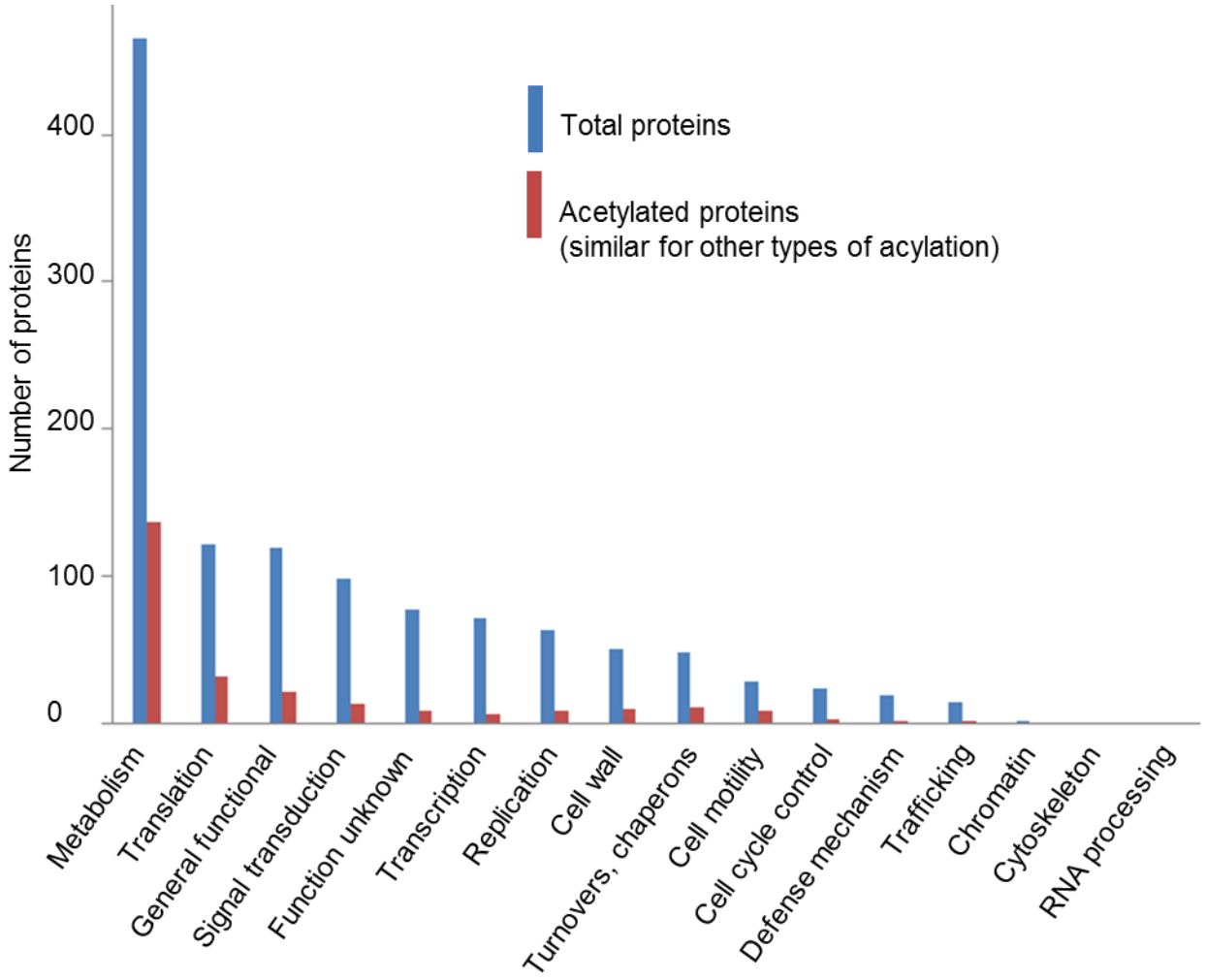
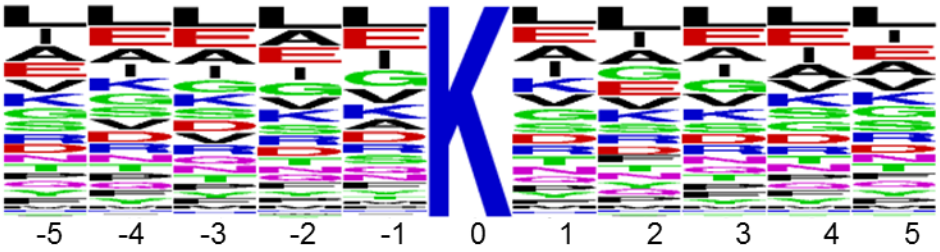



**Figure 3.3:** Pathway analysis of all detected proteins and acetylated proteins. A similar trend is also observed for the other types of acylations.

*S. wolfei* all Lys
- 44,797 sequences




*S. wolfei* acetyled Lys
– 881 sequences

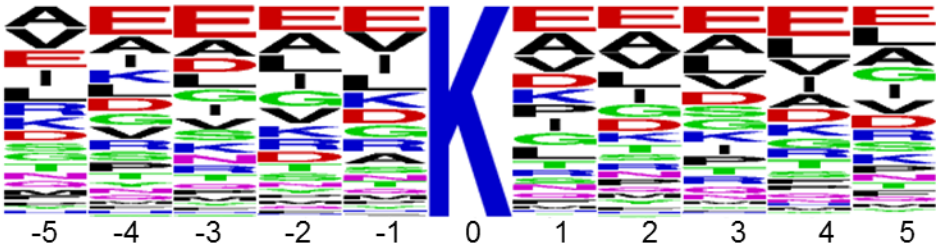



**Figure 3.4:** Lysine acetylation motif in *S. wolfei.* Butyrylated and 3-hydroxybutyrylated lysines also show a similar glutamate-enrichment motif.

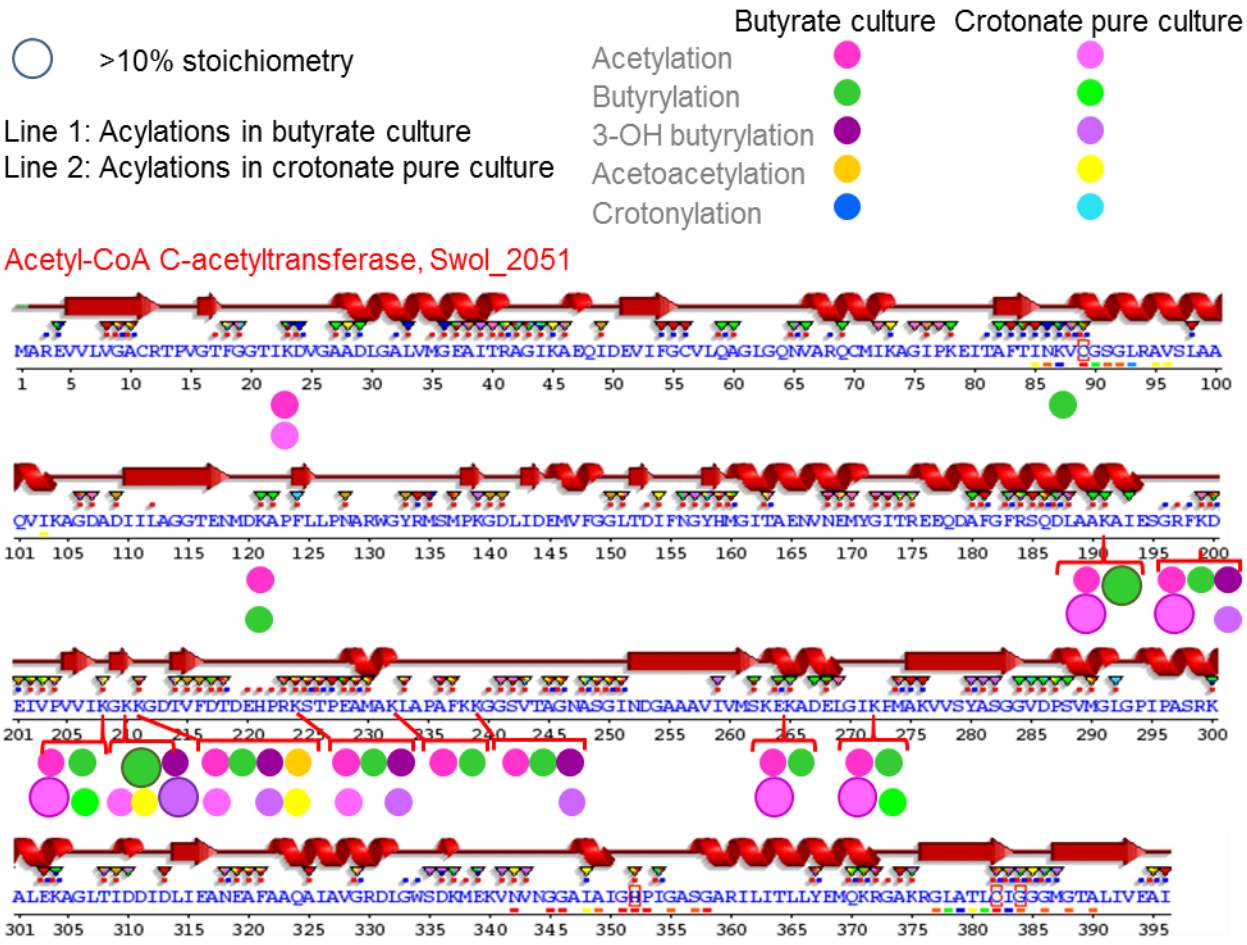



**Figure 3.5:** Mapping lysine acylation sites onto a representative protein that has many lysine acylations.

| Enzyme | Enzymatic activity | | |
|---|---|---|---|
| | But | Cro | Pure |
| 4-OHbutyryl-CoA transferase | 16.6 | 0.01 | 0.9 |
| crotonyl-CoA transferase | | 14.7 | 0.8 |
| acyl-CoA dehydrogenase | 2.9 | | 2.6 |
| 3OHbutyryl-CoA dehydratase | 224 | | 110 |
| 3OHbutyryl-CoA dehydrogenase | 10.2 | | 28.4 |
| acetyl-CoA C-acetyltransferase | 7.9 | | 6.2 |
| phosphate acetyltransferase | 11.9 | | 143 |
| acetate kinase | <1.1 | | 3.4 |

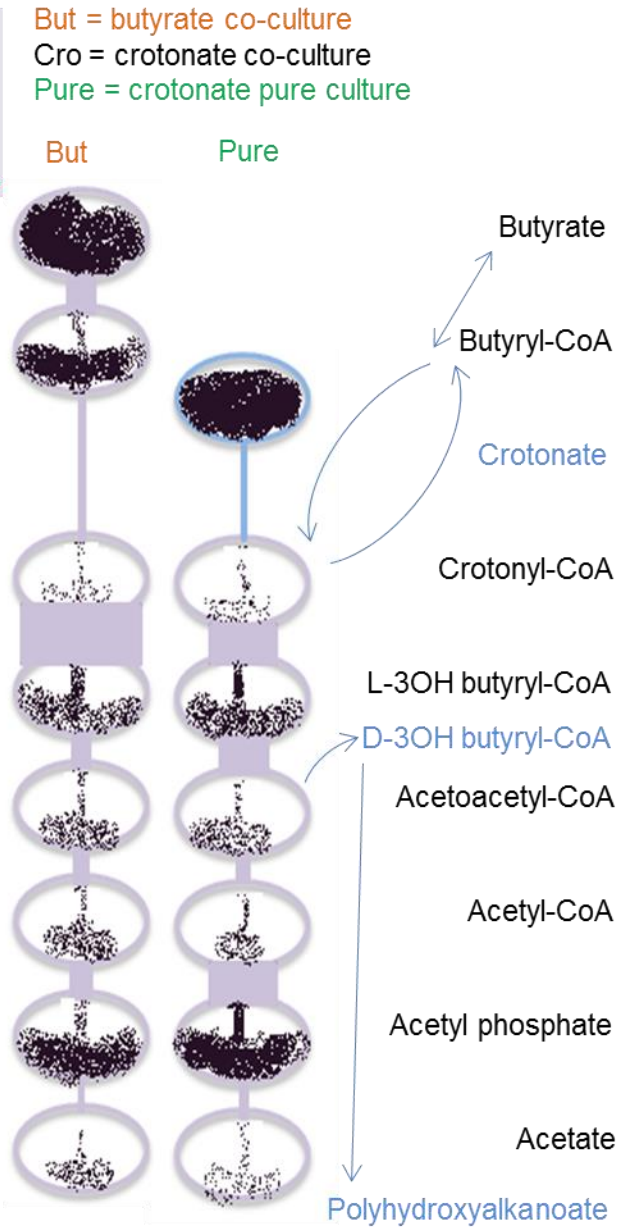



**Figure 3.6:** Specific activities (µmol/min per mg of protein) of β-oxidation enzymes from the cell extract (Mcinerney & Wofford, 1992)(Wofford, Beaty, & McInerney, 1986) and the predicted level of corresponding acyl-CoAs. The relative enzymatic activities are represented in the widths of the "necks" of the "vessels" and the relative amounts of acyl-CoAs are represented by the amounts of sand.

**Table 3.1:** List of characteristic immonium product ions and ion intensities for observed acylation-specific product ions

| **Acylation type** | **Immonium ions (*m/z*)** | **Ion intensity** |
|---|---|---|
| Acetylation | 143.118, 126.091 | 126: strong, 143: medium |
| Butyrylation | 171.149, 154.123 | 154: strong, 171: medium |
| 3-OH butyrylation | 187.144, 170.117, 169.133, 152.107 | 187, 170: medium, 169, 152: weak |
| Crotonylation | 169.133, 152.107 | 152: medium, 169: weak |
| Acetoacetylation | 185.128, 168.102 | 185: medium, 168: weak |

**Table 3.2:** List of proteins with highest and lowest tendency for acylation

| Protein type | Highly acylated | Weakly acylated |
|---|---|---|
| Surface displayed proteins | **ALL** *highly expressed* copper amine oxidase and S-layer proteins are highly acylated | **SOME** *moderately expressed* copper amine oxidase and S-layer isoforms are weakly acylated |
| Stress-induced proteins | **ALL** (chaperone DnaK, DnaJ, Hsp20, GroEL, GroES, alkaline shock and stress induced protein) are highly acylated | |
| Reverse electron transport proteins | **ALL** cytosolic ATP synthase subunits and **SOME** oxidoreductase, ETF, dehydrogenase, and hydrogenase are highly acylated | **SOME** oxidoreductase, ETF, dehydrogenase, and hydrogenase isoforms are weakly acylated |
| β-oxidation enzymes | **ALL** 4-OHbutyrate CoA transferase isoforms, **ALL** acyl-Col A dehydrogenase isoforms, **ALL** acetyl-CoA C-acetyltransferase isoforms, phosphate acetyltransferase, acetate kinase, **SOME** 3OH butyryl-CoA hydratase, and **SOME** 3OH butyryl-CoA dehydrogenase are highly acylated | **SOME** 3OH butyryl-CoA hydratase and 3OH butyryl-CoA dehydrogenase isoforms are weakly acylated |
| tRNA synthetase | **ALL** S, M, **W**, D, P, **I, T**, F, G, **L** and **Q** tRNA synthetase isoforms are highly acylated | **SOME** K tRNA synthetase isoforms are weakly acylated |
| Ribosomal subunits | **MANY** 30S subunits and **SOME** 50S subunits (S3, S4, S5, S13, S17, L7/L12, L19) are highly acylated | **MANY** 50S subunits and **SOME** for 30S subunits (S15, S21, L6, L11, L15, L18, L30, L35, L36) are weakly acylated |
| Other translational proteins | Elongation factors such as EF-Ts, EF-Tu, EF-G are highly acylated | |
| Chemotaxis proteins | **ALL** *membrane-associated* chemotaxis proteins (methyl-accepting chemotaxis protein) are highly acylated | **ALL** *soluble* chemotaxis proteins (some histidine kinase CheA, *all* response regulator CheY, "adaptor" protein CheW, methyl-transferase CheR) are weakly acylated |
| Transporter | **None, e**xcept for $K^+$ transporter, are highly acylated | **ALL** other transporter are weakly acylated |
| Other integral and outer membrane proteins | | **ALL** *externally oriented* formate dehydrogenase and cytochrome are weakly acylated |
| DNA-associated proteins | | **ALL** DNA binding, replication and transcription proteins are weakly acylated |

**Table 3.3:** The correlation between acylation level and acyl-CoA level

| Organism | Growth condition | Observation | Possible explanation |
|---|---|---|---|
| *S. cerevisiae* | Mutate the gen converting acetyl-CoA to citrate. Grow on glucose substrate | Acetyl-CoA level increases ~1.8 fold. Acylation level increases 1.8 fold (B. T. Weinert et al., 2014) | Cells are unable to convert acetyl-CoA to citrate (B. T. Weinert et al., 2014) |
| Patient fibroblast | Malonyl-CoA decarboxylase deficient cells<br>Or propionyl-CoA carboxylase deficient cells | Increase lysine malonylation (Pougovkina, Te Brinke, Wanders, Houten, & de Boer, 2014)<br>Or increase lysine propionylation (Pougovkina, Te Brinke, Wanders, et al., 2014) | Dysfunction in metabolizing acyl-CoA resulting in acyl-CoA accumulation (Pougovkina, Te Brinke, Wanders, et al., 2014) |
| Mice | Short-chain acyl-CoA dehydrogenase deficient cells | Increase lysine butyrylation (Pougovkina, Te Brinke, Wanders, et al., 2014) | Dysfunction in metabolizing butyryl-CoA resulting in butyryl-CoA accumulation (Pougovkina, Te Brinke, Wanders, et al., 2014) |
| Mice | Fasting | Increase lysine acetylation (Pougovkina, Te Brinke, Ofman, et al., 2014) | Increase β-oxidation that produces acetyl-CoA (Pougovkina, Te Brinke, Ofman, et al., 2014) |
| Mice | Fasting, very-long chain acyl-CoA dehydrogenase deficient | No change in lysine acetylation level (Pougovkina, Te Brinke, Ofman, et al., 2014) | Defective β-oxidation and not producing acetyl-CoA (Pougovkina, Te Brinke, Ofman, et al., 2014) |
| Cells and tissues | Exposure to high concentration of sodium glutarate, diet rich in tryptophan and glutaryl-CoA dehydrogenase deficiency | Increase lysine glutarylation (Tan et al., 2014) | Glutarate is activated to glutaryl-CoA.<br>Lysine and tryptophan degradation pathway yields glutaryl-CoA. Dysfunction in metabolizing glutaryl-CoA resulting in glutaryl-CoA accumulation (Tan et al., 2014) |
| *S. cerevisiae* | Grow on lactose or mutate enzymes in TCA cycle | Increase lysine succinylation (B. T. Weinert et al., 2013) | Increase succinyl-CoA produced in the mitochondria (B. T. Weinert et al., 2013) |
| *E. coli* | Supplement with pyruvate, glucose or succinate | Increase lysine succinylation in the order of adding pyruvate < glucose < succinate (Colak et al., 2013) | Pyruvate participates in the TCA cycle which generates succinyl-CoA. When glucose is added, glycolysis increases, adding more pyruvate to the TCA cycle. Succinate directly increases the amount of succinyl-CoA in the TCA cycle. Succinate directly increases the amount of succinyl-CoA in the TCA cycle |
| *E. coli* | Growth-arrested cells undergo glycolysis<br>Or enzymes participating in synthesis and degradation of acetyl phosphate are mutated | Increase lysine acetylation (Verdin & Ott, 2013) | Increase acetyl phosphate (Verdin & Ott, 2013) |

# CHAPTER 4: STUDYING LYSINE ACYLATION AS A NEW MEANS TO ELUCIDATE SUBSTRATE METABOLISM IN *SYNTROPHUS ACIDITROPHICUS*

## 4.1. Abstract

Benzene rings, the second most abundant carbon structure after glucose, are found abundantly as building units of plant lignin, crude oils and proteins and in industrial waste. After all, only amino acids can be metabolized limitedly by plants and animals; the other aromatic compounds must rely on microbial degradation and may pose as prominent environmental pollutants. Aromatic compound degradation pathways can be divided into 2 main groups: one follows the pathway of photoheterotrophic *Rhodopseudomonas* bacterium and the other follows the pathways of facultative anaerobic *Thauera-Azoarcus* (and strictly anaerobic *Geobacter*) bacteria. *Syntrophus aciditrophicus,* one of the strictly anaerobic syntrophic bacteria participating in the "assembly-line" to cycle organic matter back into the environment, has genes supporting the *Thauera-Azoarcus* pathway. Published enzymatic activities of *S. aciditrophicus* cell extracts, however, seem to support the *Rhodopseudomonas* pathway. The emerging field of lysine acylation has shown the tight correlation between the presence and level of acyl-CoA metabolites and the corresponding lysine modifications. Our study for *S. aciditrophicus* lysine acylations showed many lysine acylations (most are first detected in any organism) resulting from intermediates of aromatic degradation, providing additional evidence for the parallel presence of both aromatic degradation pathways in this organism. Having both options available likely provides an additional mechanism to maximize ATP production in such organisms living at the limit of thermodynamic possibility.

## 4.2. Introduction

Aromatic compounds, found in lignin, flavonoids, quinones, aromatic amino acids, or constituents of fossil fuels, occur widely in nature; lignin, indeed, is the second most abundant organic polymer after cellulose. Because of the stabilizing resonance energy of the benzene ring system, aromatic compounds have become valuable for industrial usage. The persistence of these compounds, however, also pose a major environmental problem because many of them are toxic and most cannot be degraded by plants and animals (Carmona et al., 2009). Their degradation mostly relies on microbes. Aerobic organisms use $O_2$ as a powerful means to oxidize and cleave the aromatic ring. Anaerobic organisms lack $O_2$ and usually convert substrates to the universal benzoyl-CoA to facilitate the transfer of electrons to the ring for cleavage (Fuchs et al., 2011).

Benzoate, the model compound used to study the major central pathway for anaerobic aromatic compound degradation, can be catabolized anaerobically via two major pathways used by the facultative anaerobic phototrophic α-Proteobacterium, *Rhodopseudomonas palustris*, and the facultative anaerobic denitrifying *Thauera aromatica* (Fuchs et al., 2011) (**Figure 4.1**). These two pathways are different from each other only by the enzymes involved and the number of steps taken to cleave the ring into the 7-carbon compound 3-hydroxypimeloyl-CoA. Both organisms activate benzoate by an ATP-dependent benzoate-CoA ligase, consuming an equivalent of two ATPs as the released pyrophosphate is further hydrolyzed to inorganic phosphate. They drive benzoyl-CoA reduction with a stoichiometric ATP hydrolysis catalyzed by benzoyl-CoA reductases, adding *up to four ATP units hydrolyzed* before the aromatic ring is cleaved. Following ring cleavage, β-oxidation yields one molecule of $CO_2$ and three molecules of acetyl-CoA. *R. palustris* uses photosynthesis to supply ATP for benzoate metabolism and acetyl-CoA functions as building blocks for biosynthetic reactions rather than as an energy supply (J. Gibson & S Harwood, 2002). *T. aromatica* assimilates part of acetyl-CoA into cell material and oxidizes the

rest to $CO_2$, which produces energy by electron transport phosphorylation with nitrate as an electron acceptor. Strict anaerobes have been shown to follow the *Thauera* pathway, except that their benzoyl-CoA reductases are, and must be, ATP-independent (Fuchs, 2008). Since they use substrate-level phosphorylation, 3 molecules of acetyl-CoA only yield 3 ATP equivalents during the process. Two ATP molecules are used for benzoate activation and the last ATP has to be shared among regenerating oxidizing equivalents and cellular growth. With such little energy needed to support growth, how these strict anaerobes distribute and conserve energy is still understudied.

This study focuses on *Syntrophus aciditrophicus*, a Gram-negative strictly anaerobic syntrophic bacterium usually found in methanogenic microbiota. It grows on short-chain fatty acids and certain aromatic compounds (McInerney et al., 2007) generated by fermentative microbes in the "assembly-line" of cycling organic compounds back into the environment in the form of methane and carbon dioxide. There is genomic evidence that *S. aciditrophicus* has genes involved in the *Thauera* aromatic degradation pathway (McInerney et al., 2007) and the enzymes responsible for the *Thauera*-specific steps were expressed and confirmed (Peters, Shinoda, McInerney, & Boll, 2007). Earlier enzymatic activities from analysis of cellular extracts, however, suggested the *Rhodopseudomonas* aromatic degradation pathway to be present (Elshahed et al., 2001). Clearly there is a gap between what *S. aciditrophicus* is expected to do and what actually happens inside the cells.

Metabolic evidence for the aromatic degradation intermediates would be necessary to elucidate how the two pathways are involved in cellular activities. Previous studies used trimethylsilyl-labeling of *S. aciditrophicus* metabolites for GC-MS (gas chromatography – mass spectrometry) detection (Mouttaki et al., 2007). This labeling method, however, does not distinguish the salt and CoA ester forms. The over-abundance of substrate and the non-linear distribution of acyl-CoA metabolites could pose another problem for metabolite detection. Inspired

by recent discoveries of novel lysine modification analogues to lysine acetylation (termed “lysine acylation”) and by the close relationship between acyl-CoA esters and lysine acylations, we seek similar modifications in *S. aciditrophicus* to elucidate its aromatic degradation pathway(s).

## 4.3. Materials and Methods

**2D-PAGE proteomic dataset** *S. aciditrophicus* proteins from whole cell lysates were separated by two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) spanning pH 3-10 isoelectric points. Hundreds of gel spots were excised. Gel-embedded proteins were digested with trypsin and analyzed by LC-MS/MS. LC-MS/MS employed a quadrupole time-of-flight system (Applied Biosystems/Sciex QSTAR Pulsar XL mass spectrometer) equipped with a Dionex/LC Packings nano-LC system (Jupiter Proteo C12 resin, 75 µm x 150 mm).

**Protein extraction & digestion** *S. aciditrophicus* was cultured axenically on either crotonate or benzoate/crotonate mixture (Elshahed & McInerney, 2001). Cells were lysed and trypsin digested by the eFASP method (Erde et al., 2014). Recovered peptides were dried in a vacuum concentrator for subsequent immunoprecipitations and mass spectrometry measurement.

**High throughput proteomic dataset** Tryptic peptides were pre-fractionated using strong cation exchange. Fractions were analyzed by liquid chromatography-tandem mass spectrometry (LC-MS/MS) on an EASY nLC1000 (Thermo Scientific) coupled to a hybrid mass spectrometer (Q-Exactive, Thermo Scientific). Peptides from the eFASP sample processing (generated as described in the above paragraph) were desalted on an Acclaim PepMap100 C18 nano-trap column (Dionex, 75 um x 2 cm, Product # 164535) and separated on an Acclaim PepMap RSLC analytical column (Dionex, 75 um x 25 cm, Product # ES802).

**Immuno-enriched proteomic dataset** Tryptic peptides were incubated with anti-crotonyllysine or anti-glutaryllysine antibodies (PTM Biolabs, Product # PTM-502 and PTM-1151, Chicago, IL) conjugated on Dynabeads (Invitrogen, Product # 10003D and #10001D) in NET buffer (100 mM NaCl, 20 mM Tris, 0.5 mM EDTA, 0.005% Tween-20, pH 8.0). The bound complexes were washed three times with NET buffer before elution with 0.1 M citrate pH 3. Peptides were desalted with C18 StageTips (Rappsilber, Mann, & Ishihama, 2007) before mass spectrometry analysis as in the high throughput proteomic dataset.

**Immunoblotting** *Syntrophus aciditrophicus* grown axenically on benzoate/crotonate mixture and other organisms were lysed in NETN buffer (100 mM NaCl, 20 mM Tris, 0.5 mM EDTA, 0.4% NP-40, pH 8.0) in the presence of protease inhibitors (Sigma, Product #P8340, St. Louis, MO). Immunoblotting was performed with anti-crotonyllysine and anti-glutaryllysine primary antibodies (PTM Biolabs, Product # PTM-502 and PTM-1151, Chicago, IL). Secondary antibodies were purchased from LI-COR (LI-COR, Product # 827-08364 and 827-08365, Lincoln, NE) and the blots were imaged with a LI-COR Odyssey Infrared Scanner (LI-COR, Odyssey CLx).

**Database search** All of the acyl-CoA metabolites from aromatic compound degradation (**Figure 4.1**) were used to calculate the corresponding variable protein modification masses. Mass spectral data were exported into mgf format and searched with Mascot version 2.5 (Matrix Science, Boston, MA) against the *Syntrophus aciditrophicus* protein database downloaded from ftp://ftp.ncbi.nih.gov/genomes/ and protein sequences of common contaminants. Carbamidomethylation on cysteine was specified as a fixed modification. Oxidation on methionine and predicted lysine acylations were divided into groups of 7 (the maximum modification number Mascot allows) and specified as variable modifications. Trypsin/P was specified as a cleavage enzyme, and the search allowed for up to two missed cleavages. The precursor mass error was set to ±5 ppm and product mass error was set to ±0.01 Da. All modified peptides were confirmed by the presence of the characteristic acyllysine ions.

## 4.4. Results

### 4.4.1. Many types of lysine acylations were observed in highly fractionated (2D-PAGE) proteins

Two-dimensional protein separation on polyacrylamide gels greatly reduces protein sample complexity before data analysis (about three proteins per gel spot) and thus reduces the false discovery rate in the detection of novel protein modifications. Hypothesizing that a bacterium's pool of acyl-CoA esters could suggest potential acyl-lysine modifications hidden within MS/MS data, we searched this dataset for predicted lysine acylations and observed lysine crotonylation and glutarylation *for the first time in bacteria*. The modifications were confirmed from the characteristic product ions that are similar to lysine acetylation-specific ions (Trelle & Jensen, 2008)(J. Y. Kim, Kim, Kwon, Lee, & Yoo, 2002) and by immunoblotting (**Figure 4.2**).

We also observed 3-hydroxypimeloylated lysine in many peptides. Spectra containing this modification feature product ions with water loss from condensing the carboxylic and hydroxyl groups and with many other characteristic ions (**Figure 4.3**). 2,3-Didehydropimelylated lysine was also observed. Although this modification is a water-loss form of 3-hydroxypimeloylated lysine and peptide ions with a 3-hydroxypimeloylated lysine usually loses water inside the mass spectrometer, it must have arisen biologically. The same peptide with these 2 modifications eluted at different time points in a single chromatographic run, suggesting that the water loss happens before entrance to the mass spectrometer (**Figure 4.4**). 2,3-Didehydropimelylation, without the hydroxyl group, is more hydrophobic than 3-hydroxypimeloylation and thus was retained longer on the reversed-phase C18 column.

### 4.4.2. High throughput high mass accuracy shotgun proteomic data revealed a fuller set of lysine acylations

In this proteomic dataset, total proteins, without gel electrophoresis separation, were trypsin digested into peptides before being injected into the Q-Exactive instrument. This mass spectrometer has higher sensitivity and a faster detection rate than the one used for 2D-PAGE separated proteins, thus allowing for high throughput detection of peptides and peptide modification discovery. The high mass accuracy of the instrument distinguishes lysine acetylation (+42.0106 Da) from lysine trimethylation (+42.0470 Da) and makes novel modification discovery more reliable. We also relied on the detection of characteristic immonium ions for acyllysine to confirm the modifications (**Figure 4.5**).

Besides a large number of acetylated, 3-hydroxypimeloylated, crotonylated and glutarylated peptides, we detected more examples of 2,3-didehydropimeloyled peptides eluting later than the 3-hydroxypimeloylated peptides (**Table 4.1**) and discovered more novel lysine acylations. Benzoyl-, cyclohexane-1-carboxyl-, 3-ketopimeloyl-, and 3-hydroxybutyryl-lysine (as well as 3-hydroxypimeloyl-lysine) were *for the first time shown in any organism* (**Figures 4.6** to **4.9**).

### 4.4.3. Lysine acylations in *S. aciditrophicus* are regulated by two sirtuins

*Syntrophus aciditrophicus* contains two genes encoding for sirtuins: SYN_42 and SYN_1020. Both were cloned and expressed by the UCLA DOE Protein Expression Facility. Both could remove the crotonyl group from a synthetic peptide (**Figure 4.10**) and the crotonyl, glutaryl (**Figure 4.10**), and butyryl group from *S. aciditrophicus* proteins. They could remove malonylation, a modification not detected by immunoblotting in *S. aciditrophicus,* but on proteins from pea seeds poorly (data not shown). We did not test the deacylase activities of these sirtuins towards other lysine acylations due to the limitation to have synthetic peptides prepared and the requirement to

find available antibodies. In general, both putative sirtuins can remove crotonylation, glutarylation, and butyrylation. SYN_42 seems to remove negative charged groups better (crotonyl and butyryl) and SYN_1020 removes neutral groups (glutaryl) better. This trend is similar to the slight preference of SIRT3 for neutral groups (Bao et al., 2014)(M. D. Hirschey et al., 2010) and SIRT5 for negative charge groups (Tan et al., 2014)(Du et al., 2011).

## 4.5. Discussion

### 4.5.1. Lysine acylation analysis as a new means to establish metabolism

Lysine acylations have been shown to be tightly coupled to metabolic flux and nutrient availability (Papanicolaou et al., 2014). Hypothesizing that the acyl-CoA metabolite pools of an organism can suggest the presence of the corresponding lysine acylations, we detected lysine acetylation, crotonylation, glutarylation, 3-hydroxybutyrylation, 3-hydroxypimeloylation, 3-ketopimeloylation, 2,3-didehydro-pimeloylation, cyclohexane-1-carboxylation and benzoylation.

Our study showed the detection of lysine crotonylation and glutarylation for the first time in bacteria and these were confirmed with acyl-lysine characteristic ions and pan-specific antibodies. The other acylations are first seen in any organism and are confirmed with acyl-lysine characteristic ions only. These acylations, in turn, help confirm the presence of the corresponding acyl-CoA esters and suggest their relative abundance levels.

Using lysine acylation analysis to probe the acyl-CoA metabolite pool of an organism would be beneficial especially for studying syntrophic bacteria and other fatty acid and anaerobic aromatic degradation bacteria, which have to activate their substrate into the CoA forms. Syntrophic bacteria are essential for the complete conversion of natural polymers such as polysaccharides, proteins, nucleic acids, and lipids to $CO_2$ and $CH_4$ and are responsible for the rate-limiting step in global carbon recycling (Sieber et al., 2014). Understanding their biological

processes is important to understand methane production, improve anaerobic digesters for fuel production, and facilitate removal of toxic aromatic compounds.

Emerging evidence on lysine acylations regulating metabolism at the protein activities and protein expression levels (Q. Wang et al., 2010b) (Papanicolaou et al., 2014) (Sabari et al., 2015)(Sabari et al., 2015)(Kosono et al., 2015)(M. Hirschey & Zhao, 2015) suggests that analyzing acylated proteins can provide insight into metabolism regulation. This might provide important clues into the energy distribution, energy conservation and cellular response for organisms such as syntrophic bacteria living at the thermodynamic limit.

### 4.5.2. *Syntrophus aciditrophicus* may utilize both aromatic degradation pathways

Although genes homologous to enzymes specific to the *Rhodopseudomonas* type pathway were not detected in the *S. aciditrophicus* genome (Mouttaki et al., 2007), activities of these enzymes were detected in cell extract of *S. aciditrophicus* grown in benzoate co-culture with *Desulfovibrio* and/or in crotonate pure culture (**Table 4.4**) (Elshahed et al., 2001). Trimethylsilyl-derivatives of pimeloylate and 2-hydroxycyclohexane carboxylate were observed in crotonate/acetate pure culture and those were probably derived from the CoA forms because the culture broth contained low levels of those (Mouttaki et al., 2007). The presence of 2,3-didehydropimeloylation in our data suggests the presence of 2,3-didehydropimeloyl-CoA, a metabolite specific for *Rhodopseudomonas* type pathway. Together, the data suggest the steps of the *Rhodopseudomonas* type pathway are performed inside the cells.

SYN_01653, SYN_01654, and SYN_01655 were cloned and confirmed to be responsible for, accordingly, cyclohex-1,5-diene-1-carboxyl-CoA hydratase, 6-ketocyclohex-1-ene-1-carboxyl-CoA hydratase, and 6-hydroxycyclohex-1-ene-1-carboxyl-CoA dehydrogenase activities (Peters et al., 2007). Those steps are unique for the *Thauera* type pathway. Our high throughput

proteomic data show high expression levels for these proteins, suggesting this pathway is also performed inside the cells.

Pimelylate, together with cyclohex-1-1ene carboxylate and cyclohexane carboxylate are usually detected in *S. aciditrophicus* crotonate and benzoate/crotonate pure culture (Mouttaki et al., 2007) (Elshahed & McInerney, 2001), suggesting that reactions in the *Rhodopseudomonas* type pathway could be used to produce cyclohexane carboxylate (**Figure 4.11**), thus disposing reducing equivalents in the absence of a syntrophic partner. Having separate sets of enzymes for benzoate degradation (the *Thauera* pathway) and for disposing reducing equivalents (the *Rhodopseudomonas* pathway) could make regulating reactions in response to nutrient availability more efficient.

The co-occurrence of both pathways in an organism might be more prevalent than previously thought. *Ferroglobus placidus*, a hyperthermophilic archaeal organism, has genomic and transcriptomic evidence for the *Rhodopseudomonas* pathway (Holmes, Risso, Smith, & Lovley, 2012) and enzymatic activities of cellular extracts and some heterologously expressed proteins showed that the cells use the *Thauera* pathway (Schmid, René, & Boll, 2015). Joining the two pathways with enzymes converting between cyclohex-1,5-diene-1-carboxyl-CoA and cyclohex-1-ene-1-carboxylCoA for cyclohexane carboxylate production also seems to be popular among all cyclohexane carboxylate-degrading anaerobic bacteria, including denitrifying, Fe(III)-reducing, sulfate-reducing, and fermenting bacteria (Kung, Meier, Mergelsberg, & Boll, 2014).

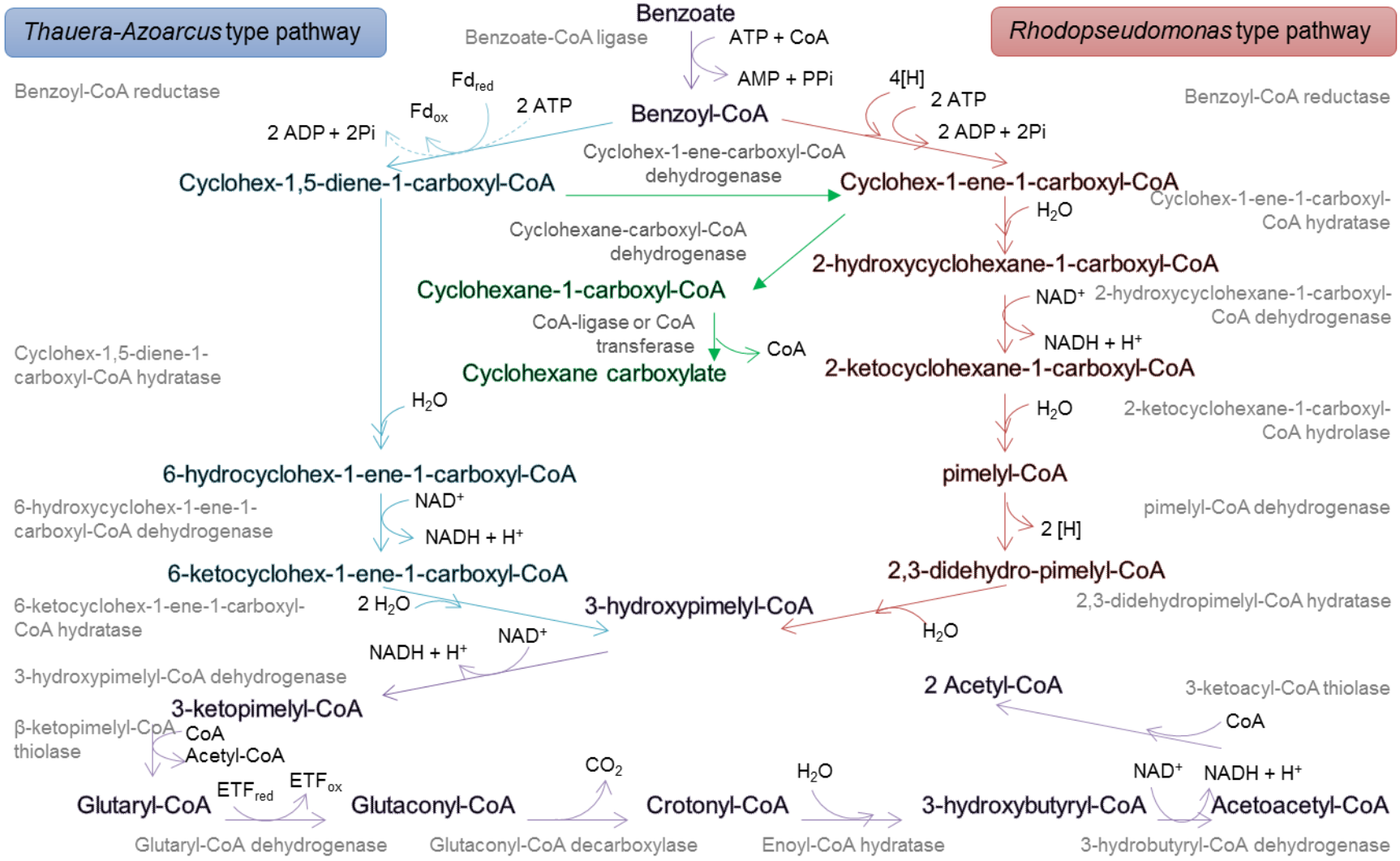



**Figure 4.1:** Two main aromatic degradation pathways in anaerobic organisms. Facultative anaerobic nitrogen-fixing *Thauera* and *Azoarcus* bacteria and strictly anaerobic bacteria use the pathway on the left (blue arrows). Facultative anaerobic photoheterotrophic *Rhodopseudomonas* bacterium uses the pathway on the right (red arrows). Some strict anaerobes can convert cyclohex-1,5-diene-1-carboxyl-CoA to cyclohexane carboxylate (green arrows).

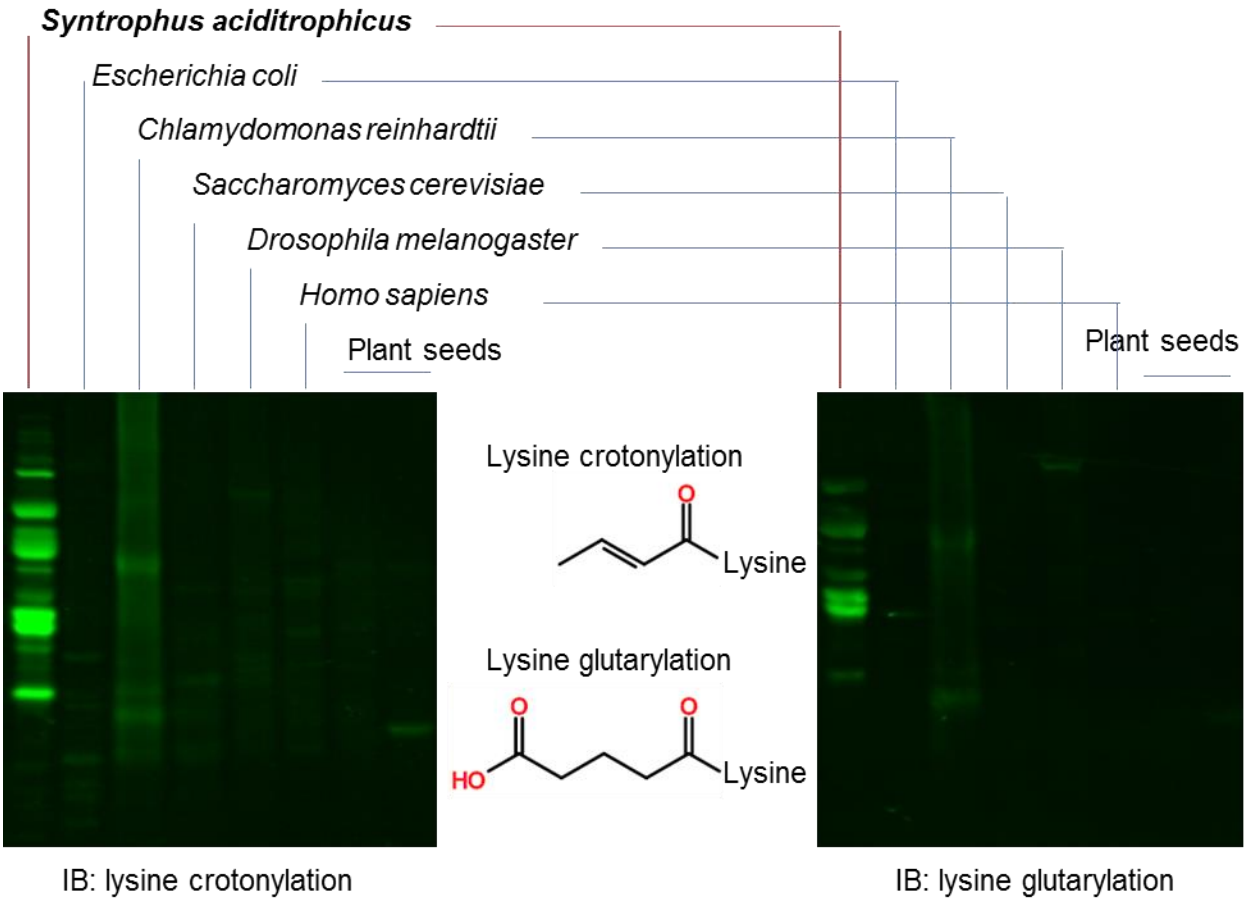



**Figure 4.2:** Immunoblotting using pan-specific crotonyl- and glutaryl-lysine antibodies confirmed the presence and prevalence of lysine crotonylation and glutarylation in *S. aciditrophicus*.

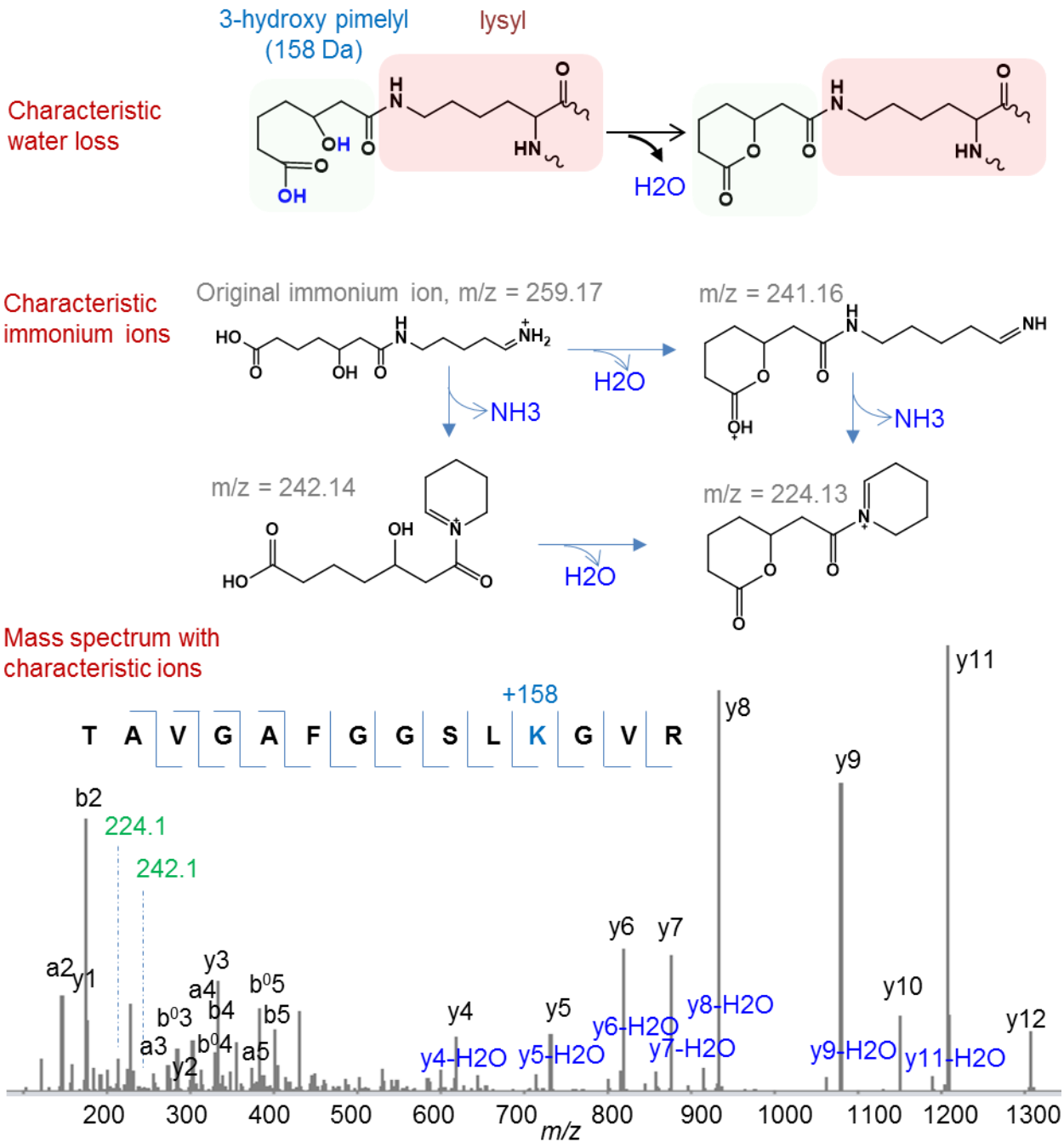
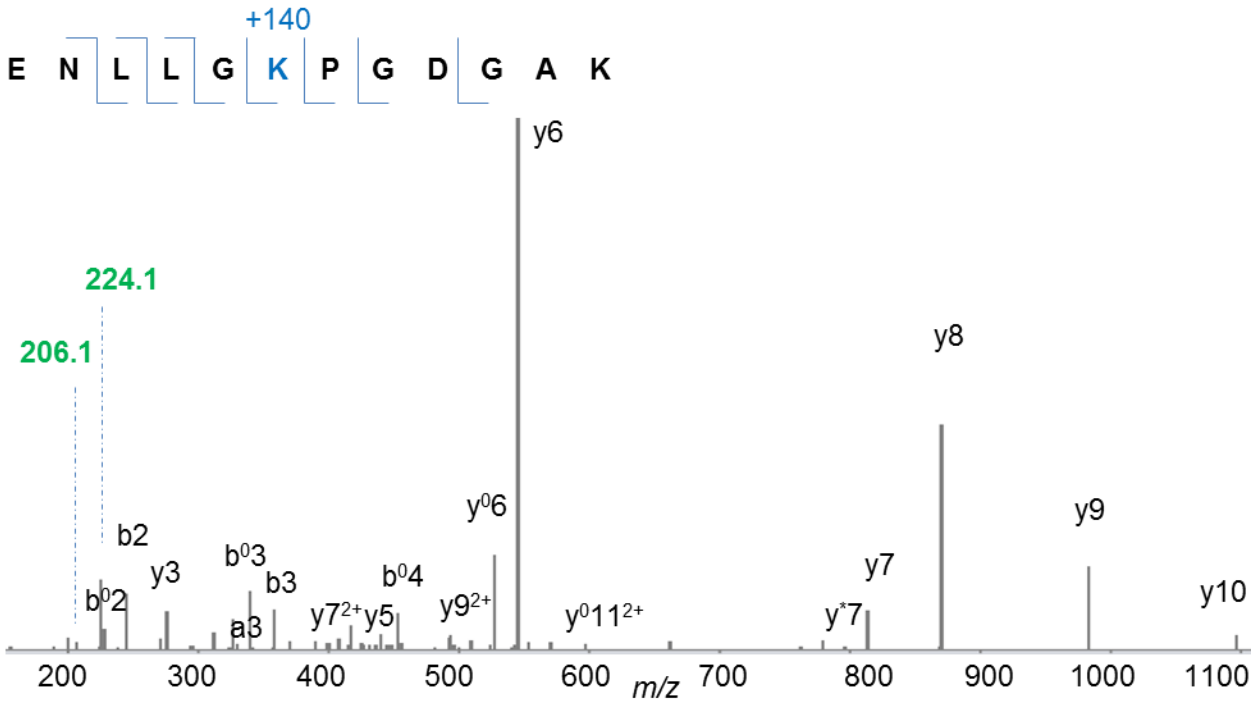



**Figure 4.3:** (Top) Structures of 3-hydroxypimelyllysine before and after water loss. (Middle) Structures of immonium ions of 3-hydroxypimelyllysine. (Bottom) Product ion mass spectrum of a peptide from SYN_02642 (acetyl-CoA acetyltransferase) having 3-hydroxypimelylation. The spectrum contains peaks for strong water-loss and characterstic ions at m/z 224.1 and 242.1.

Mass spectrum of a peptide with 2,3-didehydropimelylation




Comparing retention times of a peptide bearing different modifications

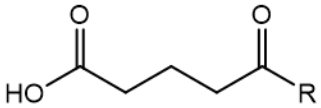
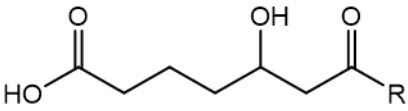
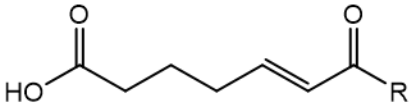
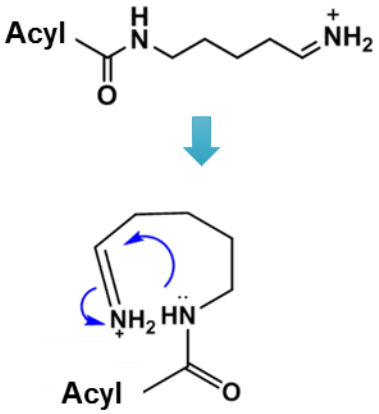
| Peptide | Modification | Retention time |
|---|---|---|
| R=ENLLGKPGDGAK | | 15.0-15.2 min |
|  | Glutarylation (K + 114 Da) | 21.6-21.7 min |
|  | 3-hydroxypimelylation (K + 158 Da) | 21.5-21.6 min |
|  | **2,3-didehydropimelylation (K + 140 Da)** | **22.5-22.6 min** |

**Figure 4.4:** (Top) Product ion mass spectrum of a 2,3-didehydropimelylated peptide originating from SYN_00480, an acyl-CoA dehydrogenase. The spectrum contains acyllysine characteristic immonium ions at m/z 206.1 and 224.1. (Bottom) Comparing elution times in a single chromatographic run of the same peptide bearing different modifications. 2,3-Didehydropimelylated peptide is eluted later than the 3-hydroxypimelylated one, suggesting they are two separate peptides.

Characteristic immonium ion of acyl-lysine

(analog to acetyl-lysine immonium ion m/z 143.118)




Major immonium ion

(analog to major acetyl-lysine immonium ion m/z 126.092)

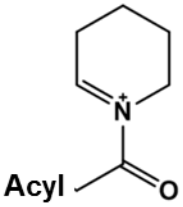



**Figure 4.5:** Structures of characteristic immonium ions for acyllysine. Most of the analogs to the acetyllysine immonium ion m/z 143.118 undergo ammonia loss. Depending on the structure of the acyl group, some can have further water or ammonia losses.

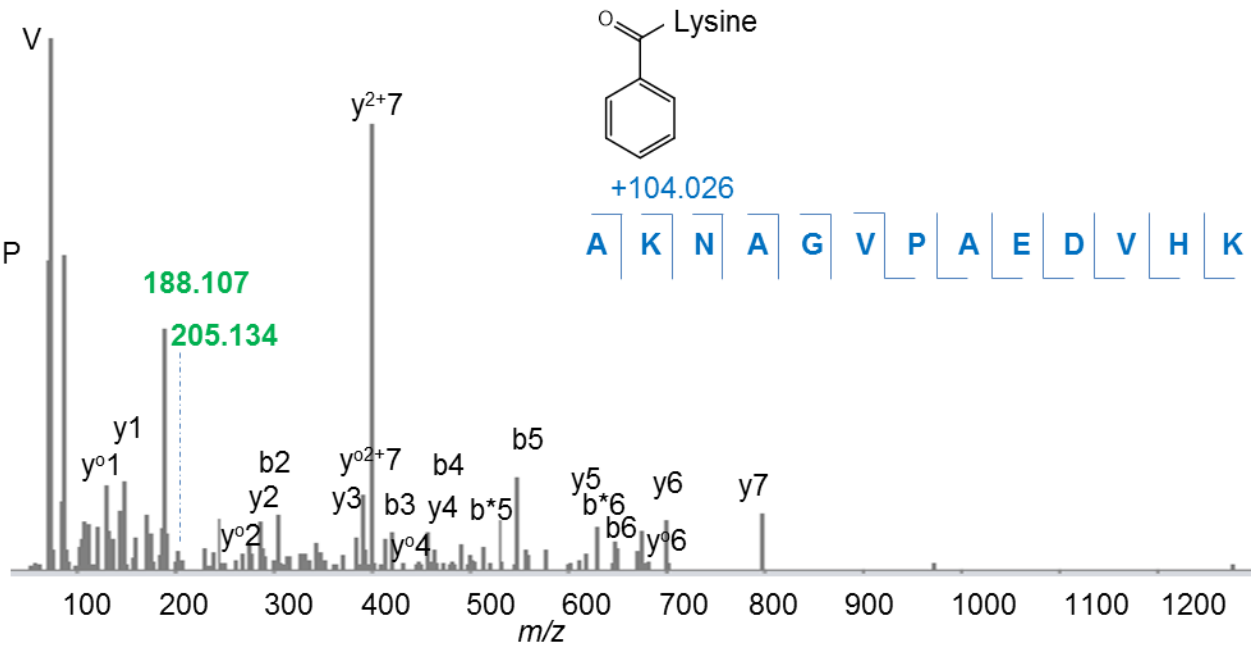



**Figure 4.6:** Product ion mass spectrum of a benzoylated peptide originating from SYN_00481, a glutaconyl-CoA decarboxylase. The spectrum contains acyllysine characteristic ions at m/z 188.107 and 205.134.

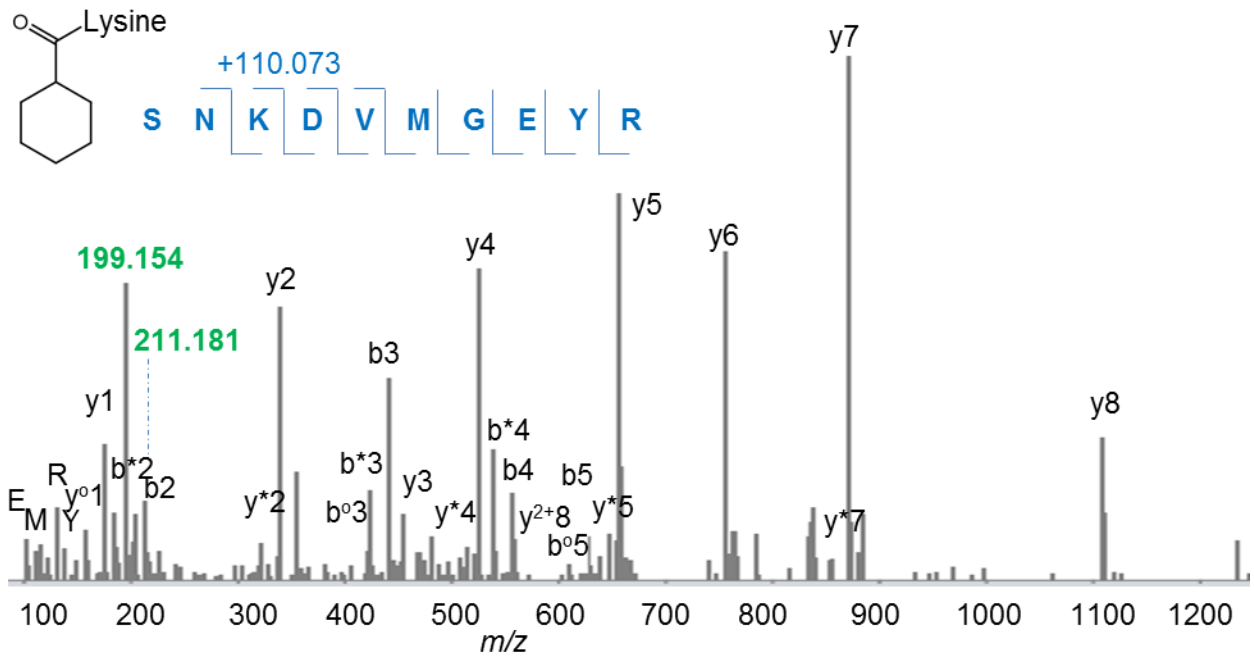



**Figure 4.7:** Product ion mass spectrum of a cyclohexane-1-carboxylated peptide originating from SYN_01655, a 6-hydroxycyclohex-1-ene-1-carboxyl-CoA dehydrogenase. The spectrum contains acyllysine characteristic ions at m/z 199.154 and 211.181.

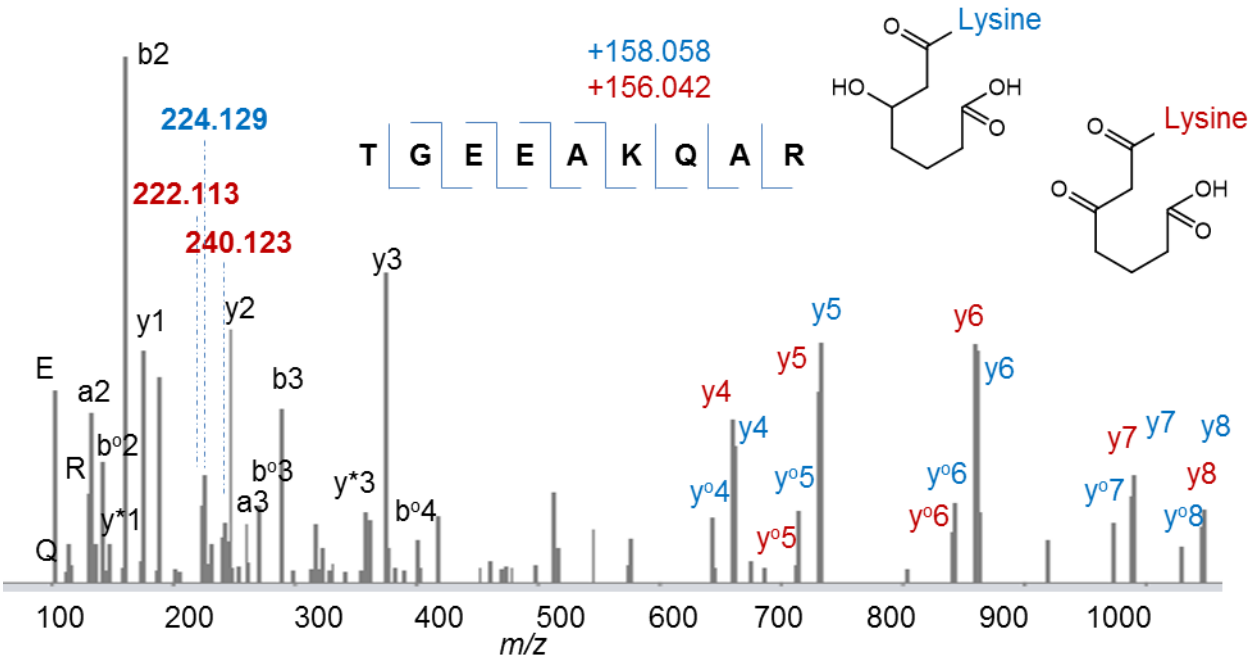



**Figure 4.8:** Product ion mass spectrum of a 3-ketopimelylated peptide (red) originating from SYN_01654, an enoyl-CoA hydratase. The spectrum also shows the fragments for the same peptide with 3-hydroxypimelylation (blue), which was eluted and detected at the same time. The spectrum contains acyllysine characteristic ions at m/z 222.113 and 240.123 for 3-ketopimelyllysine and 224.129 for 3-hdyroxypimelyllysine.

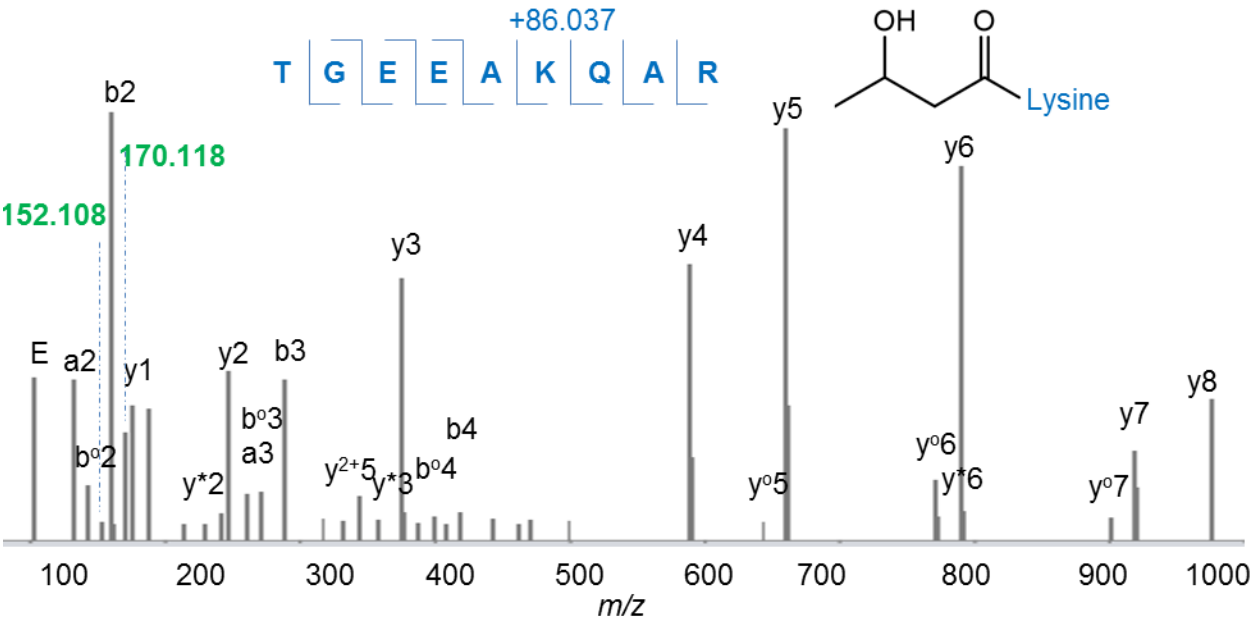



**Figure 4.9:** Product ion mass spectrum of a 3-hydroxybutyrylated peptide originating from SYN_01654, an enoyl-CoA hydratase. The spectrum contains acyllysine characteristic ions at m/z 152.108 and 170.118..

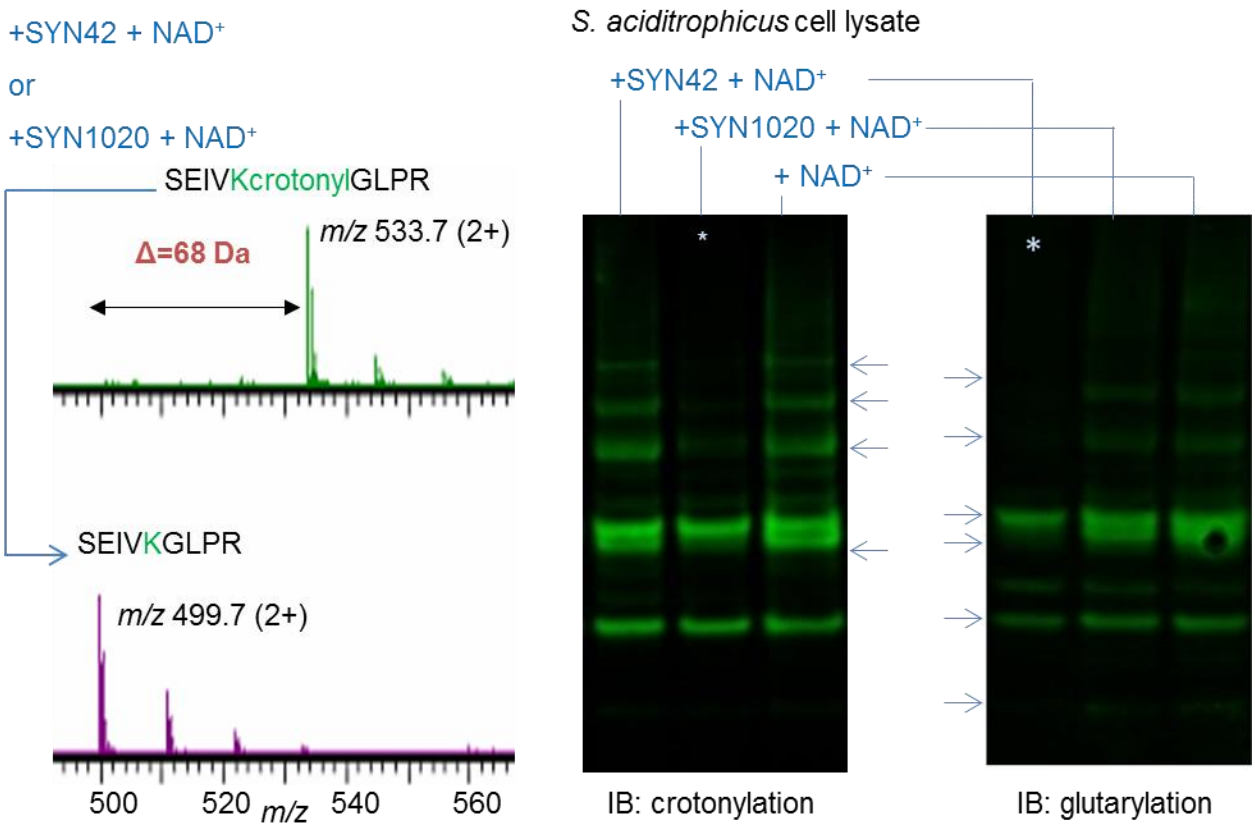



**Figure 4.10:** *S. aciditrophicus* contains two putative sirtuin genes: SYN_42 and SYN_1020. Both were expressed in *E. coli* and their activities were tested against a synthetic peptide substrate bearing crotonylation and against the protein extract of *S. aciditrophicus.* The synthetic peptide incubated with or without SYN_42 or SYN_1020 was analyzed by mass spectrometry (left). The protein extract incubated with or without SYN_42 or SYN_1020 was analyzed by immunoblotting with pan-specific crotonyl- and glutaryl-lysine antibodies (right).

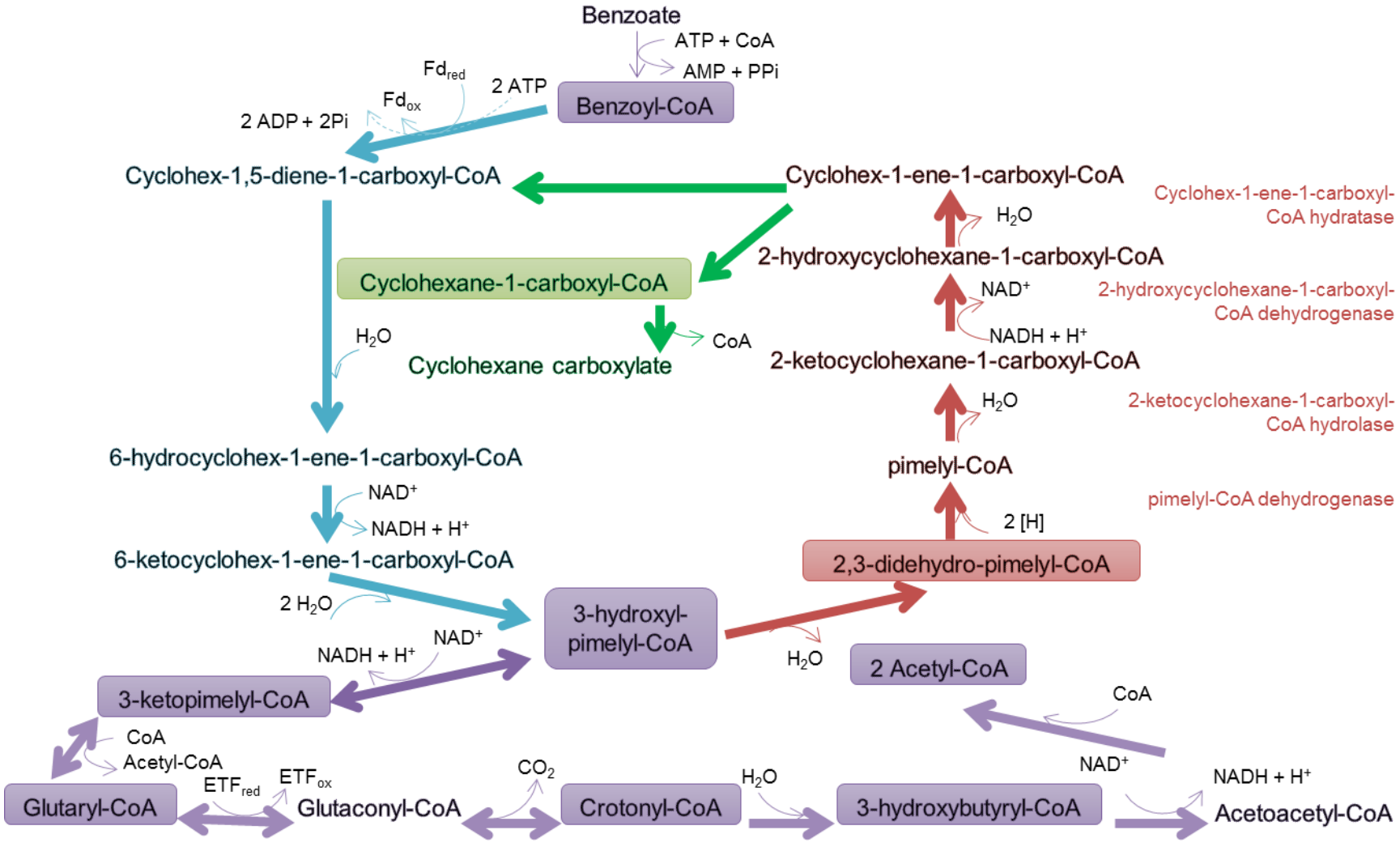



**Figure 4.11:** Suggested enzymatic directions for aromatic degradation in *S. aciditrophicus*. Red colored enzymes were detected by their activities in the cellular extract (Elshahed et al., 2001). The corresponding lysine acylations were detected for the boxed acyl-CoAs.

**Table 4.1**: Comparing elution times of 2,3-didehydropimeloylated and 3-hydroxypimeloylated peptides

| Dataset | Sample name | Protein name | Peptide | Modification | Retention time (minutes) |
|---|---|---|---|---|---|
| 2D-PAGE | Spot G06 | SYN_00481 | SEIVKGLPR | glutarylated lysine | 22.6 & 23.2 |
| | | | | acetylated lysine | 21.5 |
| | | | | crotonylated lysine | 24 |
| | | | | *3-hydropimeloylated lysine* | *22.1* |
| | | | | *2,3-didehydropimeloylated lysine* | *23.2* |
| | Spot H01 | SYN_00480 | ENLLGKPGDGAK | none | 15.1 |
| | | | | glutarylated lysine | 21.7 |
| | | | | *3-hydropimeloylated lysine* | *21.6* |
| | | | | *2,3-didehydropimeloylated lysine* | *22.6* |
| High throughput | benzoate + crotonate pure culture | SYN_00480 | SSPTGEIYLDNVKVPK | acetylated lysine | 62.6 |
| | | | | *3-hydropimeloylated lysine* | *64* |
| | | | | *2,3-didehydropimeloylated lysine* | *65.3* |
| | crotonate pure culture | SYN_02586 | NTPGLSVGKIEHK | none | 13 & 22-25 |
| | | | | acetylated lysine | 38 |
| | | | | butyrylated lysine | 45.4 |
| | | | | *3-hydropimeloylated lysine* | *41.7 & 42.0* |
| | | | | *2,3-didehydropimeloylated lysine* | *42.8 & 43.2* |

**Table 4.2:** Number of acylated proteins detected by high throughput proteomics and immunoprecipitation with either pan-specific crotonyl- or glutaryl-lysine antibodies.

| Lysine modification | # modified proteins from high throughput proteomic data | # modified proteins from enrichment with antibodies against crotonylation and glutarylation |
|---|---|---|
| Acetylation (K+42.011) | 37 | 4 |
| 3hydroxypimeloylation (K+158.058) | 17 | |
| 3-hydroxybutyrylation (K+86.037) | 12 | |
| Glutarylation (K+114.0317) | 8 | 45 |
| Crotonylation (K+68.026) | 7 | 50 |
| Cyclohexane-1-carboxylation (K+110.073) | 6 | |
| 2,3-didehydropimeloylation (K+140.047) | 3 | |
| Butyrylation (K+70.042) | 3 | 5 |
| 3-ketopimeloylation (K+156.042) | 1 | |
| Benzoylation (K+104.026) | 1 | 2 |

**Table 4.3:** Some enzymes are highly acylated with different types of lysine acylations

| Protein | Description | Benzoyl- | Cyclohexane-1-carboxyl- | 2,3-didehydropimeloyl- | 3-hydroxypimeloyl- | 3-ketopimeloyl- | Glutaryl- | 3-hydroxybutyryl- | Crotonyl- | Acetyl- |
|---|---|---|---|---|---|---|---|---|---|---|
| SYN_01654 | 6-ketocyclohex-1-ene-1-carboxyl-CoA hydratase | Y | Y | | Y | Y | Y | Y | Y | Y |
| SYN_02635 | acetyl-CoA synthetase | | Y | Y | Y | | Y | Y | Y | Y |
| SYN_00480 | acyl-CoA dehydrogenase | | Y | Y | Y | | Y | Y | Y | Y |
| SYN_02586 | acyl-CoA dehydrogenase | | Y | Y | Y | | Y | | Y | Y |
| SYN_02896 | benzoate--CoA ligase | | Y | | Y | | Y | Y | Y | Y |
| SYN_01655 | 6-hydroxycyclohex-1-ene-1-carboxyl-CoA dehydrogenase | Y | Y | | | | Y | Y | Y | Y |
| SYN_00481 | glutaconyl-CoA decarboxylasesubunit A | Y | | | | | Y | Y | Y | Y |
| SYN_01681 | acetyl-CoA acetyltransferase | | | | Y | | Y | Y | Y | Y |

**Table 4.4:** Specific enzymatic activities found in *S. aciditrophicus* cell extract (Elshahed et al., 2001)

| Aromatic degradation pathway type | Enzyme | Specific activity (nmol/min/mg of protein) | |
|---|---|---|---|
| | | Benzoate-grown coculture | Crotonate-grown pure culture |
| Shared | Benzoyl-CoA ligase | 141 | 103 |
| ***Rhodopseudomonas* type** | **Cyclohex-1-ene-1-carboxyl-CoA hydratase** | **57** | **2** |
| ***Rhodopseudomonas* type** | **2-hydroxycyclohexane carboxyl-CoA dehydrogenase** | **317** | **41** |
| ***Rhodopseudomonas* type** | **2-ketocyclohexane carboxyl-CoA dehydrogenase** | **391** | **ND** |
| ***Rhodopseudomonas* type** | **Pimelyl-CoA dehydrogenase** | **1,144** | **581** |
| Shared | Glutaryl-CoA dehydrogenase | 630 | 395 |
| Shared | Enoyl-CoA hydratase | 18,780 | 23,880 |
| Shared | L-3-hydroxyacyl-CoA dehydrogenase | 1,244 | 1,488 |
| Shared | 3-ketoacyl-CoA thiolase | 287 | 323 |
| Shared | Phosphotransacetylase | 128 | ND |
| Shared | Acetate kinase | 21 | 10 |

# CHAPTER 5: IMPROVING PEAK MATCHING AND MODIFICATION LOCALIZATION FOR TOP-DOWN PROTEOMICS

## 5.1. Abstract

The top-down proteomics field has been gaining more attention in the past decade for studying protein modifications, protein structures, protein complex compositions, and interactions between subunits or with ligands. Top-down MS analysis of intact proteins has become possible for complex biological samples to study protein isoforms and dynamic modifications. Mass spectrometric methods have seen great improvements but data analysis, the core and basis for all these studies, has lagged behind.

Using mass spectrometric data acquired with a 15 Tesla solariX$^{TM}$ FTMS, currently offering the ultimate in mass resolution and accuracy, we are developing a software package that improves peak matching and detection of protein modifications. Multiple criteria such as types of fragments (b, c, y, z, b+, c·, water-loss, ammonia-loss, and especially internal product ions), mass error distribution, and especially fragmentation “hot spots” and spatially clustered fragments are used to evaluate peak matching results. Applying these criteria has significantly helped reduce the number of matched peaks in scrambled protein sequences or in random proteins similar in length, i.e. reducing false positive matches. Our software also localizes regions containing protein modification(s) and improves data display to the user. Future work will improve scoring methods and allow the software to perform “unsupervised machine learning” to evaluate outputs from this first round of searching and ultimately optimize the final results.

## 5.2. Introduction

Genome sequencing reveals heritable encoded information that is propelling understanding of the living world. Cellular proteomes should, in theory, be contained within the encoded sequences and reflect the abundance of their corresponding transcripts. The complexity of biological systems and proteomes, however, expands beyond that. The level of protein expression depends on not only the level of mRNA transcripts, but also on rates of protein synthesis and degradation. Besides limitations in genome annotation and sequencing/assembly that can yield spurious sequences, proteins exist with alternative splicing isoforms, N- and C-terminal processing, amino acid mutations and dynamic modifications. Bottom-up proteomics, the currently dominant proteomic approach, identifies and characterizes protein sequences and post-translational modifications (PTMs) by proteolytically digesting proteins prior to analysis by mass spectrometry. The subsequent alignments of identified peptides to predicted protein sequences in this approach does not distinguish which peptide fragment arises from which protein isoform. The incomplete recovery of all proteolytic peptide sequences limits the ability to reveal/distinguish intact protein molecules subjected to different biological processing.

Top-down proteomics, the emerging complement to bottom-up proteomics, analyzes protein information without proteolytic cleavage. Thus, it preserves many protein features such as intact mass and sequence, protein isoforms, modification stoichiometry, protein structure, protein complex composition, and interaction among subunits or with ligands (Zhou, Ning, E. Starr, Abu-Farha, & Figeys, 2012). Proof-of-concept for large-scale interrogation of intact proteins in a complex biological sample has been achieved, revealing many protein isoforms and dynamic modifications altered for different cellular stages (Tran et al., 2011). Protein separation methods, mass spectrometric methods for protein fragmentation and complex dissociation in the gas phase, and mass spectrometry resolving power have seen major improvements but data analysis, the core and basis for all these studies, has lagged behind.

No consensus peak matching methods to improve peak assignment and to reduce false positive matches are available. Top-down proteomics so far relies mainly on intact protein accurate mass and matching to few fragments. Top-down MS software packages to match observed peaks to potential protein fragments are mostly based on bottom-up detection algorithms (X. Liu et al., 2012). Most of these software packages require and rely heavily on the detection and the mass of precursor intact protein ions. Many precursor ions, however, can be pre-fragmented in the ionization source or on the way to the mass analyzer, potentially leading to a wrong identification. Current top-down MS algorithms, similar to those employed for bottom-up proteomics, primarily ascribe protein cleavage products to series containing either the amino- or carboxy-terminus (b, c, y, and z series). Thus, only a few product ions can be matched to support an identification; the rest belong to internal fragments and are frequently ignored. Data possessing a low percentage of matched ions is more likely to have been ascribed to an incorrect sequence and no means for determining a false-discovery rate for top-down proteomics has been established so far.

Our initial goals for this point are to increase the number of matched product ions and to reduce the number of false positive matches. For proof-of-concept, we analyzed top-down MS data acquired from highly purified recombinant proteins on a 15 Tesla solariX$^{TM}$ FTMS, currently offering the best mass resolution and accuracy performance commercially available. A list of all ions potentially generated by our instrument for the test samples was calculated: a, b, y, c, z, b+, c·, water-loss, ammonia-loss and internal fragments. The last one is usually ignored or poorly processed by current software because the larger the protein, the more internal fragments it can generate and, thus, the more product ions that could be falsely matched. To overcome this limitation, we applied a stringent set of criteria: (1) very low mass error (according to the mass error distribution of matched peaks, our mass error seems to stay in the ±2 ppm range); (2) fragmentation “hot spots” (according to the matched ion, the fragmentation/cleavage usually

happens after aspartic acid and before proline); and (3) spatially clustered fragments (we argue that if a protein region is susceptible to fragmentation, we should observe more than 2 fragments there). The last criterion does not typically apply to primary ions (a, b, y, c and z series) that arise from a single cleavage. For internal fragments, however, which arise from at least 2 cleavages within the sequence, the criterion is critical to reducing the number of potentially matched internal fragments by half and to distinguish a real hit from the false ones (scrambled sequences of the real protein and for random proteins having the same mass range). Besides matching more product ions and estimating false discovery rate, this approach removes the dependence on (limitation of) the accurate m/z for an intact protein.

Software to identify proteoforms lags behind even more. Purified protein samples, even affinity-tagged, recombinant proteins, are often regarded as homogenous, possessing only the expected sequence and with few specific modifications. In reality, biologically-produced protein samples are much more complicated. Proteoforms could arise from less than 100% modification stoichiometry, combination of many variable modifications, or sequence mutations. Narrowly selecting precursor masses for fragmentation should, in theory, should resolve products from different protein isoforms based on precursor *m/z*. However the oscillating electric fields applied to cylindrical rods used in quadrupole-band filter are limited in their ability to narrowly isolate *a population of ions* having an *m/z* close to the targeted value. Such isolation windows, when aiming to enhance product ion signals, are obviously limited in their ability to eliminate intact proteins with small mass differences (i.e. from small modifications) and thus the product ions arise from non-homogeneous precursor ions.

Our second goal is to identify proteoforms in a mass spectrum. This can be achieved with a better matching and scoring method (above) in combination with generating and localizing regions having amino acid mutations or modifications. Since amino acid mutations are basically changes in predicted amino acid mass, we treat them as modifications.

Current software packages calculate and report protein modifications based on “amino acid residue location”, which works better in bottom-up proteomics because peptides are short and peptide sequence coverage is high. When this approach is applied in top-down proteomics, it gives all possible amino acid residues a chance to have modifications and thus increases the search space and gives a false impression on protein modification sites. A few software packages can become occupied with and over-iterate a few top hits and fail to reveal other modified regions. We instead calculate all possible modifications for “each protein fragment”, perform product ion matching, and then align matched peptides, including modified and non-modified ones, to the protein sequence so that users can visually judge which modification regions are more probable.

Future work will improve the scoring method and will allow for “unsupervised machine learning” on matched peaks to optimize fragmentation type, mass error windows, fragmentation “hot spots”, and spatial cluster sizes to optimize the number of real matches. That will further false discovery rate calculations and identification of protein modifications.

## 5.3. Materials and Methods

**Materials** Glutamate dehydrogenase (GDH) from bovine liver (330 kDa) was purchased from Sigma-Aldrich (St. Louis, MO).

**Sample Preparation** GDH was dissolved in MilliQ water to a concentration of 100 μM and then buffer exchanged three times with 200 mM ammonium acetate solution (300 μL each time) using Amicon centrifugal filters (Millipore Inc., Billerica, MA) with a molecular weight cut-off (MWCO) of 50~100 K. The buffer exchanged protein sample was then diluted with 200 mM ammonium acetate solution to a concentration of 20 μM for native nano-ESI MS analysis.

**FT-ICR MS Analysis** Protein solution was loaded into metal-coated borosilicate capillaries (Au/Pd-coated, 1 μm I.D.; Thermo Fisher Scientific, West Palm Beach, FL) and sprayed at a flow

rate of 10 - 40 nL/min through a nanospray ion source. The experiments were performed using a 15-T Bruker SolariX FT-ICR MS with an infinity cell. The ESI capillary voltage was set to 0.9~1.2 kV. The temperature of the drying gas was 80 ºC and the flow rate was 2.5 L/min. The RF amplitude of the ion-funnels was 300 $V_{pp}$, and the applied voltages were 210 V and 6 V for funnels 1 and 2, respectively. The voltage of skimmer 1 was 30 V and the skimmer 2 voltage was kept at 20 V. The lowest values of RF frequencies were used in all ion-transmission regions: multipole 1 (2 MHz), quadrupole (1.4 MHz), and transfer hexpole (1MHz). Ions were accumulated for 500 ms in the hexapole collision cell before being transmitted to the infinity ICR cell. The time-of-flight of 2.5 ms was used. Vacuum pressures for the different regions were ~2 mbar for the source region, ~$2\times10^{-6}$ mbar for the quadrupole region, and ~$2\times10^{-9}$ mbar for the UHV-chamber pressure. IRMPD was performed with a Synrad 30-W $CO_2$ laser (Mukilteo, WA) that was interfaced to the back of the ICR cell. IRMPD experiments were recorded at different laser energy with the irradiation time kept at 0.5 s and the laser power was varied from 30% to 60% (30 W). One hundred scans were averaged for each spectrum and all spectra were externally calibrated using cesium iodide ions.

**Software package** Our in-house software is written in Java. The product ion peak list was searched against user-defined parameters. Currently the software ranks product ions as #1 for: (1) Main ions (b, c, y, and z series); (2) Clustered ions; (3) Ions created from the “hot spots”; (4) Ions matched uniquely to a product ion and one of its ends is observed in another matched ion. Another part of the software aligns matched peaks having rank #1 onto protein sequences and allows users to easily see the covered regions.

## 5.4. Results and Discussion

### 5.4.1. Software interface and results display

The “Match Fragments” window (**Figure 5.1**) allows users to input parameters such as protein sequence, product ion peak list, MS/MS mass error, types of modifications and the maximum number of modifications, types of ions, and whether to use fragmentation “hot spots” and spatial clustering and parameters for those. The output contains matched ions with different labels such as: types of ions, fragment positions, mass error, spatially cluster, fragmentation sites, and the number of other fragments matched to the same product ions. The search results are organized into Excel tables so that users can easily view, filter or sort to check and accept peak assignment.

The “Display Matched Fragments” window (**Figure 5.2**) allows users to input protein sequence, display region, and the link to “Match Fragments” result. It allows users to easily visualize the protein regions that are susceptible to fragmentation and are modified.

### 5.4.2. Detecting mass error window and fragmentation “hot spots”

The search for the GDH protein example was performed first with a 10 ppm mass error window and the resulting distribution of values by which the masses of matched peaks deviated from calculations was evaluated. The mass error seems to be within +/- 2 ppm (**Figure 5.3**), consistent with the known mass accuracy of the experimental data. The cleavage sites for matched product ion matches were also analyzed, revealing that the preferential cleavage sites lie *C*-terminal to aspartic acid and *N*-terminal to proline (**Figure 5.3**)**.**

Tandem MS analyses of proteins differ from those of peptides because the formers’ extended structure makes dissociation along sequence segments extremely non-uniform. The low percent sequence coverages obtained from intact protein dissociations render bottom-up

algorithms less applicable. Top-down MS data characteristics present a different advantage, however, because certain regions are prone to fragmentation and they should be present as clusters of at least 2-3 cleavage sites. Applying this criterion to internal fragments, (frequently ignored due to matching challenges), we increased the number of matched ions and percent sequence coverage, while reducing false-positive protein identifications.

Matching all calculated ions to product ions gives an average of 3.3 matches per observed ion (528 potential fragments are matched to 159 product ions). After selecting for all matched calculated ions clustered together (cluster size >=3 and an arbitrary cluster radius =9), we disregard 5/6 of the possible matches and are left with 1.1 matches per observed ion (91 potential clustered fragments are matched to 79 product ions) (**Table 5.1**). This helps determine sequence coverage with higher confidence.

The number of internal fragments that match “false” sequences (scrambled sequences of the “real” hit and sequences of proteins within SwissProt that are similar in size) is lower with the clustering criterion (less than 10 clustered ions per false hit as compared to 91 clustered ions for a true hit) (**Table 5.1**). This difference will be advantageous for reducing false-discovery rates and preferable to present calculations that compare the number of terminal ion matches, which have lower hit numbers and smaller differences between correct and incorrect matches. An algorithm relying on clustered internal fragments should increase the success rate for ascribing protein identities to tandem mass spectra along with the number of product ions assigned within those spectra, especially for large proteins (>50 kDa) and can also be applied to complex protein samples.

Using stringent clustering parameters (longer radii, higher product ions targets) is important for differentiating correct from incorrect protein matches, but certain regions within a protein sequence may yield few or no product ions and/or not all of their resultant products may

be detected. Thus, after differentiating “true” protein hits from “false” ones, employing stringent criteria, additional product ions can be ascribed by correlation to the “true” sequence with weaker or no clustering requirements. In principle, the scoring of protein hits can be based entirely on the number of product ions that meet or exceed cluster thresholds (perhaps, in relationship to the precursor mass), or combined with the total number of product ions assigned within mass tolerance.

### 5.4.3. Searching for variable modifications using Big Mascot (Matrix Science, UK)

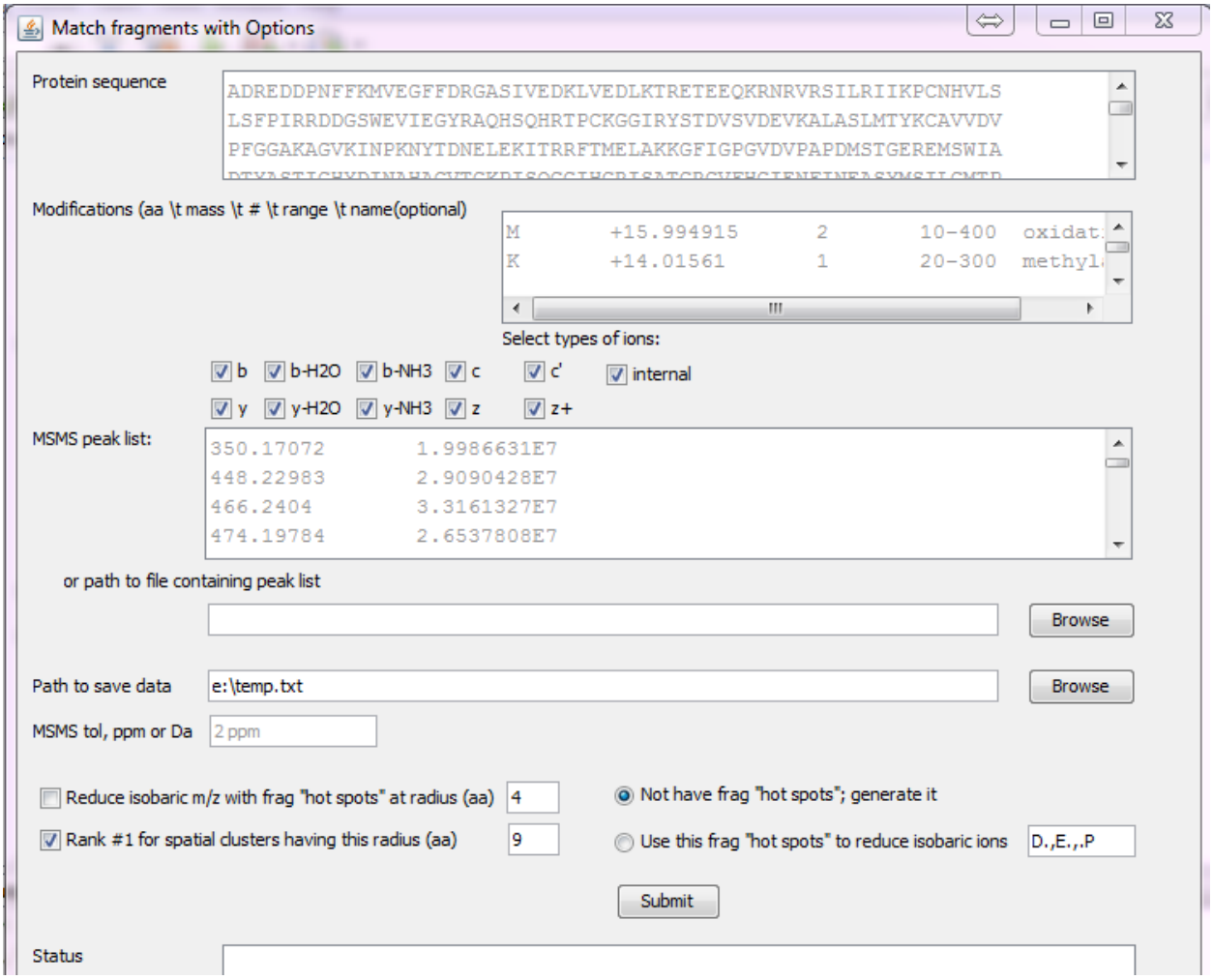
Current software packages to search top-down data for proteins with variable modifications are not satisfying. **Figure 5.4** (Top) displays the search result of the GDH example using the commercial Big Mascot software. First, the software requires the precursor m/z to narrow down the list of potential proteins. We picked an arbitrary m/z for the protein and specified a large MS mass error window (1% or 557 Da). Oxidation (M) and methylation (K) are specified as variable modifications and the software returns a highly methylated N-terminus result with methylation on K33, K41, K53, K90, K105 and K114 and oxidation on M150. Their modification assignments are supported only by 15 internal fragments that do not cluster. The Mascot result display page got stalled and the lengthy display of all possible internal fragments is difficult to interpret easily. Alternative peak assignments to the submitted query are: (i) oxidation on M12 and M111 and methylation on K90 and K105; (ii) oxidation on M233, M238 and M260 and methylation on K400, K420 and K423; (iii) oxidation on M12 and M111 and methylation on K41, K53, K90 and K105; (iv) oxidation on M12, M111 and methylation on K245; (v) oxidation on M12 and M111 and methylation on K90, K105, K114, K126, and K130. Their assigned modification sites are in proximity to each other and the Mascot software gets “stuck” at making “different combinations for a specific set of sites” instead of searching for “other regions containing modifications”.

### 5.4.4. Searching for variable modifications using our software

When preparing the list of potentially modified fragments, our software calculates all possibilities “for each fragments” instead of “for each residues”. For example, a b6 ion with the sequence KKAKDE having at most 1 methylation on K is recorded as a single entry for the b6-K+14-1 site instead of 3 alternative entries (K+14KAKDE, KK+14AKDE and KKAK+14DE) having the same m/z. This eliminates biases toward a specific set of residues and sites outputted from the Big Mascot search and allows users to view how the modified peptides could potentially cover the whole protein sequence (**Figure 5.4** Bottom). When clustering matched modified fragments, the software allows the narrowing down of the modified regions, as shown in **Figure 5.5**.

## 5.5. Conclusions

Utilizing the fragmentation property of intact proteins and frequently ignored internal fragments, we have shown that those features could be useful to improve peak matching and reduce false positive hits for top-down MS. Applying this to protein modification search, we have seen improvements compared to the commercial Big Mascot software. In subsequent steps, we will continue to improve the scoring method and protein modification search. “Unsupervised machine learning” will also be applied to optimize the search results. Our overall goal will be to provide an efficient software platform that will be useful to the top-down proteomics community.




| | A | B | C | D | E | F | G | H | I | J | K |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | start | end | ion type | cal m/z | obs m/z | intensity | delta ppm | match label | rank | sequence | |
| 29 | 489 | 501 | y14 | 1630.859 | 1630.864 | 3.78E+07 | 2.91 | 10 share end: : 3 matches | 1 | K.VFRVYNEAGVTFT.- | |
| 30 | 100 | 114 | itnIC | 1636.862 | 1636.868 | 1.12E+07 | 4.23 | 7 share end: 7 cluster: 4 matches | 1 | V.SVDEVKALASLMTYK.C | |
| 31 | 288 | 301 | itnIC | 1636.869 | 1636.868 | 1.12E+07 | -0.57 | 4 share end: 3 cluster: 4 matches | 1 | D.PKELEDFKLQHGTI.L | |
| 32 | 23 | 175 | itnIC:M_15.99492_2,K_14.01561_1 | 17263.94 | 17263.87 | 3.36E+07 | -3.53 | 5 share end: 3 cluster: 14 matches | 1 | S.IVEDKLVEDLKTRETEEQKRNRVRS | |
| 33 | 99 | 114 | itnIC:K_14.01561_1 | 1749.946 | 1749.953 | 2.97E+07 | 4.50 | 7 share end: 7 cluster: 4 matches | 1 | D.VSVDEVKALASLMTYK.C | |
| 34 | 99 | 114 | itnIN:K_14.01561_1 | 1749.946 | 1749.953 | 2.97E+07 | 4.50 | 5 share start: 4 cluster: 4 matches | 1 | D.VSVDEVKALASLMTYK.C | |
| 35 | 208 | 368 | itnIC:M_15.99492_1,K_14.01561_1 | 17505 | 17504.96 | 1.14E+07 | -1.97 | 9 share end: 9 cluster: 22 matches | 1 | G.IHGRISATGRGVFHGIENFINEASY | |
| 36 | 19 | 175 | itnIC:M_15.99492_2,K_14.01561_1 | 17635.13 | 17635.14 | 4.61E+07 | 0.50 | 5 share end: 3 cluster: 10 matches | 1 | D.RGASIVEDKLVEDLKTRETEEQKR | |
| 37 | 206 | 368 | itnIC:M_15.99492_2,K_14.01561_1 | 17635.04 | 17635.14 | 4.61E+07 | 5.63 | 9 share end: 9 cluster: 10 matches | 1 | Q.GGIHGRISATGRGVFHGIENFINE | |
| 38 | 22 | 36 | itnIN:K_14.01561_1 | 1769.964 | 1769.964 | 4.83E+07 | -0.14 | 3 share start: 3 cluster: 6 matches | 1 | A.SIVEDKLVEDLKTRE.T | |
| 39 | 354 | 368 | itnIC | 1769.962 | 1769.964 | 4.83E+07 | 1.25 | 9 share end: 9 cluster: 6 matches | 1 | T.PEADKIFLERNIMVI.P | |
| 40 | 22 | 37 | itnIN:K_14.01561_1 | 1871.012 | 1871.011 | 8658735 | -0.58 | 3 share start: 3 cluster: 10 matches | 1 | A.SIVEDKLVEDLKTRET.E | |
| 41 | 289 | 304 | itnIC:K_14.01561_1 | 1871.006 | 1871.011 | 8658735 | 2.57 | 5 share end: 4 cluster: 10 matches | 1 | P.KELEDFKLQHGTILGF.P | |
| 42 | 353 | 368 | itnIC | 1871.01 | 1871.011 | 8658735 | 0.74 | 9 share end: 9 cluster: 10 matches | 1 | T.TPEADKIFLERNIMVI.P | |
| 43 | 285 | 301 | itnIC:K_14.01561_1 | 1936.018 | 1936.034 | 2.13E+07 | 8.62 | 4 share end: 3 cluster: 4 matches | 1 | D.GIDPKELEDFKLQHGTI.L | |
| 44 | 288 | 304 | itnIC | 1954.043 | 1954.045 | 7.82E+07 | 0.60 | 5 share end: 4 cluster: 4 matches | 1 | D.PKELEDFKLQHGTILGF.P | |
| 45 | 139 | 154 | itnIN | 1961.064 | 1961.063 | 6919554 | -0.39 | 6 share start: 5 cluster: 10 matches | 1 | D.NELEKITRRFTMELAK.K | |
| 46 | 357 | 372 | itnIC | 1961.057 | 1961.063 | 6919554 | 3.29 | 3 share end: 3 cluster: 10 matches | 1 | A.DKIFLERNIMVIPDLY.L | |
| 47 | 359 | 374 | itnIC:M_15.99492_1 | 1961.057 | 1961.063 | 6919554 | 3.29 | 3 share end: 3 cluster: 10 matches | 1 | K.IFLERNIMVIPDLYLN.A | |

**Figure 5.1**: View of the software interface (top) and result display (bottom) for “Match Fragments”. For display purposes, data were searched with less stringent criteria.

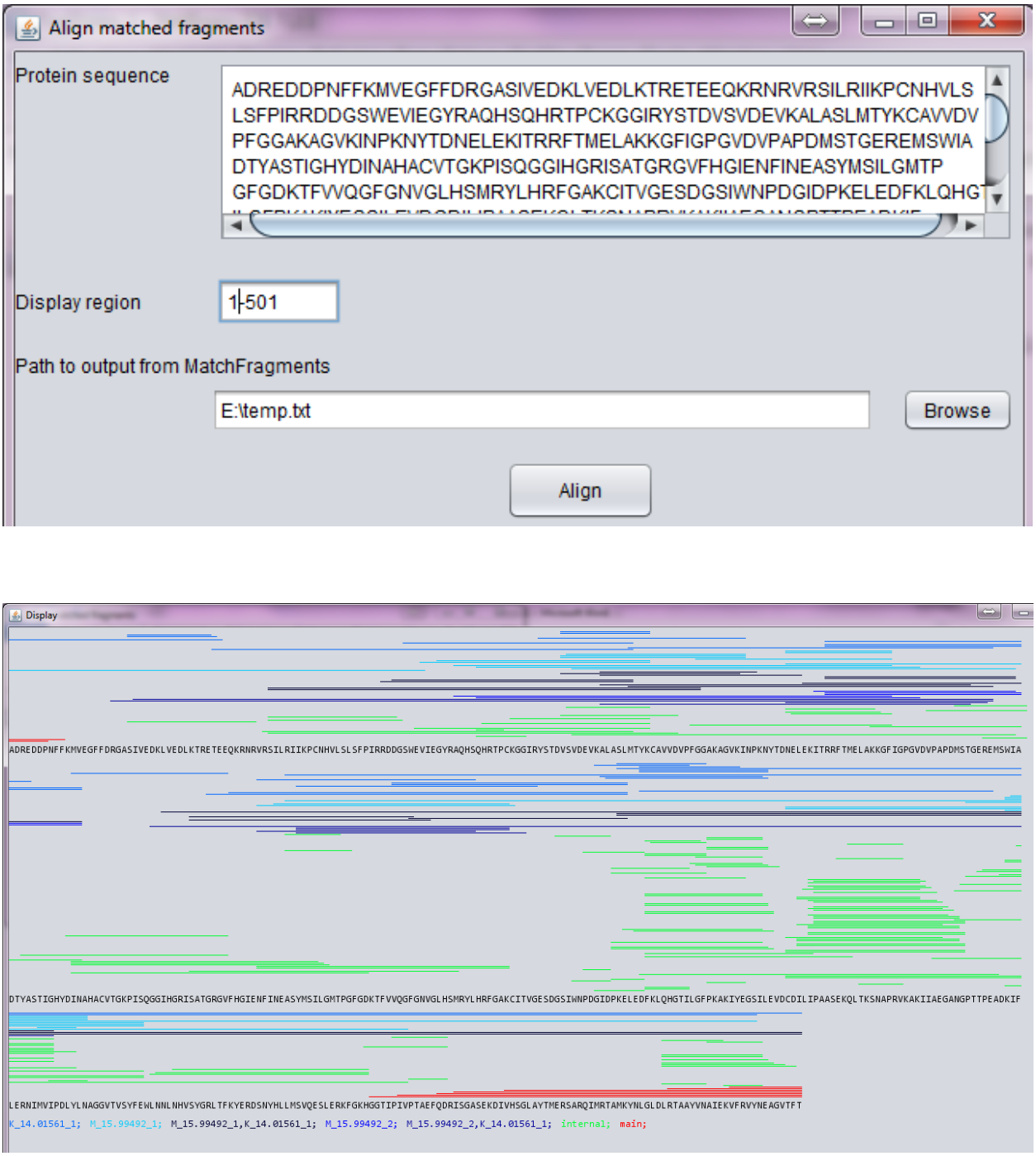
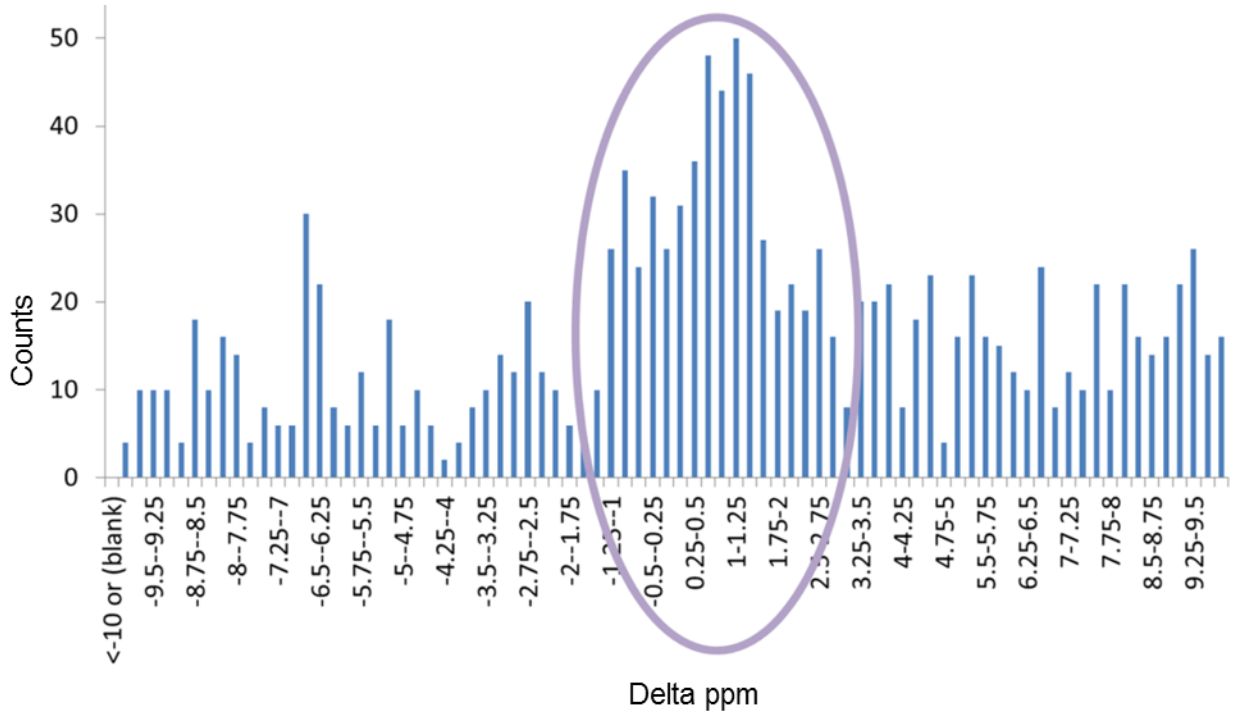



**Figure 5.2:** Software interface (top) and result display (bottom) for “Align Matched Fragments”. For display purposes, data were searched with less stringent criteria.



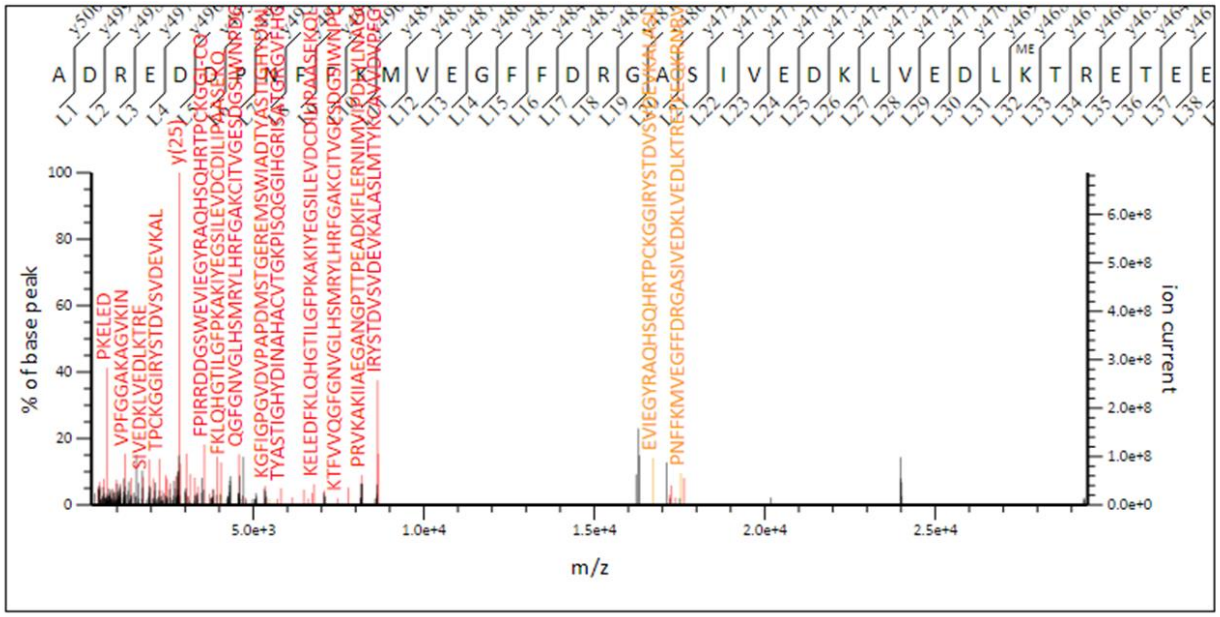

Occurrence of fragmentation on C-terminal side of X

| amino acid X. | # fragments having X. | # X in protein sequence | # X. fragments per X |
|---|---|---|---|
| D. | 182 | 29 | 6 |
| I. | 129 | 38 | 3 |
| P. | 61 | 21 | 3 |
| E. | 88 | 35 | 3 |
| L. | 76 | 31 | 2 |
| K. | 68 | 32 | 2 |
| T. | 49 | 29 | 2 |
| A. | 59 | 35 | 2 |
| C. | 10 | 6 | 2 |
| V. | 56 | 34 | 2 |
| N. | 36 | 22 | 2 |
| F. | 37 | 23 | 2 |
| R. | 43 | 30 | 1 |
| G. | 65 | 46 | 1 |
| Q. | 14 | 10 | 1 |
| M. | 18 | 13 | 1 |
| Y. | 22 | 18 | 1 |
| S. | 28 | 31 | 1 |
| H. | 11 | 14 | 1 |
| W. | 2 | 4 | 1 |

Occurrence of fragmentation on N-terminal side of X

| amino acid .X | # fragments having .X | # X in protein sequence | # .X fragments per X |
|---|---|---|---|
| .P | 144 | 21 | 7 |
| .I | 117 | 38 | 3 |
| .D | 86 | 29 | 3 |
| .G | 103 | 46 | 2 |
| .L | 66 | 31 | 2 |
| .W | 8 | 4 | 2 |
| .V | 64 | 34 | 2 |
| .F | 43 | 23 | 2 |
| .Q | 17 | 10 | 2 |
| .S | 52 | 31 | 2 |
| .C | 10 | 6 | 2 |
| .E | 58 | 35 | 2 |
| .K | 52 | 32 | 2 |
| .A | 56 | 35 | 2 |
| .H | 22 | 14 | 2 |
| .M | 20 | 13 | 2 |
| .R | 44 | 30 | 1 |
| .T | 40 | 29 | 1 |
| .N | 30 | 22 | 1 |
| .Y | 14 | 18 | 1 |

**Figure 5.3:** MS/MS mass error distribution (Top) and fragmentation “hot spots” analysis (Bottom) for GDH.




```
Monoisotopic mass of neutral peptide Mr(calc): 55791.3068
Variable modifications:
K33     : Methyl (K)
K41     : Methyl (K)
K53     : Methyl (K)
K90     : Methyl (K)
K105    : Methyl (K)
K114    : Methyl (K)
M150    : Oxidation (M), with neutral losses 63.9983(shown in table), 0.0000
```

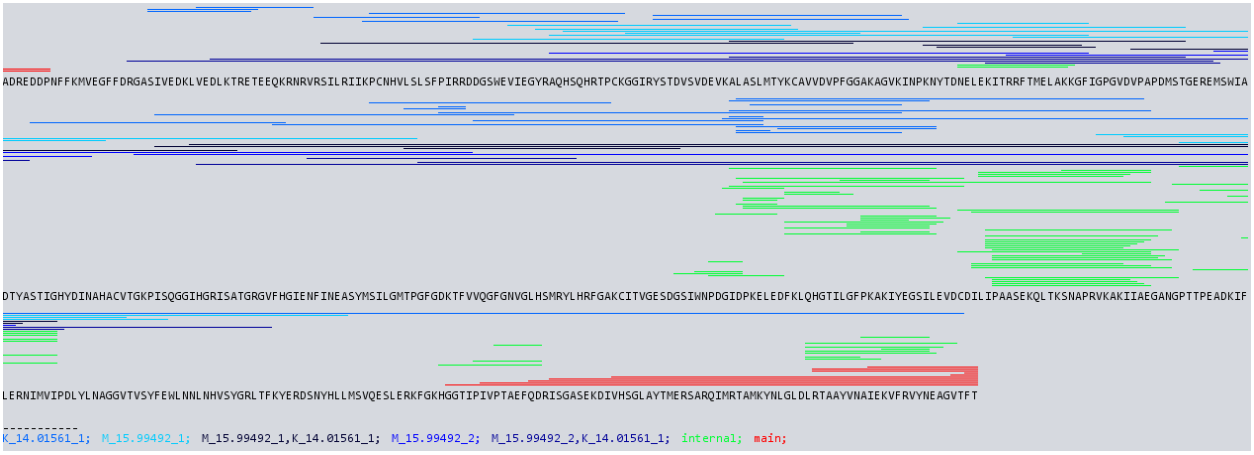



**Figure 5.4:** (Top) Display the search results using Big Mascot for the GDH example with an arbitrary precursor MH+ = 55727, MS mass error = 1%, MSMS mass error = 2ppm, oxidation (M) and methylation (K). (Bottom) Search results using our software with MSMS mass error = 2ppm, at most 2 oxidations (M), 1 methylation (K). Spatial clustering has not been applied to fragments containing modifications.

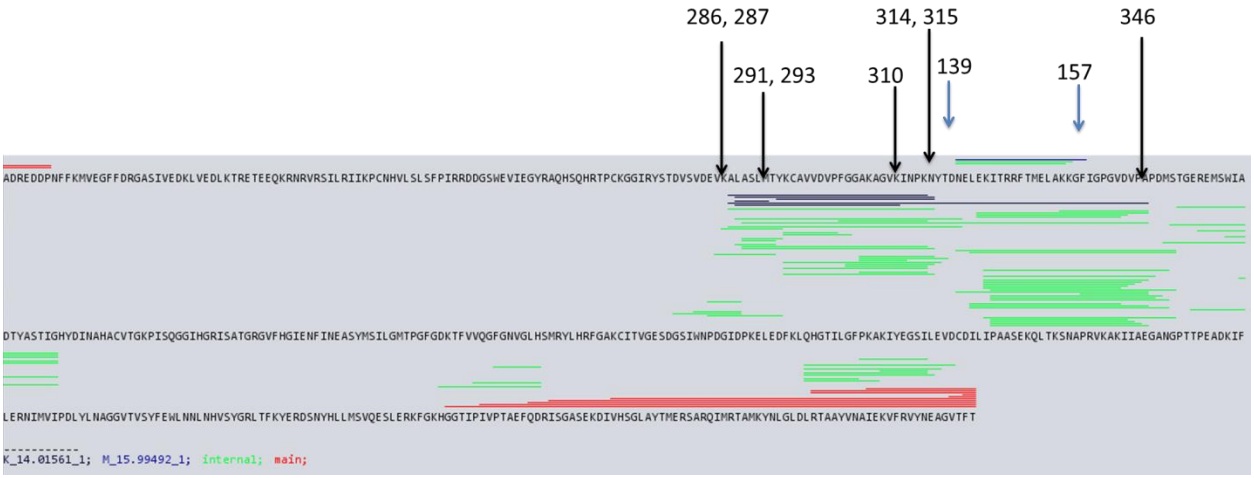



**Figure 5.5:** Re-display Figure 5.4 (Bottom) with clustered modified fragments. The numbers are positions of detected peptides.

**Table 5.1:** Comparing matched peak number in "real" versus "false" protein hits for the GDH example

| | # matched cal ion | # obs ion matched | % obs ion matched (out of 280) | ave matched cal ion / obs ion | # matched terminal ions | # clustered ions | # obs ion matched | ave clustered ions / obs ion |
|---|---|---|---|---|---|---|---|---|
| "true" hit | 528 | 159 | 57 | 3.3 | 12 | 91 | 79 | 1.2 |
| scrambled "true" | 215 | 75 | 27 | 2.9 | 1 | 3 | 3 | 1.0 |
| scrambled "true" | 206 | 66 | 24 | 3.1 | 0 | 0 | 0 | |
| ~ 500 aa protein | 142 | 50 | 18 | 2.8 | 0 | 0 | 0 | |
| ~ 500 aa protein | 241 | 59 | 21 | 4.1 | 1 | 8 | 5 | 1.6 |
| ~ 500 aa protein | 308 | 86 | 31 | 3.6 | 0 | 3 | 3 | 1.0 |
| ~ 500 aa protein | 119 | 45 | 16 | 2.6 | 5 | 0 | 0 | |

# REFERENCES


AbouElfetouh, A., Kuhn, M. L., Hu, L. I., Scholle, M. D., Sorensen, D. J., Sahu, A. K., … Wolfe, A. J. (2014). The E. coli sirtuin CobB shows no preference for enzymatic and nonenzymatic lysine acetylation substrate sites. *MicrobiologyOpen*, 1–18. http://doi.org/10.1002/mbo3.223

Amos, D. A., & McInerney, M. J. (1991). Composition of Poly-Beta-Hydroxyalkanoate from Syntrophomonas- Wolfei Grown on Unsaturated Fatty-Acid Substrates. *Archives of Microbiology*, *155*, 103–106.

Baeza, J., Smallegan, M. J., & Denu, J. M. (2015). Site-specific reactivity of non-enzymatic lysine acetylation. *ACS Chemical Biology, 10*(1), 122-128. http://doi.org/10.1021/cb500848p

Bao, X., Wang, Y., Li, X., Li, X.-M., Liu, Z., Yang, T., … Li, X. D. (2014). Identification of "erasers" for lysine crotonylated histone marks using a chemical proteomics approach. *eLife*, *3*, 1–18. http://doi.org/10.7554/eLife.02999

Baranov, P. V, Gurvich, O. L., Fayet, O., Prère, M. F., Miller, W. A., Gesteland, R. F., … Giddings, M. C. (2001). RECODE: a database of frameshifting, bypassing and codon redefinition utilized for gene expression. *Nucleic Acids Research*, *29*(1), 264–267. http://doi.org/10.1093/nar/29.1.264

Beaty, P. S., & McInerney, M. J. (1987). Growth of Syntrophomonas wolfei in pure culture on crotonate. *Archives of Microbiology, 147,* 389-393. http://doi.org/10.1007/BF00406138

Carmona, M., Zamarro, M. T., Blázquez, B., Durante-Rodríguez, G., Juárez, J. F., Valderrama, J. A., … Díaz, E. (2009). Anaerobic catabolism of aromatic compounds: a genetic and genomic view. *Microbiology and Molecular Biology Reviews : MMBR*, *73*(1), 71–133. http://doi.org/10.1128/MMBR.00021-08

Chen, Y., Sprung, R., Tang, Y., Ball, H., Sangras, B., Kim, S. C., … Zhao, Y. (2007). Lysine propionylation and butyrylation are novel post-translational modifications in histones. *Molecular & Cellular Proteomics : MCP*, *6*(5), 812–819.

Cheng, Z., Tang, Y., Chen, Y., Kim, S., Liu, H., Li, S. S. C., … Zhao, Y. (2009). Molecular characterization of propionyllysines in non-histone proteins. *Molecular & Cellular Proteomics : MCP*, *8*(1), 45–52.

Choudhary, C., Kumar, C., Gnad, F., Nielsen, M. L., Rehman, M., Walther, T. C., … Mann, M. (2009). Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science (New York, N.Y.)*, *325*(5942), 834–840.

Colak, G., Xie, Z., Zhu, A. Y., Dai, L., Lu, Z., Zhang, Y., … Tan, M. (2013). Identification of lysine succinylation substrates and the succinylation regulatory enzyme CobB in Escherichia coli. *Molecular & Cellular Proteomics : MCP*, *12*, 3509–20.

Dai, L., Peng, C., Montellier, E., Lu, Z., Chen, Y., Ishii, H., … Zhao, Y. (2014). Lysine 2-hydroxyisobutyrylation is a widely distributed active histone mark. *Nature Chemical Biology*, *10*(5), 365–370.

Desuccinylation, S. L., Park, J., Chen, Y., Tishkoff, D. X., Peng, C., Tan, M., … Zhao, Y. (2013). SIRT5-Mediated Lysine Desuccinylation Impacts Diverse Metabolic Pathways. *Molecular Cell*, *50*(6), 919–930. http://doi.org/10.1016/j.molcel.2013.06.001

Du, J., Zhou, Y., Su, X., Yu, J. J., Khan, S., Jiang, H., … Lin, H. (2011). Sirt5 Is a NAD-Dependent Protein Lysine Demalonylase and Desuccinylase. *Science, 334,* 806-809. http://doi.org/10.1126/science.1207861

Elshahed, M. S., Bhupathiraju, V. K., Wofford, N. Q., Nanny, M. A., & 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. *Applied and Environmental Microbiology*, *67*, 1728–1738. http://doi.org/10.1128/AEM.67.4.1728-1738.2001

Elshahed, M. S., & McInerney, M. J. (2001). Benzoate Fermentation by the Anaerobic Bacterium Syntrophus aciditrophicus in the Absence of Hydrogen-Using Microorganisms. *Applied and Environmental Microbiology*, *67*, 5520–5525. http://doi.org/10.1128/AEM.67.12.5520-5525.2001

Erde, J., Loo, R. R. O., & Loo, J. A. (2014). Enhanced FASP (eFASP) to increase proteome coverage and sample recovery for quantitative proteomic experiments. *Journal of Proteome Research*, *13*, 1885–95. http://doi.org/10.1021/pr4010019

Friedmann, D. R., & Marmorstein, R. (2013). Structure and mechanism of non-histone protein acetyltransferase enzymes. *FEBS Journal, 280*, 5570-5581. http://doi.org/10.1111/febs.12373

Fuchs, G. (2008). Anaerobic metabolism of aromatic compounds. In *Annals of the New York Academy of Sciences* (Vol. 1125, pp. 82–99). http://doi.org/10.1196/annals.1419.010

Fuchs, G., Boll, M., & Heider, J. (2011). Microbial degradation of aromatic compounds — from one strategy to four. *Nature Reviews Microbiology, 9,* 803-816.

Genschel, U. (2004). Coenzyme A biosynthesis: Reconstruction of the pathway in archaea and an evolutionary scenario based on comparative genomics. *Molecular Biology and Evolution*, *21*, 1242–1251. http://doi.org/10.1093/molbev/msh119

Gesteland, R. F., & Atkins, J. F. (1996). Recoding: dynamic reprogramming of translation. *Annual Review of Biochemistry*, *65*, 741–768. http://doi.org/10.1146/annurev.biochem.65.1.741

Gibson, G. E., Xu, H., Chen, H.-L., Chen, W., Denton, T., & Zhang, S. (2015). Alpha-ketoglutarate dehydrogenase complex-dependent succinylation of proteins in neurons and neuronal cell lines. *Journal of Neurochemistry, 134*(1), 86-96. http://doi.org/10.1111/jnc.13096

Gibson, J., & S Harwood, C. (2002). Metabolic diversity in aromatic compound utilization by anaerobic microbes. *Annual Review of Microbiology*, *56*, 345–369.

Glozak, M. A., Sengupta, N., Zhang, X., & Seto, E. (2005). Acetylation and deacetylation of non-histone proteins. *Gene, 363*(1-2), 15-23.

Gregoretti, I. V., Lee, Y. M., & Goodson, H. V. (2004). Molecular evolution of the histone deacetylase family: Functional implications of phylogenetic analysis. *Journal of Molecular Biology*, *338*, 17–31. http://doi.org/10.1016/j.jmb.2004.02.006

Okanishi, H., Kim, K., Masui, R., & Kuramitsu, S. (2014). Lysine propionylation is a prevalent post-translational modification in *Thermus thermophilus*. *Molecular & Cell Proteomics, 13*, 2382-2398.

Hirschey, M. D., Shimazu, T., Goetzman, E., Jing, E., Schwer, B., Lombard, D. B., … Verdin, E. (2010). SIRT3 regulates mitochondrial fatty-acid oxidation by reversible enzyme deacetylation. *Nature*, *464*(7285), 121–125. http://doi.org/10.1038/nature08778

Hirschey, M., & Zhao, Y. (2015). Metabolic regulation by lysine malonylation, succinylation and glutarylation. *ASBMB*, *5*, 1–22.

Holmes, D. E., Risso, C., Smith, J. A., & Lovley, D. R. (2012). Genome-scale analysis of anaerobic benzoate and phenol metabolism in the hyperthermophilic archaeon *Ferroglobus placidus*. *The ISME Journal, 6,* 146-157. http://doi.org/10.1038/ismej.2011.88

Jackson, B. E., & McInerney, M. J. (2002). Anaerobic microbial metabolism can proceed close to thermodynamic limits. *Nature*, *415*, 454–456. http://doi.org/10.1038/415454a

Jeffers, V., & Sullivan, W. J. (2012). Lysine acetylation is widespread on proteins of diverse function and localization in the protozoan parasite Toxoplasma gondii. *Eukaryotic Cell*, *11*(6), 735–742. http://doi.org/10.1128/EC.00088-12

Kim, J. Y., Kim, K. W., Kwon, H. J., Lee, D. W., & Yoo, J. S. (2002). Probing lysine acetylation with a modification-specific marker ion using high-performance liquid chromatography/electrospray-mass spectrometry with collision-induced dissociation. *Analytical Chemistry*, *74*, 5443–5449. http://doi.org/10.1021/ac0256080

Kim, M., Le, H., McInerney, M. J., & Buckel, W. (2013). Identification and characterization of re-citrate synthase in syntrophus aciditrophicus. *Journal of Bacteriology*, *195*, 1689–1696. http://doi.org/10.1128/JB.02185-12

Kim, S. C., Sprung, R., Chen, Y., Xu, Y., Ball, H., Pei, J., … Zhao, Y. (2006). Substrate and Functional Diversity of Lysine Acetylation Revealed by a Proteomics Survey. *Molecular Cell*, *23*(4), 607–618.

Kirschke, S., Bousquet, P., Ciais, P., Saunois, M., Canadell, J. G., Dlugokencky, E. J., … Zeng, G. (2013). Three decades of global methane sources and sinks. *Nature Geoscience*, *6*(10), 813–823. http://doi.org/10.1038/ngeo1955

Kosono, S., Tamura, M., Suzuki, S., Kawamura, Y., Yoshida, A., Nishiyama, M., & Yoshida, M. (2015). Changes in the Acetylome and Succinylome of Bacillus subtilis in Response to Carbon Source. *Plos One*, *10*, e0131169. http://doi.org/10.1371/journal.pone.0131169

Kryukov, G. V, & Gladyshev, V. N. (2004). The prokaryotic selenoproteome. *EMBO Reports*, *5*(5), 538–543. http://doi.org/10.1038/sj.embor.7400126

Kung, J. W., Meier, A.-K., Mergelsberg, M., & Boll, M. (2014). Enzymes involved in a novel anaerobic cyclohexane carboxylic acid degradation pathway. *Journal of Bacteriology, 196*(20), 3367-3374. http://doi.org/10.1128/JB.02071-14

Kung, J. W., Seifert, J., von Bergen, M., & Bolla, M. (2013). Cyclohexanecarboxyl-coenzyme a (CoA) and cyclohex-1-ene-1-carboxyl-CoA dehydrogenases, two enzymes involved in the fermentation of benzoate and crotonate in syntrophus aciditrophicus. *Journal of Bacteriology*, *195*, 3193–3200. http://doi.org/10.1128/JB.00322-13

Kursula, P., Ojala, J., Lambeir, A., & Wierenga, R. K. (2002). The Catalytic Cycle of Biosynthetic Thiolase : A Conformational Journey of an Acetyl Group through Four Binding Modes and Two Oxyanion Holes, *Biochemistry, 41*(52), 15543–15556.

Liu, B., Lin, Y., Darwanto, A., Song, X., Xu, G., & Zhang, K. (2009). Identification and characterization of propionylation at histone H3 lysine 23 in mammalian cells. *The Journal of Biological Chemistry*, *284*(47), 32288–32295.

Liu, X., Sirotkin, Y., Shen, Y., Anderson, G., Tsai, Y. S., Ting, Y. S., … Pevzner, P. A. (2012). Protein identification using top-down spectra. *Molecular & Cellular Proteomics*, *11*, M111.008524. http://doi.org/10.1074/mcp.M111.008524

Lobanov, A. V, Turanov, A. A., Hatfield, D. L., & Gladyshev, V. N. (2010). Dual functions of codons in the genetic code. *Critical Reviews in Biochemistry and Molecular Biology*, *45*, 257–265. http://doi.org/10.3109/10409231003786094

Löffler, C., Kuntze, K., Vazquez, J. R., Rugor, A., Kung, J. W., Böttcher, A., & Boll, M. (2011). Occurrence, genes and expression of the W/Se-containing class II benzoyl-coenzyme A reductases in anaerobic bacteria. *Environmental Microbiology*, *13*, 696–709. http://doi.org/10.1111/j.1462-2920.2010.02374.x

Long, M. (2000). A new function evolved from gene fusion. *Genome Research*, *10*(11), 1655–1657. http://doi.org/10.1101/gr.165700

Lundby, A., Lage, K., Weinert, B. T., Bekker-Jensen, D. B., Secher, A., Skovgaard, T., … Olsen, J. V. (2012). Proteomic Analysis of Lysine Acetylation Sites in Rat Tissues Reveals Organ Specificity and Subcellular Patterns. *Cell Reports*, *2*(2), 419–431.

McCarthy, A. D., & Hardie, D. G. (1984). Fatty-acid synthase - an example of protein evolution by gene fusion. *Trends BiochemSci*, *9*, 60–63.

McInerney, M. J., Rohlin, L., Mouttaki, H., Kim, U., Krupp, R. S., Rios-Hernandez, L., … Gunsalus, R. P. (2007). The genome of Syntrophus acidotrophicus: Life at the thermodynamic limit of microbial growth. *Proceedings of the National Academy of Science USA*, *104*, 7600–7605.

McInerney, M. J., Sieber, J. R., & Gunsalus, R. P. (2009). Syntrophy in anaerobic global carbon cycles. *Current Opinion in Biotechnology, 20,* 623-632. http://doi.org/10.1016/j.copbio.2009.10.001

Mcinerney, M. J., & Wofford, N. Q. (1992). Enzymes involved in crotonate metabolism in *Syntrophomnas wolfei*. *Archives of Microbiology, 158*, 344–349.

Metz, B., Kersten, G. F. a, Hoogerhout, P., Brugghe, H. F., Timmermans, H. a M., Jong, A. De, … Jiskoot, W. (2004). Identification of Formaldehyde-induced Modifications in Proteins, *Journal of Biological Chemistry, 279*(8), 6235–6243. http://doi.org/10.1074/jbc.M310752200

Miao, J., Lawrence, M., Jeffers, V., Zhao, F., Parker, D., Ge, Y., … Cui, L. (2013). Extensive lysine acetylation occurs in evolutionarily conserved metabolic pathways and parasite-specific functions during Plasmodium falciparum intraerythrocytic development. *Molecular Microbiology*, *89*, 660–675. http://doi.org/10.1111/mmi.12303

Montgomery, D. C., Sorum, A. W., & Meier, J. L. (2015). Defining the Orphan Functions of Lysine Acetyltransferases. *ACS Chemical Biology*, *10*, 85–94. http://doi.org/10.1021/cb500853p

Mouttaki, H., Nanny, M. a, & McInerney, M. J. (2007). Cyclohexane carboxylate and benzoate formation from crotonate in Syntrophus aciditrophicus. *Applied and Environmental Microbiology*, *73*(3), 930–8. http://doi.org/10.1128/AEM.02227-06

Na, S., Bandeira, N., & Paek, E. (2012). Fast Multi-blind Modification Search through Tandem Mass Spectrometry. *Molecular & Cellular Proteomics*, *11*(4), M111.010199. http://doi.org/10.1074/mcp.M111.010199

Namy, O., Rousset, J. P., Napthine, S., & Brierley, I. (2004). Reprogrammed Genetic Decoding in Cellular Gene Expression. *Molecular Cell*, *13*(2), 157–168. http://doi.org/10.1016/S1097-2765(04)00031-0

Okanishi, H., Kim, K., Masui, R., & Kuramitsu, S. (2013). Acetylome with structural mapping reveals the significance of lysine acetylation in Thermus thermophilus. *Journal of Proteome Research*, *12*, 3952–3968. http://doi.org/10.1021/pr400245k

Papanicolaou, K. N., O'Rourke, B., & Foster, D. B. (2014). Metabolism leaves its mark on the powerhouse: recent progress in post-translational modifications of lysine in mitochondria. *Frontiers in Physiology*, *5*(September), 301. http://doi.org/10.3389/fphys.2014.00301

Peng, C., Lu, Z., Xie, Z., Cheng, Z., Chen, Y., Tan, M., … Zhao, Y. (2011). The First Identification of Lysine Malonylation Substrates and Its Regulatory Enzyme. *Molecular & Cellular Proteomics, 10*(12), M111.012658. http://doi.org/10.1074/mcp.M111.012658

Peters, F., Shinoda, Y., McInerney, M. J., & 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. *Journal of Bacteriology*, *189*(3), 1055–1060. http://doi.org/10.1128/JB.01467-06

Pougovkina, O., Te Brinke, H., Ofman, R., Van Cruchten, A. G., Kulik, W., Wanders, R. J. A., … De Boer, V. C. J. (2014). Mitochondrial protein acetylation is driven by acetyl-CoA from fatty acid oxidation. *Human Molecular Genetics*, *23*, 3513–3522. http://doi.org/10.1093/hmg/ddu059

Pougovkina, O., Te Brinke, H., Wanders, R. J. a, Houten, S. M., & de Boer, V. C. J. (2014). Aberrant protein acylation is a common observation in inborn errors of acyl-CoA metabolism. *Journal of Inherited Metabolic Disease*, *37*(5), 709–14. http://doi.org/10.1007/s10545-014-9684-9

Ramponi, G., Manao, G., & Camici, G. (1975). Nonenzymatic acetylation of histones with acetyl phosphate and acetyl adenylate. *Biochemistry*, *14*, 2681–2685.

Rappsilber, J., Mann, M., & Ishihama, Y. (2007). Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nature Protocols*, *2*, 1896–1906. http://doi.org/10.1038/nprot.2007.261

Sabari, B. R., Tang, Z., Huang, H., Yong-Gonzalez, V., Molina, H., Kong, H. E., … Allis, C. D. (2015). Intracellular Crotonyl-CoA Stimulates Transcription through p300-Catalyzed Histone Crotonylation. *Molecular Cell*, *58*(2), 203–215. http://doi.org/10.1016/j.molcel.2015.02.029

Schink, B. (1997). Energetics of syntrophic cooperation in methanogenic degradation. *Microbiology and Molecular Biology Reviews : MMBR*, *61*, 262–280.

Schink, B., & Friedrich, M. (1994). Energetics of syntrophic fatty acid oxidation. *FEMS Microbiology Reviews*, *15*, 85–94. http://doi.org/10.1016/0168-6445(94)90105-8

Schmid, G., René, S. B., & Boll, M. (2015). Enzymes of the benzoyl-coenzyme A degradation pathway in the hyperthermophilic archaeon *Ferroglobus placidus*. *Environmental Microbiology*. http://doi.org/10.1111/1462-2920.12785

Shi, L., & Tu, B. P. (2015). Acetyl-CoA and the regulation of metabolism: mechanisms and consequences. *Current Opinion in Cell Biology*, *33*, 125–131. http://doi.org/10.1016/j.ceb.2015.02.003

Sieber, J. R., Crable, B. R., Sheik, C. S., Hurst, G. B., Rohlin, L., Gunsalus, R. P., & McInerney, M. J. (2015). Proteomic analysis reveals metabolic and regulatory systems involved in the syntrophic and axenic lifestyle of Syntrophomonas wolfei. *Frontiers in Microbiology*, *6*(February), 1–9. http://doi.org/10.3389/fmicb.2015.00115

Sieber, J. R., Le, H. M., & Mcinerney, M. J. (2014). The importance of hydrogen and formate transfer for syntrophic fatty, aromatic and alicyclic metabolism. *Environmental Microbiology*, *16*, 177–188. http://doi.org/10.1111/1462-2920.12269

Sieber, J. R., Sims, D. R., Han, C., Kim, E., Lykidis, A., Lapidus, A. L., … McInerney, M. J. (2010). The genome of Syntrophomonas wolfei: new insights into syntrophic metabolism and

biohydrogen production. *Environmental Microbiology*, *12*(8), 2289–301. http://doi.org/10.1111/j.1462-2920.2010.02237.x

Smith, L. M., & Kelleher, N. L. (2013). Proteoform: a single term describing protein complexity. *Nature Methods*, *10*(3), 186–7. http://doi.org/10.1038/nmeth.2369

Stahl, G., McCarty, G. P., & Farabaugh, P. J. (2002). Ribosome structure: Revisiting the connection between translational accuracy and unconventional decoding. *Trends in Biochemical Sciences, 27,* 178-183. http://doi.org/10.1016/S0968-0004(02)02064-9

Sund, J., Andér, M., & Aqvist, J. (2010). Principles of stop-codon reading on the ribosome. *Nature*, *465*, 947–950. http://doi.org/10.1038/nature09082

Tan, M., Luo, H., Lee, S., Jin, F., Yang, J. S., Montellier, E., … Zhao, Y. (2011). Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell*, *146*(6), 1016–1028.

Tan, M., Peng, C., Anderson, K. A., Chhoy, P., Xie, Z., Dai, L., … Zhao, Y. (2014). Lysine glutarylation is a protein posttranslational modification regulated by SIRT5. *Cell Metabolism*, *19*(4), 605–617.

Touriol, C., Bornes, S., Bonnal, S., Audigier, S., Prats, H., Prats, A. C., & Vagner, S. (2003). Generation of protein isoform diversity by alternative initiation of translation at non-AUG codons. *Biology of the Cell, 95,* 169-178. http://doi.org/10.1016/S0248-4900(03)00033-9

Tran, J., Zamdborg, L., Ahlf, D., Lee, J., Catherman, A., Durbin, K., … Kelleher, N. . (2011). Mapping intact protein isoforms in discovery mode using top-down proteomics. *Nature, 480,* 254-258. http://doi.org/10.1038/nature10575

Trelle, M. B., & Jensen, O. N. (2008). Utility of immonium ions for assignment of epsilon-N-acetyllysine-containing peptides by tandem mass spectrometry. *Analytical Chemistry*, *80*, 3422–3430. http://doi.org/10.1021/ac800005n

Verdin, E., & Ott, M. (2013). Acetylphosphate: A novel link between Lysine Acetylation and intermediary metabolism in bacteria. *Molecular Cell, 51,* 132-134. http://doi.org/10.1016/j.molcel.2013.07.006

Wagner, G. R., & Hirschey, M. D. (2014). Nonenzymatic Protein Acylation as a Carbon Stress Regulated by Sirtuin Deacylases. *Molecular Cell*, *54*(1), 5–16. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/24725594

Wagner, G. R., & Payne, R. M. (2013). Widespread and enzyme-independent Nε-acetylation and Nε-succinylation of proteins in the chemical conditions of the mitochondrial matrix. *The Journal of Biological Chemistry*, *288*(40), 29036–45.

Wang, L., Tang, Y., Cole, P. A., & Marmorstein, R. (2008). Structure and chemistry of the p300/CBP and Rtt109 histone acetyltransferases: implications for histone acetyltransferase evolution and function. *Current Opinion in Structural Biology, 18,* 741-747. http://doi.org/10.1016/j.sbi.2008.09.004

Wang, Q., Zhang, Y., Yang, C., Xiong, H., Lin, Y., Yao, J., … Zhao, G.-P. (2010a). Acetylation of metabolic enzymes coordinates carbon source utilization and metabolic flux. *Science (New York, N.Y.)*, *327*(5968), 1004–1007.

Wang, Q., Zhang, Y., Yang, C., Xiong, H., Lin, Y., Yao, J., … Zhao, G.-P. (2010b). Acetylation of metabolic enzymes coordinates carbon source utilization and metabolic flux. *Science (New York, N.Y.)*, *327*, 1004–1007. http://doi.org/10.1126/science.1179687

Weinert, B., Iesmantavicius, V., Wagner, S., Schölz, C., Gummesson, B., Beli, P., … Choudhary, C. (2013). Acetyl-Phosphate is a critical determinant of Lysine Acetylation in E.coli. *Molecular Cell*, *51*, 265–272. http://doi.org/10.1016/j.molcel.2013.06.003

Weinert, B. T., Iesmantavicius, V., Moustafa, T., Schölz, C., Wagner, S. a, Magnes, C., … Choudhary, C. (2014). Acetylation dynamics and stoichiometry in Saccharomyces cerevisiae. *Molecular Systems Biology*, *10*, 716. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/24489116

Weinert, B. T., Schölz, C., Wagner, S. A., Iesmantavicius, V., Su, D., Daniel, J. A., & Choudhary, C. (2013). Lysine succinylation is a frequently occurring modification in prokaryotes and eukaryotes and extensively overlaps with acetylation. *Cell Reports*, *4*, 842–851. http://doi.org/10.1016/j.celrep.2013.07.024

Weinert, B. T., Wagner, S. A., Horn, H., Henriksen, P., Liu, W. R., Olsen, J. V, … Choudhary, C. (2011). Proteome-wide mapping of the Drosophila acetylome demonstrates a high degree of conservation of lysine acetylation. *Science Signaling*, *4*, ra48. http://doi.org/10.1126/scisignal.2001902

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

Yang, X. J. (2004). The diverse superfamily of lysine acetyltransferases and their roles in leukemia and other diseases. *Nucleic Acids Research*.

Yuan, H., & Marmorstein, R. (2013). Histone acetyltransferases: Rising ancient counterparts to protein kinases. *Biopolymers, 99,* 98-111. http://doi.org/10.1002/bip.22128

Yvon-Durocher, G., Allen, A. P., Bastviken, D., Conrad, R., Gudasz, C., St-Pierre, A., … del Giorgio, P. a. (2014). Methane fluxes show consistent temperature dependence across microbial to ecosystem scales. *Nature*, *507*(7493), 488–91. http://doi.org/10.1038/nature13164

Zhang, Z., Tan, M., Xie, Z., Dai, L., Chen, Y., & Zhao, Y. (2011). Identification of lysine succinylation as a new post-translational modification. *Nature Chemical Biology*, *7*(1), 58–63.

Zhou, H., Ning, Z., E. Starr, A., Abu-Farha, M., & Figeys, D. (2012). Advancements in top-down proteomics. *Analytical Chemistry, 84,* 720-734. http://doi.org/10.1021/ac202882y