Integrated Top-down and Bottom-up Mass Spectrometry Characterization of *Escherichia coli* Ribosomal Protein Heterogeneity - Identification of Protein Isoforms and Post-translational Modifications

Kevin R. Ramkissoon

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Microbiology and Immunology.

Chapel Hill
2010

Approved by:
Advisor: Morgan C. Giddings, Ph.D.
Reader: Robert B. Bourret, Ph.D.
Reader: Janne G. Cannon, Ph.D.
Reader: Aravinda M. DeSilva, Ph.D.
Reader: Robert A. Nicholas, Ph.D.

Abstract

KEVIN R. RAMKISSOON: Integrated Top-down and Bottom-up Mass Spectrometry Characterization of *Escherichia coli* Ribosomal Protein Heterogeneity - Identification of Protein Isoforms and Post-translational Modifications
(Under the direction of Morgan C. Giddings)

The bacterial genome exhibits notable plasticity but is relatively static when compared to the proteome. Protein expression can vary significantly depending on environmental factors, growth stage or stochastic processes within cells. This highly variable character, coupled with the large dynamic range of protein expression levels and the complexity achieved through processes such as post-translational modification (PTM), necessitate accurate, sensitive and high-throughput methods of analysis. The primary aim of this research was to develop an integrated experimental and analysis workflow that combines the analytical power of top-down and bottom-up mass spectrometry towards protein isoform and PTM characterization. We apply this approach to a comprehensive characterization of *Escherichia coli* ribosomal protein isoform heterogeneity. Our findings uncovered a significant level of heterogeneity in the post-translational modification of a number of ribosomal proteins, revealing a possible mechanism for the regulation of ribosomal protein function both within and beyond the ribosome.

To my family.

For their unwavering love,

support and understanding.

Acknowledgements

The list of individuals I owe gratitude is long. I thank my advisor, Dr. Morgan Giddings, for her guidance along the path to becoming a scientist. The members of my committee were all pivotal to my success, offering support and scientific advice each time it was needed. I am indebted to scientific collaborators at Oak Ridge National Laboratories and at the UNC Proteomics Core Facility for assistance with data generation, data analysis and technical training. I thank Dr. Sharon Milgram for her advice and guidance that helped ensure my transition to graduate school was both smooth and successful.

I offer a very special thanks to present and past members of the Giddings lab who quickly became a second family. I count them amongst the many friends and loved ones who made my time in graduate school enjoyable and helped make UNC a home away from home.

Finally, I thank my family. Everything I have achieved thus far in life, I owe in large part to them.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

2-D PAGE	2-dimensional polyacrylamide gel electrophoresis
ACN	acetonitrile
DHB	2,5-dihydroxybenzoic acid
CID	collision-induced dissociation
ECD	electron capture dissociation
ETD	electron transfer dissociation
ESI	electrospray ionization
FA	formic acid
FT-ICR	Fourier transform ion cyclotron resonance
HPLC	high performance liquid chromatography
IEC	ion exchange chromatography
IEF	isoelectric focusing
IRMPD	infared multiphoton dissociation
LC	liquid chromatography
LC-MS	liquid chromatography-mass spectrometry
MALDI	matrix-assisted laser desorption/ionization
MeOH	methanol
MS	mass spectrometry
MS/MS	tandem mass spectrometry
PCR	polymerase chain reaction
pI	isoelectric point
PMF	peptide mass fingerprint

PPM	parts per million
PTM	post-translational modification
Q	quadrupole
RF	radio frequency
RP	reversed phase
SDS	sodium dodecyl sulphate
SEC	size exclusion chromatography
SmR	streptomycin-resistant
SmRC	streptomycin-resistant/compensated
TOF	time of flight
WT	wild-type

Chapter 1

Introduction

1.1 Bacterial Phenotypic Heterogeneity

Pathogenic bacteria are responsible for significant human morbidity and mortality across a broad spectrum of infectious human diseases including tuberculosis, diarrheal diseases and lower respiratory infections - amongst the top ten causes of death worldwide according to the World Health Organization. The plasticity of bacterial genotypes and phenotypes is one of the biggest contributors to their tremendous successes in pathogenesis, immune system evasion and their increasing tolerance to antibiotics[1-3, 4-8]. The first and critical step towards solving many of the challenges posed by bacterial pathogens is understanding the nature of their complex and highly variable physiology. Heterogeneity within bacterial populations can be achieved through a number of mechanisms, which can be broadly divided into genetically-based or phenotypically-based heterogeneity. Phenotypic heterogeneity refers to the observed phenomenon wherein individual cells within an isogenic bacterial population achieve different phenotypes whilst exposed to the same environmental conditions[9]. As with genetic heterogeneity, phenotypic heterogeneity enhances the overall fitness of a population and may sometimes be heritable[10]. On the level of an individual cell, changes in a bacteria's phenotype may be the result of stochastic mechanisms or may occur in response to

environmental changes, such as may be encountered during growth or pathogenesis. Mechanisms such as positive feedback regulation of gene expression, bi-/multi-stable gene networks or variable protein post-translational modification have all been shown to be mechanisms employed by cells in altering their phenotype[11, 12]; (figure 1.1). The results may be localized to a specific pathway or cellular function (eg. competence)[13], or alternatively result in global physiological changes (eg. sporulation)[14]. Regardless of the mechanism employed, in the vast majority of cases, the underlying ultimate step of any phenotypic change occurs at the level of protein expression and/or post-translational modification, hence making the analysis of expressed proteins a key element of any examination of bacterial heterogeneity.

1.2 The Era of the 'omics'

Over the last two decades or so there have been fundamental changes in the approaches to understanding bacterial physiology. The genomics era yielded a tremendous amount of information on, and insight into, the genetic makeup of bacteria. As important as the organisms that have been sequenced thus far, the technologies that emerged from the genomic revolution facilitated the advent of high-throughput DNA sequencing, assuring the influx of bacterial genome sequences will continue. Genome sequences provide a virtual blueprint of cellular components but provide little to no information on which genes are expressed or when. With the determination of DNA sequence no longer a bottleneck in the process of bacterial characterization, and with the challenge of decoding the meaningful information contained in the vast amount of sequences in hand mounting, science turned towards more global approaches to characterizing bacteria. The disciplines of transcriptomics, functional genomics and proteomics came to the fro in an almost

natural scientific evolution to begin to address the pressing "which?", "whens?", and "hows?" of bacterial gene and protein expression. Each of these fields take a global approach to the study of organisms by attempting to analyze the full complement, or large subsets thereof, of the mRNA transcripts (transcriptome) and proteins (proteome) produced by an organism under given conditions[15-18]. Unlike the more reductionist single-gene or pathway-focused studies, such global, high-throughput techniques have a much greater potential to identify gene and protein networks that may act in concert to engender phenotypic changes in bacteria. Such methods are however accompanied by significant challenges, particularly in the management, integration and interpretation of the vast amount of data generated.

1.3 Proteomics

The term 'proteome' was coined to describe the set of proteins encoded by a study of the complement of proteins expressed by a cell, tissue or organism. The field's overall goal is not simply to identify which proteins are expressed, but to fully characterize their expression, localization, structure and interactions towards the elucidation their functional significance to cellular survival[19]. Proteins are responsible for the majority of the biochemical processes in cells, making the ability to identify and characterize those associated with particular cellular phenotypes of critical importance. Traditional approaches to protein characterization have focused on single proteins and pathways, or a small family of related proteins at a time.

While invaluable to understanding prokaryotic biology, traditional techniques such as 1- dimensional polyacrylamide gel electrophoresis (1-D PAGE) and western blotting are not well suited to high-throughput protein analyses. One of the key limitations faced

by protein or pathway-focused studies is that context may sometimes be lost as in many cases simultaneous or sequential changes in the expression, localization or activation of multiple, often functionally unrelated proteins, are involved in effecting phenotypic change.

The advantages to a more global approach like gene expression profiling or proteome analysis stem from an enhanced ability to discern cell-wide patterns in protein expression, modification, regulation and interactions, allowing the elucidation of intracellular dynamics that may not be apparent from the analysis of proteins or pathways in isolation. However as is often the case, a greater potential for finding answers comes with greater challenges.

1.3.1 The Challenges of Proteomic Analysis

While genomes are by comparison relatively static, proteomes are constantly in flux, with protein expression, degradation and post-translational modification (PTM) prone to change during development or in response to changes in physiological and environmental conditions. Proteomic efforts face fundamentally different and more challenging problems than those faced by genome sequencing efforts[15, 20]. With technologies like polymerase chain reaction (PCR) and automated gene sequencing already established, the challenge of genome sequencing was primarily one of scale. The ability to copy and amplify nucleic acid sequences also meant that limited sample was a surmountable problem. Equivalent technologies to amplify protein from limited sample are not available. Compounding potential sample limitation problems is the inherent instability of protein compared to DNA. Even in the absence of proteases proteins may undergo degradation, including but not limited to oxidation, making the preparation,

storage and preservation of samples critically important to sample fidelity. The large dynamic range of protein expression within cells, which can be greater than 106-fold, also poses significant challenges to protein analysis. It is difficult enough to attempt to analyze proteins present only in a few copies in a cell (such as transcription factors), much less to attempt their detection within a complex milieu of other proteins, many of which are expressed at orders of magnitude higher abundance (such as ribosomes). Proteins may also be modified post-translationally. Post-translational modification of proteins may be static, or as with their expression, may be variable, adding complexity to the proteome far beyond that encoded by the genome.

1.4 Protein Post-translational Modification

Post-translational modification (PTM) of proteins is a process wherein polypeptides are enzymatically or chemically modified during or after translation. It represents a mechanism whereby cells can achieve a far greater diversity in protein expression and function than allowed for by the genetic code. PTMs, which include covalent modifications and protein cleavage, alter the structural and biochemical properties of proteins and play diverse roles in cellular physiology[21]. Hundreds of PTMs have been identified, with tens of thousands of proteins being targeted[22] (http:// dbptm.mbc.nctu.edu.tw/statistics.php). In bacteria, where alternative splicing of genes does not occur, protein PTM represents a major contributor to protein isoform heterogeneity and proteome complexity. In addition to providing a mechanism whereby cells may rapidly respond to physiological changes in cellular environment, PTMs may allow cells to forego the time-consuming and resource-expensive dependence of altering gene transcription and protein translation profiles. Amongst the more common bacterial

protein PTMs are N-terminal methionine cleavage, methylation, acetylation and phosphorylation[23-29]. Aside from its role in protein maturation, bacteria employ PTMs in the regulation of critical cellular processes such as signal transduction[30-32], chemotaxis[33-36] and protein translation[29, 37, 38]. The ability to identify expressed proteins and characterize both the presence and locations of any PTMs is essential to understanding bacterial protein heterogeneity and by extension the numerous physiological changes for which they are responsible.

1.4.1 Ribosomal Protein Post-translational Modification

The bacterial ribosome consists of protein and ribonucleic acid (RNA) in a bipartite structure, forming a macromolecular machine at the heart of the cellular biogenesis[39]. Ribosomes are responsible for all polypeptide synthesis in the cell, translating the messages encoded in bacterial genomes to protein by decoding the intermediary messenger RNA (mRNA). Both ribosomal structure and function is highly conserved in prokaryotes. It has been the focus of intense structural, biochemical and functional research for decades yet still there remains much we do not understand about its function[40-49]. Many PTMs have been discovered on ribosomal proteins but their functions remains largely a mystery[37, 38, 50]. The high level of conservation of the metabolically expensive process of ribosomal protein modification however suggests that their true significance to the cell remains to be uncovered[51].

One likely reason for this apparent incongruity is simply that the conditions under which the modifications are most apparent still remain to be uncovered. Indeed in the cases where phenotypes have been associated with of the loss of ribosomal protein modification they have been found to be growth condition associated[52-55]. Another

possible reason is that these studies have primarily focused on roles for the modifications on ribosomal assembly and translation rates.

There are two intriguing avenues through which ribosomal protein modification may yet prove to exert a strong influence that remain to be examined closely. One such avenue lies in the extraribosomal functions of ribosomal proteins i.e. functions of individual proteins beyond their roles as part of the holoribosome. Speculation and interest in extraribosomal functions of ribosomal proteins have been growing in the face of mounting evidence that ribosomal proteins exert influence far beyond the realm of translation[56-58]. While the majority of such examples are currently from eukaryotes, several ribosomal proteins have been shown to play key roles in the complex regulation of ribosome biogenesis in bacteria[42]. In the majority of elucidated cases this regulation is achieved through interaction of a protein with its own, often polycistronic, mRNA. Whether individual proteins' roles in transcriptional regulation is relegated to translation-related transcripts or extends to other mRNA species is an area of ongoing research interest.

A second area where ribosomal protein PTM may play a significant role in cellular physiology is in modulating which mRNA transcripts are translated by the ribosome under different conditions. Mauro and Edelman have proposed that the ribosome itself acts as a regulator of translation by directly and specifically interacting with cellular mRNA via both rRNA and ribosomal proteins, in a mRNA sequence-specific manner[59, 60]. In the aptly-termed *ribosome filter hypothesis* it is the ribosome as a whole, inclusive of rRNA, that regulates or 'filters' which mRNA are preferentially translated under different conditions. Thus far the majority of examples cited in support

of the filter hypothesis have been in eukaryotes. The prevailing view is that the ribosome-mRNA interactions are modulated through heterogeneity within the ribosomes in cells.

To ascertain the potential for heterogeneity within bacterial ribosomal protein as a mechanism via which the ribosome could influence translation, we sought to characterize the heterogeneity present within the ribosomal proteins of the Gram-negative model bacterium, *Escherichia coli*.

The *E. coli* ribosome consists of 3 ribosomal RNA (rRNA) molecules and 53 proteins divided into a large (50S) and a small (30S) subunit. A characterization of ribosomal heterogeneity faces many of the same challenges discussed in section 1.2.1. Ribosomal proteins are high in abundance, accounting for over 40% of the dry cell mass in *E. coli*. However potential secondary isoforms of individual ribosomal proteins may be many orders of magnitude less than that of the primary isoform. Identifying any low-abundant isoforms would be additionally stymied by other highly-abundant ribosomal protein species. These factors combine to make the sensitivity of protein detection critical to the success of any heterogeneity analysis.

The prevalance of post-translational modification on ribosomal proteins adds to the challenge of protein identification and characterization, not only because it increases the complexity of the protein mixture in question, but also because PTMs result in a final product whose physical and chemical properties cannot be accurately predicted from the known gene sequences. As such, any successful characterization of *E. coli* ribosomal protein heterogeneity would also require the ability to identify protein isoforms due to any combination of PTMs.

1.5 Mass Spectrometry

The multiple challenges presented by the propsed analysis of ribosomal protein heterogeneity necessitated the use of analytical techniques that offered high sensitivity, accuracy and resolution in order to identify and distinguish between various combinations of PTMs and the resulting protein isoforms. Mass spectrometry (MS) provided just such a technique.

Mass spectrometry is currently one of the most powerful and widely employed analytical methods for the high-throughput analysis of proteins and other biomolecules[61-63]. It involves determination of the precise mass of ionized, gaseous phase analytes. It facilitates rapid analysis of large, complex samples with sufficient sensitivity to detect low-abundance proteins. Critically, MS accomplishes all this at a high enough resolution to distinguish between analytes that may differ only by a modification to, or substitution of, a single amino acid in the polypeptide chain. In the proteomics arena, mass spectrometry is most frequently employed in the analysis of proteins and peptides. The information generated from mass spectrometric analyses can be employed not only to identify proteins present in a sample, but also measure their relative abundances, identify post-translational modifications and protein isoforms, probe their structure, as well as investigate intra- and inter-molecular interactions.

1.5.1 Top-down and Bottom-up Mass Spectrometry Analysis

MS-based protein analyses can be divided into two general approaches – top-down and bottom-up mass spectrometry. The term **top-down** mass spectrometry is used to refer to methods wherein intact proteins are introduced to the mass spectrometer for analysis, sometimes coupled with fragmentation of the protein within the instrument to

obtain sequence information[64, 65]. **Bottom-up** mass spectrometry refers to the analysis of peptides rather than intact proteins and includes MudPIT analysis[66]. In bottom-up analyses protein mixtures are typically digested with a protease, such as trypsin, and the resulting peptides subjected to MS analysis. Compared to proteins, peptides are less troublesome to ionize and analyze by MS. This is partly due to the much smaller mass and chemical space they occupy. Bottom-up analysis also offers multiple opportunities for identifying proteins in a sample as multiple peptides are produced from the digestion of each protein.

While both methods are extremely powerful and useful towards protein analysis, each has inherent benefits and limitations as it pertains to the tasks of protein identification and characterization. Top-down MS allows for analysis of proteins in a form closest to their native cellular state. To assign protein identities from top-down mass measurements the experimental masses are compared to databases populated with protein molecular weight values derived from previous experimental measurements or predicted by an *in silico* translation of gene sequences (figure 1.2). While efficient front-end separation coupled with accurate, high-resolution mass measurements may be sufficient to allow for protein identification in some cases, distinguishing between isobaric proteins (i.e. proteins with the same molecular weight), and full characterization are not possible by this method alone.

Further confounding top-down MS-based protein identification are the numerous biological processes that result in changes to a protein's mass from that which is encoded by its gene sequence. Genetic mutations can result in amino acid substitutions or the production of truncated protein products. Protein post-translational modification (PTM)

may also result in differences between a mature protein's mass and that which is predicted from its gene sequence. Bottom-up MS has the advantage of being able to distinguish between isobaric proteins as identification is not based on a single mass measurement but on the sum of measurements of constitutive peptides.

While two proteins may have similar or identical masses, differences in their amino acid sequence usually result in a unique pattern of peptides when digested with cleavage site-specific proteases. The resulting peptide masses or 'peptide mass fingerprint; (PMF) facilitate identification of the originating protein[67-74]. In PMF, a list of experimentally measured peptide masses is compared to a theoretical mass list generated through an *in silico* digestion of either a database of known protein sequences or alternatively, a database of proteins generated through an *in silico* translation of the entire genome[75] (figure 1.3.) Statistical analysis of the resulting peptide mass matches enables determination of the most likely protein(s) from which the peptides originated.

While PMF is a fairly standard technique for protein identification, analyses based solely on bottom-up methods have inherent limitations for protein characterization as well. The primary obstacle to protein characterization using bottom-up data is that such methods rarely achieve 100% amino acid sequence coverage of the protein. The result is that post-translational modifications may go potentially unobserved with no indication, leaving them in the category of "unknown unknowns." Another critical limitation is that all context as to the combinations in which any observed PTMs may occur is lost. While the observation of modified peptides may indicate that PTM of the protein occurs, some modifications may occur concurrently while others may be mutually exclusive. This leads to significant challenges in distinguishing between variable protein isoforms. The sum of

the limitations is a reduced ability to discern important physical and biochemical characteristics on the biologically active, intact protein.

1.5.2 Tandem Mass Spectrometry

Whether proteins or peptides are introduced to the mass spectrometer, further structural data on the sequence of amino acids that constitute the peptide/proteins can be gained from fragmentation of the molecules within the mass spectrometer in **tandem mass spectrometry** (MS/MS)[76, 77]. Tandem mass spectrometry (MS/MS) applies fragmentation combined with a second stage of mass spectrometry to obtain sequence and structure data directly from introduced proteins or peptides.

Tandem mass spectrometry (MS/MS) is most widely applied in bottom-up proteomics analyses where peptides are fragmented within the mass spectrometer and mass of the resulting fragments measured. The breaking of intramolecular peptide bonds may be achieved through any number of processes that most commonly include collision-induced dissociation (CID), infared multiphoton dissociation (IRMPD), electron capture dissociation (ECD) and electron transfer dissociation (ETD)[78-81]. Each method produces a characteristic distribution of fragment ions[82-86] with fragmentation possible both along and outside of the peptide backbone (figure 1.4). Fragmentation of the peptide bonds produces b-ions and y-ions resulting in a spectrum-representation of polypeptide amino acid sequence (figure 1.5). Ideally MS/MS would allow for near complete sequencing of a polypeptide chain facilitating the localization of PTMs to specific amino acid residues. In practice however, 100% sequence coverage is seldom achieved. Short amino acid sequences, or *sequence tags*, can however be sufficient for unambiguous identification of the parent peptide ion, and often, the originating protein.

Fragmentation can also be applied to obtain structural data from intact proteins[87-89]. However difficulties associated with creating and analyzing gas-phase ions of large proteins make such analyses challenging. The instrumentation to fragment proteins, as well as the informatic tools to interpret the resulting data, are also still under development. As such, the fragmentation of intact proteins has yet to become a mainstream analytical method within the proteomics community.

1.6 Instrumentation

While the instrumentation employed in mass spectrometry experiments varies, there are several constant parameters that demand consideration. These include *mass accuracy*, which refers to the difference between the measured mass and the calculated mass of the analyte, (typically measured as percent error or parts-per-million (ppm)), *detection sensitivity* - a measure of how well an instrument can detect small concentrations of an analyte, and instrument *resolution* which is a measure of how well adjacent masses, represented as peaks in a mass spectrum, can be distinguished from each other (typically measured as the full width at half maximum (FWHM) of the peaks.) The types of, and the configurations in which, the three basic components of a mass spectrometer - the ion source, mass analyzer and detector - are employed largely determine the overall performance characteristics of a mass spectrometer (figure 1.6).

The ion source is responsible for the aerosolization, vaporization and ionization of analytes. The two ionization methods I employed in my dissertation work were electrospray ionization (ESI) and matrix-assisted laser desorption/ionizaion (MALDI)[90-93]. Both are examples of gentle or 'soft' ionization techniques, which circumvent the propensity for large biomolecules like proteins to fragment when ionized. Electrospray

ionization is the primary technique used for the introduction of liquid solubilized protein to the mass spectrometer. In ESI solubilized protein is dispersed into highly charged, aerosolized droplets by the application of a high electric potential between the liquid in a thin metal capillary and a counter electrode. Subsequent to aerosolization, the liquid in which the proteins are suspended is evaporated away leaving behind the charged analyte in the gaseous phase (figure 1.7). To enhance both ionization and evaporation, proteins are often solubilized in a mixture of water and highly volatile organic solvents such as acetonitrile (ACN) or methanol (MeOH). The solutions typically also contain a charge/ proton source such as formic acid (FA) to enhance ionization and increase the concentration of charge on the droplets produced. ESI is particularly useful for producing charged macromolecules such as proteins and peptides. In MALDI, analytes are sublimated from the solid phase directly to a vapor out of a dry, crystalline matrix using a laser. The most common matrices used are composed of one of alpha-cyano-4-hydroxycinnamic acid (alpha-cyano), 3,5-dimethoxy-4-hydroxycinnamic acid (sinapinic acid) or 2,5-dihydroxybenzoic acid (DHB) dissolved in a solvent much like ESI. The matrix enhances vaporization and ionization, providing a proton source while protecting the analyte from direct exposure to the laser, which would otherwise result in its disintegration. While MALDI primarily results in the production of singly charged ions , ESI by nature produces more complex spectra consisting of a distribution multiply charged ions. As a result, ESI of a single protein species results in spectra which contain multiple peaks at different mass-to-charge ratios (m/z) due to the multiple charge states in which the ions exist (see figure 1.8).

The detector provides the output of a mass spectrometer, measuring the current

produced upon detection of ions by one of several methods. The type of mass analyser employed has a strong influence on mass accuracy, sensitivity and resolution. Three commonly employed mass analyzers in the proteomics arena are - time-of -flight (TOF), quadrupole (Q) and Fourier transform ion cyclotron resonance (FT-ICR). I applied each of these technologies in my research, leveraging the strengths of each.

Time-of-flight instruments employ a constant electric field to accelerate charged analytes and measure the time required for the ions to reach the detector after traveling a known distance through a vacuum[94, 95]. The ions are separated during flight due to differences in their velocities, a direct result of differences in their mass-to-charge ratios. Quadrupole instruments effectively filter the charged analytes in an oscillating electromagnetic field, allowing only those with certain mass-to-charge ratios to reach the detector. This is typically achieved by varying a combination of radio frequency (RF) and current potentials which in turn alter the trajectories of the charged molecules as they pass between the quadrupoles[96, 97].

FT-ICR instruments measure the image current produced by the ions cyclotroning (spiraling) in a static magnetic field, in which the mass-to-charge ratio of a molecule determines the frequency of its cycle in the magnetic field[98-100]. Detectors located within the analyzer measure the electrical signal produced as ions repeatedly cycle nearby, producing a periodic signal that can be deconvoluted into a mass/charge measurement by performing a Fourier transform. Unlike TOF-based and quadrupole-based mass spectrometers, where the protein/peptide ions are ultimately deposited at the detector, ion detection in FT-ICR instruments is non-destructive, allowing ions to be 'detected' multiple times as they oscillate within the mass analyzer. This is one of the key factors

underlying the high sensitivity, mass accuracy and resolution typical of FT-ICR instruments. These characteristics make FT-ICR MS ideal for intact protein analyses where the large molecular masses are accompanied by large isotopic distributions. Isotopic distribution of protein mass is a result of the naturally occurring abundances of different isotopes of the constituent elements. The probability of slightly different intact masses due to incorporation of different elemental isotopes (e.g. carbon-12 (12C) versus the lower abundance carbon-13 (13C) isotope) increases with molecular size. The benefits of FT- ICR are however offset somewhat by the longer time required for data acquisition and costly nature and availability of the required instrumentation.

1.7 Protein Separation Techniques

The advances in mass spectrometry, particularly ionization and mass analyzer technologies, represent a major steps towards overcoming the challenges posed by the large dynamic range of protein concentration, as well as the need for high sensitivity, accuracy and resolving power. However regardless of the configuration of the instrumentation employed, conventional mass spectrometers would quickly be overwhelmed if faced with the simultaneous analysis of an entire proteome. The practical advantages of mass spectrometry to high-throughput protein analysis are only achievable if accompanied by efficient front-end separation of complex protein/peptide samples. Separation simplifies the chemical complexity of the analyte introduced to the mass analyzer at any particular time. It also facilitates MS analysis by reducing the effects of ion suppression, a phenomenon wherein one component in a mixture, be it protein, peptide or contaminant, suppresses the ionization of another. The end result of ion suppression is a disproportionate underrepresentation of the suppressed species in the

resulting mass spectra. This can be a critical problem when dealing with low-abundant protein isoforms or when accurate quantitation is necessary. Efficient front-end separation also facilitates analysis by simplifying the resulting mass spectra, often already complicated by factors such as isotopic spread and overlapping protein charge envelopes. This in turn facilitates more accurate charge state deconvolution and mass determination. Proteins and peptides can be separated by physical properties such as size or molecular weight, or by chemical properties such as isoelectric point (pI) or hydrophobicity. Depending on the complexity of samples to be analyzed, either one-dimensional or multi-dimensional separation schemes may be necessary. Two of the most powerful and common techniques employed for protein and peptide separation prior to MS analysis that can achieve orthogonal separations are 2-dimensional polyacrylamide gel electrophoresis (2-D PAGE) and high performance liquid chromatography methods (HPLC).

1.7.1 Two-dimensional Polyacrylamide Gel Electrophoresis (2-D PAGE)

Two-dimensional polyacrylamide gel electrophoresis (2-D PAGE) first separates proteins based on their isoelectric point (pI) in a process known as isoelectric focusing (IEF), then separates based on size[101-103]. 2-D PAGE offers a significant increase in resolving power over 1-D gels, and while the detection of proteins at nanogram levels is possible, important limitations remain[104]. 2-D PAGE is quite labor intensive, not systematically reproducible, and not ideally suited for high-throughput analyses. Limitations on gel loading capacities and dye sensitivity can also hinder the ability to observe low-abundance proteins and isoforms in the gel. Highly hydrophobic proteins, such as membrane proteins, as well as highly basic or acidic proteins, such as ribosomal

protein L7/12, can also prove challenging to resolve in 2-D PAGE analyses. Perhaps most critically, it is difficult to extract intact proteins from the gel matrix removing the potential for intact MS analysis. As such proteins separated by 2-D PAGE are typically subjected to in-gel protease digestion with the extracted peptides subject to bottom-up analysis. Even then, 2-D PAGE is not directly compatible with MS analysis as the detergent SDS, necessary for protein separation, can severely suppress peptide ionization if not removed.

1.7.2 Liquid Chromatography

High performance liquid chromatography (HPLC) has become the technique of choice for the separation of complex protein and peptide mixtures prior to MS analysis[105]. In conventional HPLC, the proteins or peptides are solubilized in a liquid *mobile phase* and passed through columns packed with an immobilized *stationary phase* with which they interact. It is the strength and specificity of these interactions that is primarily responsible for the resulting separation. Increased interactions with the stationary phase result in increased retardation of movement through the column. Proteins/peptides that interact the least elute from the column earlier than those which experience stronger interactions. Separation thus occurs in time rather than location, with the time an analyte elutes from the column referred to as its *retention time*. Various physical and chemical properties of proteins may be exploited in LC including size (size exclusion chromatography, SEC), hydrophobicity (reversed-phase chromatography, RP) or charge (ion-exchange chromatography, IEC)[106]. Orthogonal separation techniques i.e. separations based on mutually exclusive analyte properties, may be sequentially performed. Significant increases in separating power and resolution may be achieved

over gel-based techniques through the coupling of multiple, orthogonal LC separation methods[107]. The high reproducibility of LC-based separations is also an advantage for proteomics, allowing putative identification of proteins based on their retention characteristics alone.

Whether single or multi-dimensional in nature, a reversed-phase separation is typically the final stage prior to introduction to the mass spectrometer. This is because the mobile phase in which the proteins elute consists mainly of water and a volatile organic solvent such as ACN or methanol, making it highly suitable for electrospray ionization. In addition to this flexibility, a major advantage of LC-based separation to mass spectrometry-based proteomics research is that LC separation of proteins and peptides can be carried out online with MS analysis; i.e. separation can be coupled directly to the ion source of the mass spectrometer. It is also much less labor intensive and as such more suited for high-throughput applications.

1.8 Summary

Mass spectrometry has quickly become the marquee analytical method for high-throughput protein analysis because of its ability to achieve high accuracy, sensitivity and resolution. While protein identification using MS is a very common practice, full protein characterization still poses a number of technical and computational challenges. The analysis of intact proteins by top-down MS methods, and the analysis of peptides using bottom-up MS methods are each powerful in their own rights, however significant limitations towards protein characterization are clearly apparent when either is considered alone.

The primary aim of this research was to develop an integrated experimental and

data analysis workflow that combined the strengths of each method towards the rapid characterization of bacterial protein heterogeneity due to variable PTM. This study primarily made use of reversed-phase HPLC for both on-line and off-line protein/peptide separations and utilized multiple mass spectrometers, encompassing two methods of ionization and three types of mass analyzers, in a comprehensive analysis of the level of heterogeneity present in *E. coli* ribosomal proteins. The methodology and results of this global ribo-protein analysis are presented in Chapter 2. A more focused MS-based analysis of the variable methylation of ribosomal protein L7/12 is presented in Chapter 3.

Figure 1.1

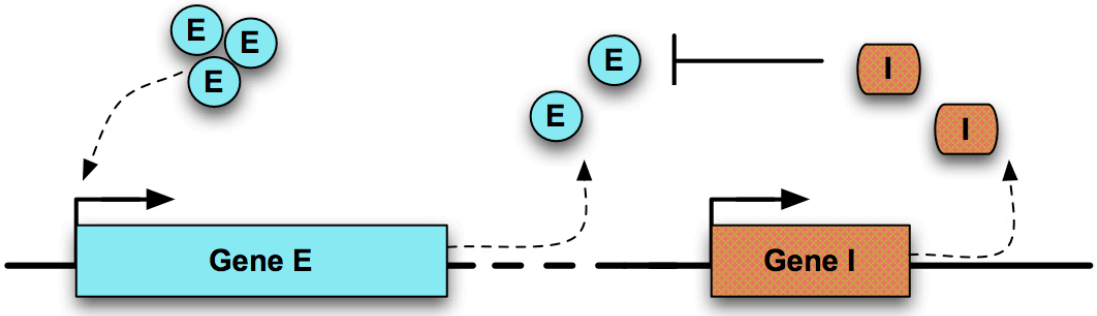

Figure 1.1. Schematic of a simple bistable switch. A simple bistable switch wherein autoregulatory feedback loops in protein expression drive the system to distinct stable phenotypic states, either in response to specific stimuli or spontaneously due to stochastic processes. Here gene product E acts directly as an enhancer of its own transcription. The concentration of protein E, regulated by the protease I, determines the 'on/off' state of the switch, switching on if the concentration of E exceeds the threshold necessary for initiation of the positive feedback loop.

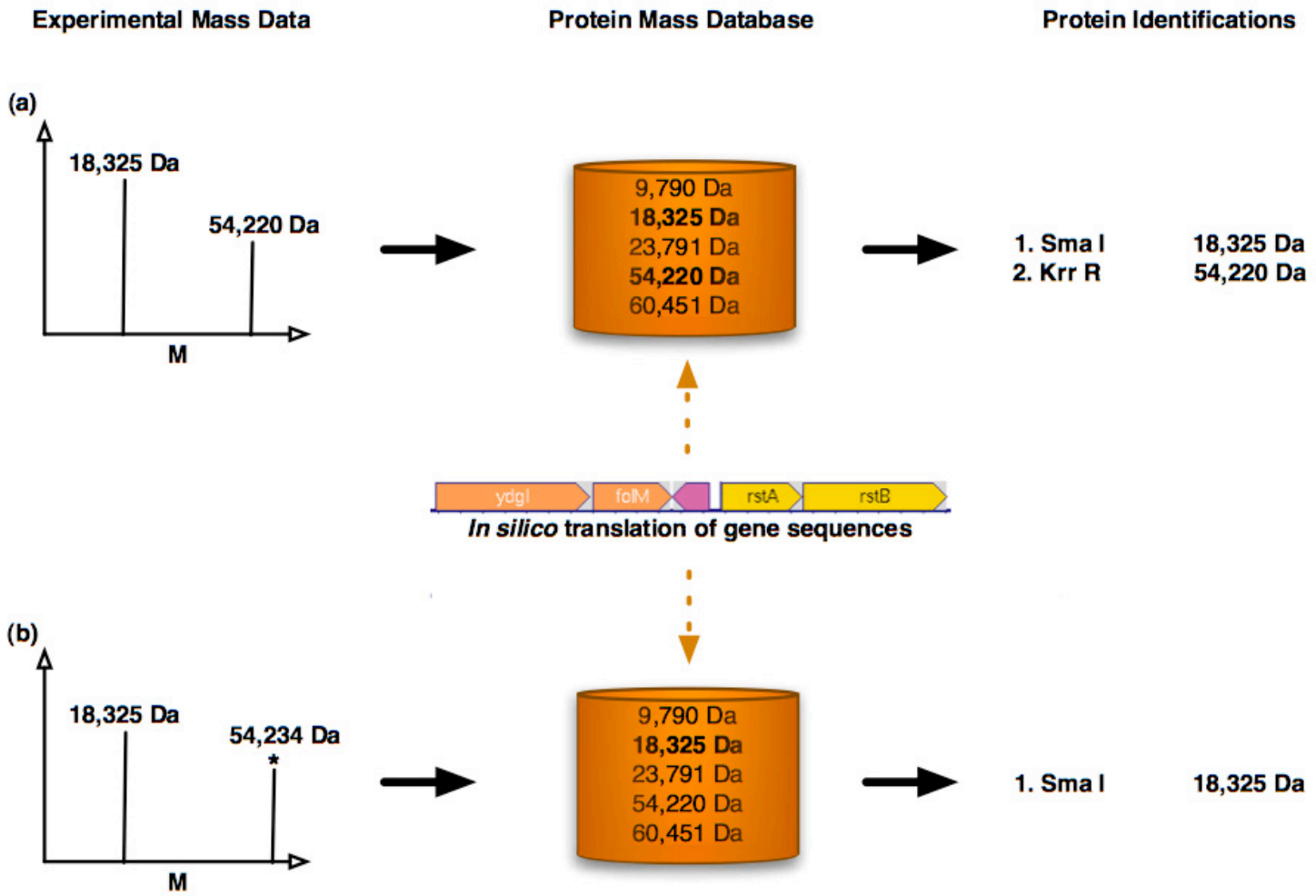
Experimental Mass Data
Protein Mass Database
Protein Identifications
(a)
18,325 Da
54,220 Da
M
9,790 Da
18,325 Da
23,791 Da
54,220 Da
60,451 Da
1. Sma I
2. Krr R
18,325 Da
54,220 Da
ydgI
fclM
rstA
rstB
In silico translation of gene sequences
(b)
18,325 Da
54,234 Da
*
M
9,790 Da
18,325 Da
23,791 Da
54,220 Da
60,451 Da
1. Sma I
18,325 Da

Figure 1.2

Figure 1.2. Protein identification using intact protein mass data. Experimental mass measures of intact proteins are compared to databases populated with protein masses derived from previous experimental data and/or an *in silico* translation of gene sequences. Masses are considered a match if they fall within a user-defined tolerance (typically a function of mass spectrometer accuracy.) Matches and identifications are typically scored and ranked based on software-specific scoring and statistical algorithms. (a) Both experimentally observed proteins Sma I and Krr R are identified through database searches. (b) Post-translational modification (*) of Krr R results in an experimental mass +14 Da larger than that encoded by its gene sequence resulting in a missed identification. PTMs can similarly result in incorrect database identifications.

Figure 1.3

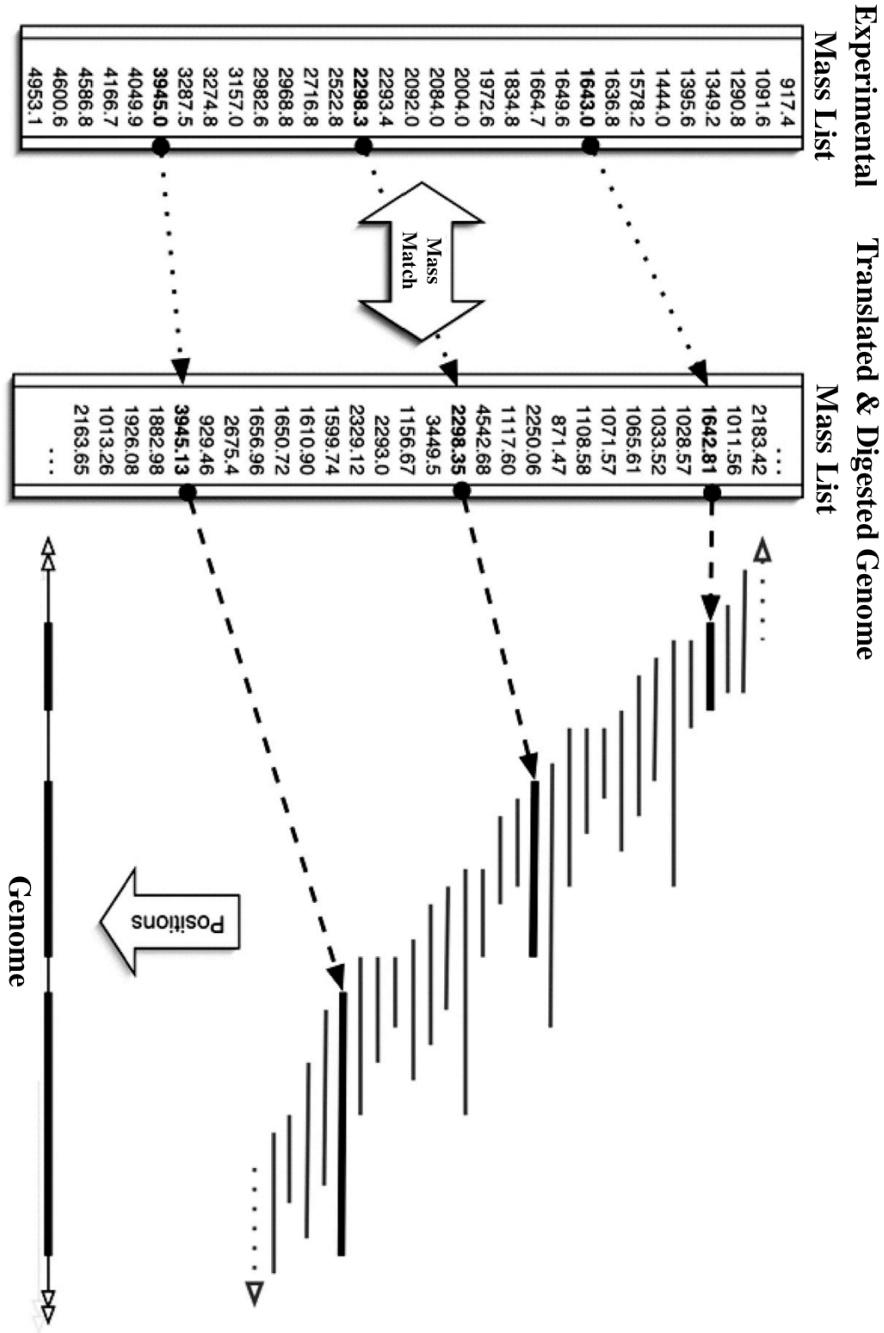
Experimental Mass List
917.4
1091.6
1290.8
1349.2
1395.6
1444.0
1578.2
1636.8
1643.0
1649.6
1664.7
1834.8
1972.6
2004.0
2084.0
2092.0
2293.4
2298.3
2522.8
2716.8
2968.8
2982.6
3157.0
3274.8
3287.5
3945.0
4049.9
4166.7
4586.8
4600.6
4953.1
Mass Match
Translated & Digested Genome Mass List
...
2183.42
1011.56
1642.81
1028.57
1033.52
1065.61
1071.57
1108.58
871.47
2250.06
1117.60
4542.68
2298.35
3449.5
1156.67
2293.0
2329.12
1599.74
1610.90
1650.72
1656.96
2675.4
929.46
3945.13
1882.98
1926.08
1013.26
2163.65
...
Positions
Genome

Figure 1.3. Protein identification from genome sequence using peptide mass data. Experimentally measured peptide masses are compared against a database of theoretical peptide masses generated from an *in silico* translation and digestion of the genome. Masses within a user-specified tolerance are considered a match. Matched peptides are mapped positionally back to the region of the genome from whence they originated. Clusters of matches are used to determine protein hits. Protein identifications may be derived directly from gene annotation if available. (Figure adapted from Giddings *et al.*[86])

Figure 1.4

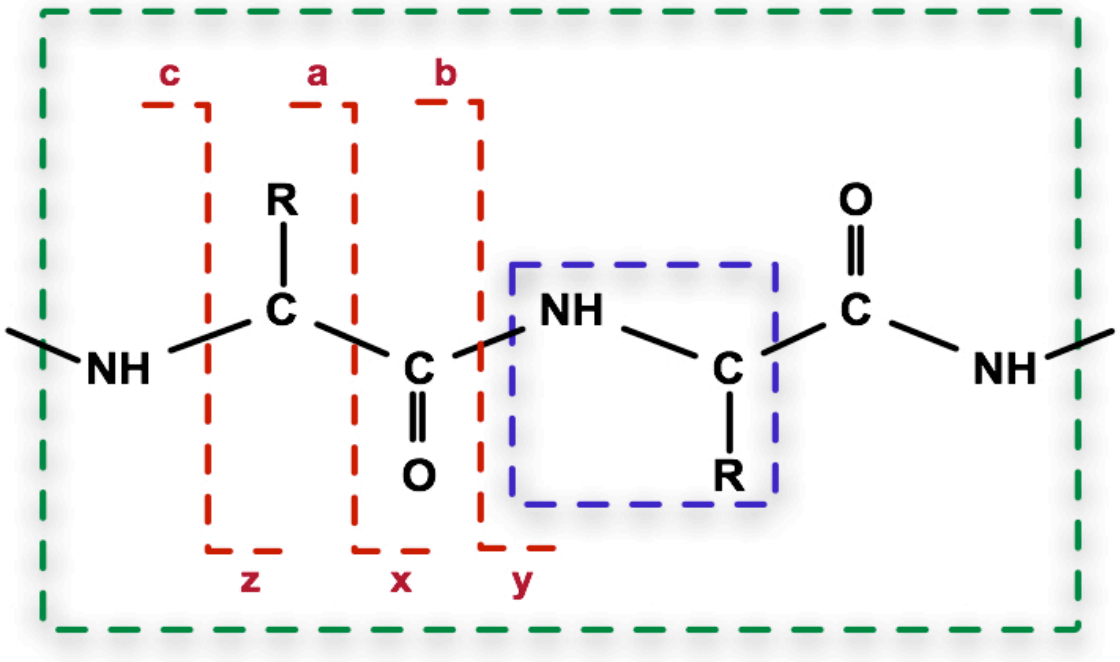

Figure 1.4. MS/MS fragmentation ions. Post-source fragmentation of peptides during MS/MS analysis produces multiple fragment ions that can be used to identify the amino acid composition of the parent ion. Breakage of a peptide bond results in the formation of b-ions and y-ions containing the N- and C-termini of the fragment respectively. Fragmentation along the peptide backbone can also produce a, b, c, x, y and z ions. Internal fragments (green box), the result of the simultaneous breakage of two peptide bonds, as well as immonium ions (blue box) are also commonly observed in MS/MS spectra. (Figure adapted from Khatun *et al.*[108])

Figure 1.5

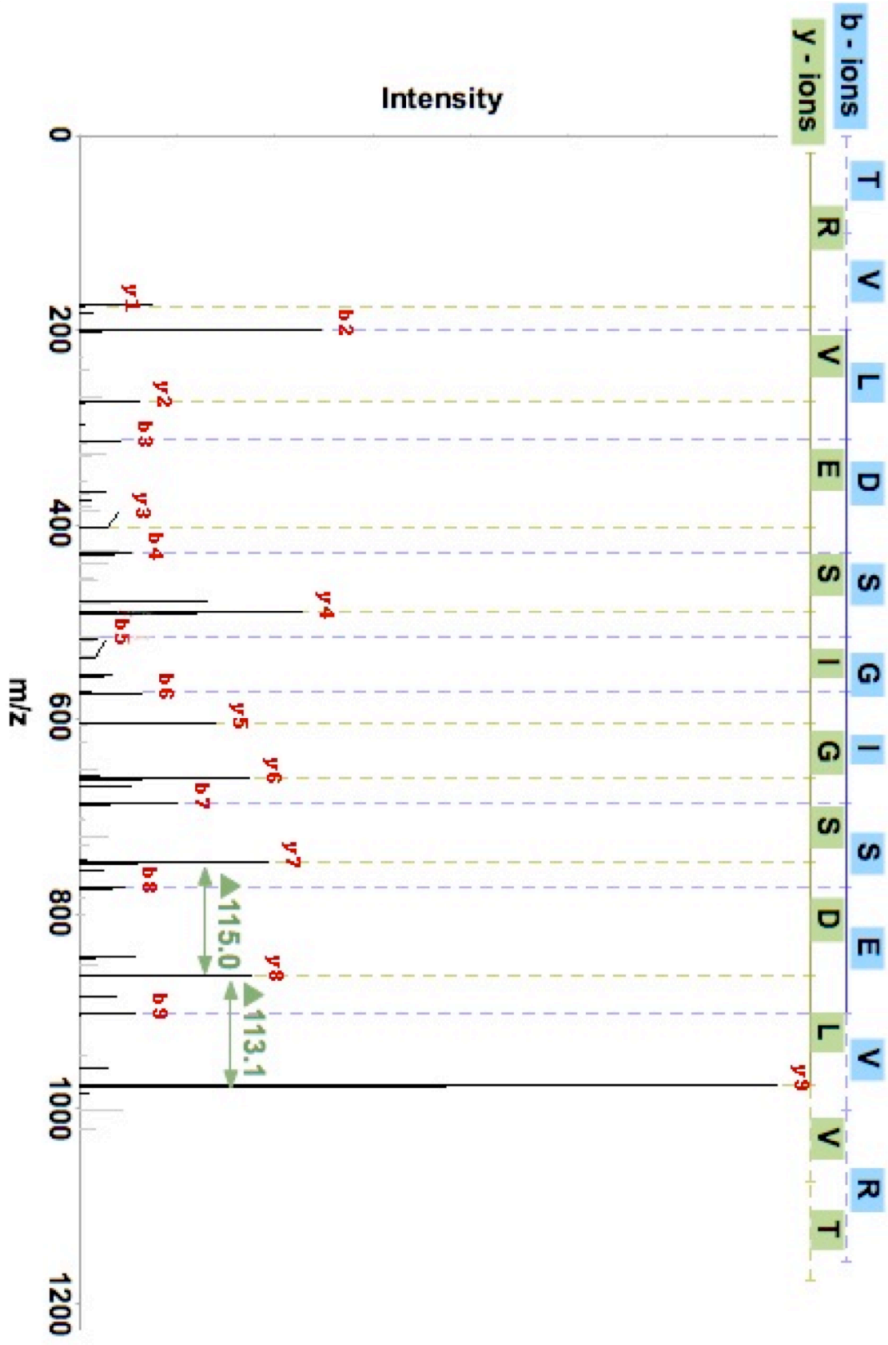
Intensity
m/z
0
200
400
600
800
1000
1200
b - ions
y - ions
T R V V L E D S S I G I G S S D E L V V R T
y1
b2
y2
b3
y3
b4
y4
b5
b6
y5
y6
b7
y7
b8
115.0
y8
b9
113.1
y9

Figure 1.5. Peptide MS/MS fragmentation spectrum. Tandem MS (MS/MS) spectrum of peptide TVLDSGISEVR illustrating amino acid ladder sequence derived from the b-ions (blue) and y-ions (green) produced. The amino acid sequence is derived from the mass differences between adjacent b- and y-ion series peaks - illustrated for the y7 - y8 and y8 - y9 ion peaks. Mass differences of 115.0 Da (y7 - y8) and 113.1 Da (y8 - y9) identified the amino acids D and L respectively. Additional peaks due to other fragmentation ion types (unlabeled, see figure 1.4) may also be used to determine amino acid sequence. The relative intensities of the different ion types present in an MS/MS spectrum is dependent on the method of fragmentation employed.

Figure 1.6

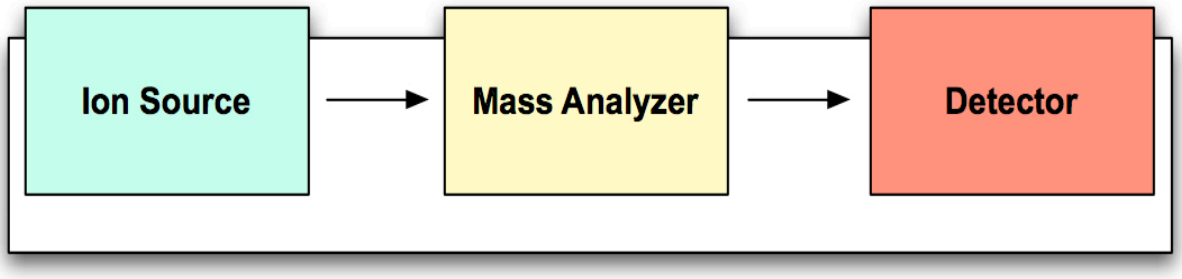

Figure 1.6. Three major components of conventional mass spectrometers. The ion source is responsible for aerosolization and ionization of the analytes. The mass analyzer is responsible for separating analytes and is ultimately dependent on the mass-to-charge ratios of the various analytes present. The detector is responsible for measuring the value of a quantifiable quality of the charged analytes thus facilitating determination of an analytes abundance.

Figure 1.7

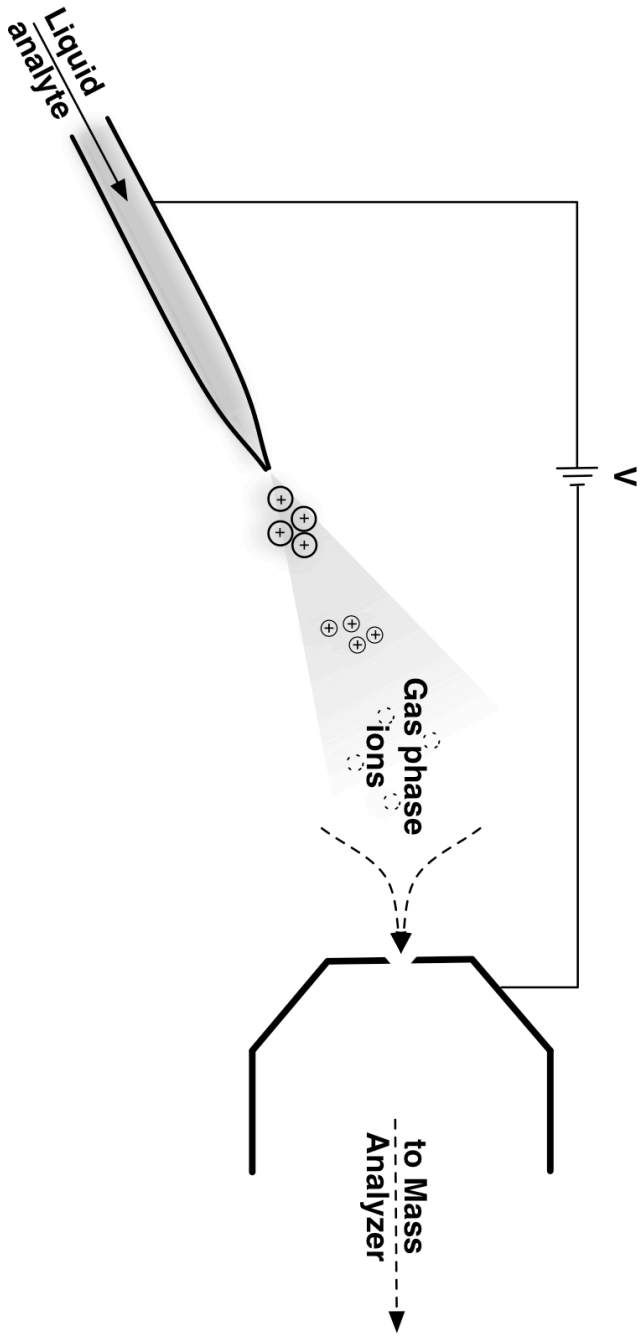

Figure 1.7 Electrospray ionization. Solubilized protein analytes are dispersed into charged, aerosolized droplets by the application of a high electric potential between the liquid in a thin metal capillary and a counter electrode. The liquid in which the proteins are solublized is rapidly evaporated in a process known as desolvation, leaving behind the charged, gas-phase molecule for MS analysis.

Figure 1.8

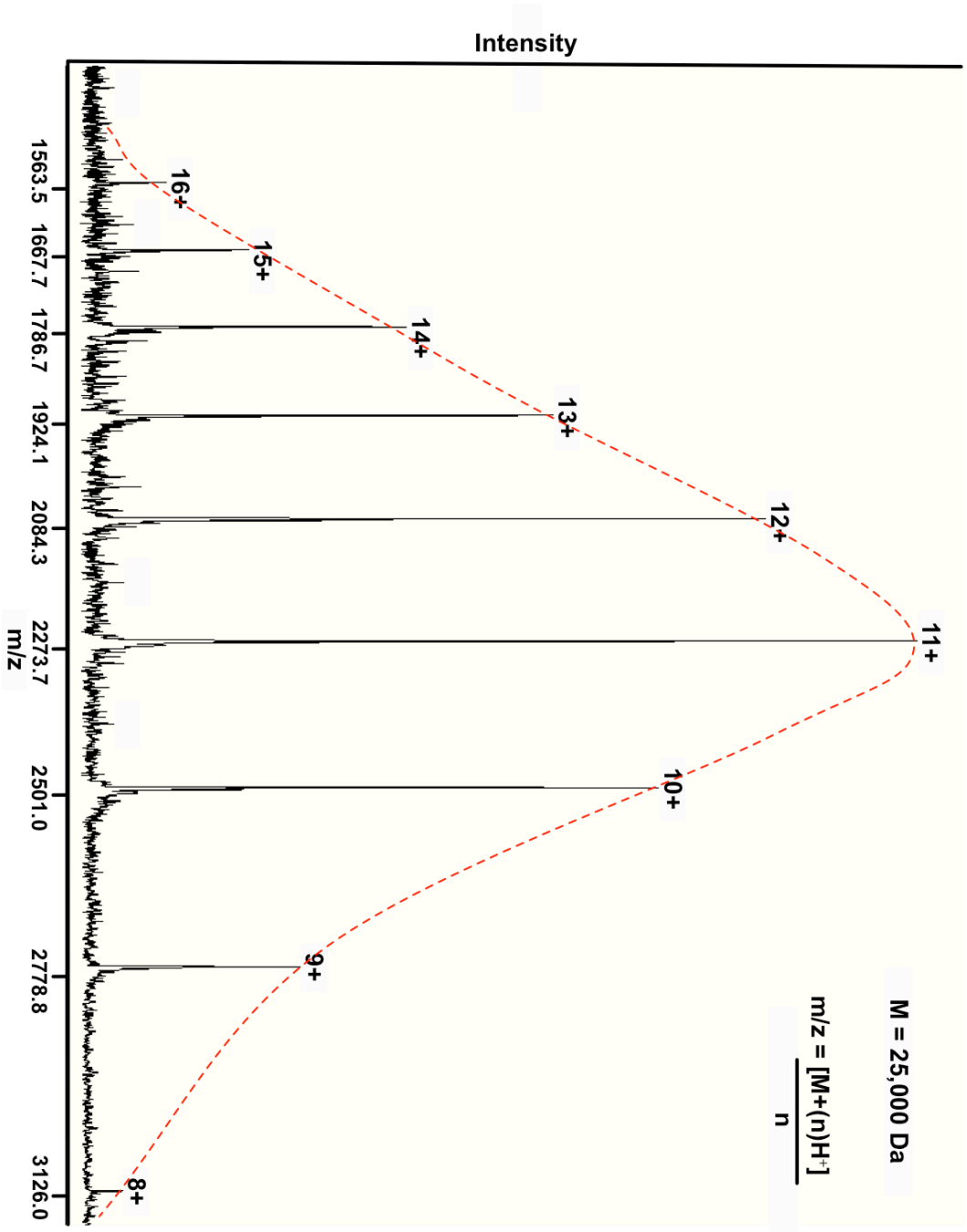
Intensity
16+
15+
14+
13+
12+
11+
10+
9+
8+
1563.5
1667.7
1786.7
1924.1
2084.3
2273.7
2501.0
2778.8
3126.0
m/z
M = 25,000 Da
m/z = [M+(n)H+]
n

Figure 1.8. Characteristic charge state envelope produced by ESI. Unlike matrix-assisted laser desorption/ionization (MALDI) which typically produces singly charged ions, electrospray ionization (ESI) of proteins results in a distribution of multiply charged states which constitute the so-called *charge state envelope of* the protein (denoted by red line). An m/z spectrum for a protein of mass (M) = 25 kDa protein is illustrated wherein 'M' is the mass of the protein,'H+' represents the mass of a hydrogen ion (proton) and 'n' is the number of protons (charge) on the ion.

References

1. Reeder, J.C. and G.V. Brown, Antigenic variation and immune evasion in Plasmodium falciparum malaria. Immunol Cell Biol, 1996. **74**(6): p. 546-54.

2. Penn, C.W. and C.J. Luke, Bacterial flagellar diversity and significance in pathogenesis. FEMS Microbiol Lett, 1992. **79**(1-3): p. 331-6.

3. Canu, A., et al., Diversity of ribosomal mutations conferring resistance to macrolides, clindamycin, streptogramin, and telithromycin in Streptococcus pneumoniae. Antimicrob Agents Chemother, 2002. **46**(1): p. 125-31.

4. Campos, L.C., et al., Genetic diversity and antibiotic resistance of clinical and environmental Vibrio Cholerae suggests that many serogroups are reservoirs of resistance. Epidemiol Infect, 2004. **132**(5): p. 985-92.

5. Fegan, M., et al., Heterogeneity, persistence, and distribution of Pseudomonas aeruginosa genotypes in cystic fibrosis patients. J Clin Microbiol, 1991. **29**(10): p. 2151-7.

6. Gefen, O. and N.Q. Balaban, The importance of being persistent: heterogeneity of bacterial populations under antibiotic stress. FEMS Microbiol Rev, 2009. **33**(4): p. 704-17.

7. Dhar, N. and J.D. McKinney, Microbial phenotypic heterogeneity and antibiotic tolerance. Curr Opin Microbiol, 2007. **10**(1): p. 30-8.

8. Andersson, D.I., Persistence of antibiotic resistant bacteria. Curr Opin Microbiol, 2003. **6**(5): p. 452-6.

9. Graumann, P.L., Different genetic programmes within identical bacteria under identical conditions: the phenomenon of bistability greatly modifies our view on bacterial populations. Mol Microbiol, 2006. **61**(3): p. 560-3.

10. Veening, J.W., et al., Bet-hedging and epigenetic inheritance in bacterial cell development. Proc Natl Acad Sci U S A, 2008. **105**(11): p. 4393-8.

11. Dubnau, D. and R. Losick, Bistability in bacteria. Mol Microbiol, 2006. **61**(3): p. 564-72.

12. Kearns, D.B. and R. Losick, Cell population heterogeneity during growth of Bacillus subtilis. Genes Dev, 2005. **19**(24): p. 3083-94.
Microbiol, 2005. **13**(10): p. 459-62.

13. Avery, S.V., Cell individuality: the bistability of competence development. Trends Microbiol, 2005. **13**(10): p. 459-62.

14. Veening, J.W., L.W. Hamoen, and O.P. Kuipers, Phosphatases modulate the bistable sporulation gene expression pattern in Bacillus subtilis. Mol Microbiol, 2005. **56**(6): p. 1481-94.

15. Tyers, M. and M. Mann, From genomics to proteomics. Nature, 2003. **422**(6928): p. 193-7.

16. Gomase, V.S. and S. Tagore, Transcriptomics. Curr Drug Metab, 2008. **9**(3): p. 245-9.

17. Hegde, P.S., I.R. White, and C. Debouck, Interplay of transcriptomics and proteomics. Curr Opin Biotechnol, 2003. **14**(6): p. 647-51.

18. Wasinger, V.C., et al., Progress with gene-product mapping of the Mollicutes: Mycoplasma genitalium. Electrophoresis, 1995. **16**(7): p. 1090-4.

19. Pandey, A. and M. Mann, Proteomics to study genes and genomes. Nature, 2000. **405**(6788): p. 837-46.

20. Wu, L. and D.K. Han, Overcoming the dynamic range problem in mass spectrometry-based shotgun proteomics. Expert Rev Proteomics, 2006. **3**(6): p. 611-9.

21. Han, K.K. and A. Martinage, Post-translational chemical modification(s) of proteins. Int J Biochem, 1992. **24**(1): p. 19-28.

22. Lee, T.Y., et al., dbPTM: an information repository of protein post-translational modification. Nucleic Acids Res, 2006. **34**(Database issue): p. D622-7.

23. Ben-Bassat, A., et al., Processing of the initiation methionine from proteins: properties of the *Escherichia coli* methionine aminopeptidase and its gene structure. J Bacteriol, 1987. **169**(2): p. 751-7.

24. Chang, S.Y., E.C. McGary, and S. Chang, Methionine aminopeptidase gene of *Escherichia coli* is essential for cell growth. J Bacteriol, 1989. **171**(7): p. 4071-2.

25. Zhang, J., et al., Lysine acetylation is a highly abundant and evolutionarily conserved modification in *Escherichia coli*. Mol Cell Proteomics, 2009. **8**(2): p. 215-25.

26. Driessen, H.P., et al., The mechanism of N-terminal acetylation of proteins. CRC Crit Rev Biochem, 1985. **18**(4): p. 281-325.

27. Macek, B., et al., Phosphoproteome analysis of *E. coli* reveals evolutionary conservation of bacterial Ser/Thr/Tyr phosphorylation. Mol Cell Proteomics, 2008. **7**(2): p. 299-307.

28. Deutscher, J. and M.H. Saier, Jr., Ser/Thr/Tyr protein phosphorylation in bacteria - for long time neglected, now well established. J Mol Microbiol Biotechnol, 2005. **9**(3-4): p. 125-31.

29. Amaro, A.M. and C.A. Jerez, Methylation of ribosomal proteins in bacteria: evidence of conserved modification of the eubacterial 50S subunit. J Bacteriol, 1984. **158**(1): p. 84-93.

30. Paul, R., et al., Activation of the diguanylate cyclase PleD by phosphorylation-mediated dimerization. J Biol Chem, 2007. **282**(40): p. 29170-7.

31. Foussard, M., et al., The molecular puzzle of two-component signaling cascades. Microbes Infect, 2001. **3**(5): p. 417-24.

32. Lowry, D.F., et al., Signal transduction in chemotaxis. A propagating conformation change upon phosphorylation of CheY. J Biol Chem, 1994. **269**(42): p.

26358-62.

33. Bourret, R.B., et al., Protein phosphorylation in chemotaxis and two-component regulatory systems of bacteria. J Biol Chem, 1989. **264**(13): p. 7085-8.

34. Borkovich, K.A. and M.I. Simon, The dynamics of protein phosphorylation in bacterial chemotaxis. Cell, 1990. **63**(6): p. 1339-48.

35. Springer, M.S., M.F. Goy, and J. Adler, Protein methylation in behavioural control mechanisms and in signal transduction. Nature, 1979. **280**(5720): p. 279-84.

36. Yan, J., et al., In vivo acetylation of CheY, a response regulator in chemotaxis of *Escherichia coli*. J Mol Biol, 2008. **376**(5): p. 1260-71.

37. Polevoda, B. and F. Sherman, Methylation of proteins involved in translation. Mol Microbiol, 2007. **65**(3): p. 590-606.

38. Arnold, R.J. and J.P. Reilly, Analysis of methylation and acetylation in *E. coli* ribosomal proteins. Methods Mol Biol, 2002. **194**: p. 205-10.

39. Moore, P.B. and T.A. Steitz, The involvement of RNA in ribosome function. Nature, 2002. **418**(6894): p. 229-35.

40. Ramakrishnan, V., Ribosome structure and the mechanism of translation. Cell, 2002. **108**(4): p. 557-72.

41. Wittmann, H.G., Components of bacterial ribosomes. Annu Rev Biochem, 1982. **51**: p. 155-83.

42. Kaczanowska, M. and M. Ryden-Aulin, Ribosome biogenesis and the translation process in *Escherichia coli*. Microbiol Mol Biol Rev, 2007. **71**(3): p. 477-94.

43. Wilson, D.N. and K.H. Nierhaus, Ribosomal proteins in the spotlight. Crit Rev Biochem Mol Biol, 2005. **40**(5): p. 243-67.

44. Ogle, J.M. and V. Ramakrishnan, Structural insights into translational fidelity. Annu Rev Biochem, 2005. **74**: p. 129-77.

45. Clemons, W.M., Jr., et al., Structure of a bacterial 30S ribosomal subunit at 5.5 A resolution. Nature, 1999. **400**(6747): p. 833-40.

46. Selmer, M., et al., Structure of the 70S ribosome complexed with mRNA and tRNA. Science, 2006. **313**(5795): p. 1935-42.

47. Gao, Y.G., et al., The Structure of the Ribosome with Elongation Factor G Trapped in the Posttranslocational State. Science, 2009.

48. Schmeing, T.M. and V. Ramakrishnan, What recent ribosome structures have revealed about the mechanism of translation. Nature, 2009. **461**(7268): p. 1234-42.

49. Schmeing, T.M., et al., The Crystal Structure of the Ribosome Bound to EF-Tu and Aminoacyl-tRNA. Science, 2009.

50. Kowalak, J.A. and K.A. Walsh, Beta-methylthio-aspartic acid: identification of a novel posttranslational modification in ribosomal protein S12 from *Escherichia coli*. Protein Sci, 1996. **5**(8): p. 1625-32.

51. Vanet, A., et al., Ribosomal protein methylation in *Escherichia coli*: the gene prmA, encoding the ribosomal protein L11 methyltransferase, is dispensable. Mol Microbiol, 1994. **14**(5): p. 947-58.

52. Lhoest, J. and C. Colson, Cold-sensitive ribosome assembly in an *Escherichia coli* mutant lacking a single methyl group in ribosomal protein L3. Eur J Biochem, 1981. **121**(1): p. 33-7.

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

54. Isono, K. and S. Isono, Ribosomal protein modification in *Escherichia coli*. II. Studies of a mutant lacking the N-terminal acetylation of protein S18. Mol Gen Genet,

1980. **177**(4): p. 645-51.

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

56. Wool, I.G., Extraribosomal functions of ribosomal proteins. Trends Biochem Sci, 1996. **21**(5): p. 164-5.

57. Warner, J.R. and K.B. McIntosh, How common are extraribosomal functions of ribosomal proteins? Mol Cell, 2009. **34**(1): p. 3-11.

58. Tchorzewski, M., B. Boldyreff, and N. Grankowski, Extraribosomal function of the acidic ribosomal P1-protein YP1alpha from Saccharomyces cerevisiae. Acta Biochim Pol, 1999. **46**(4): p. 901-10.

59. Mauro, V.P. and G.M. Edelman, The ribosome filter hypothesis. Proc Natl Acad Sci U S A, 2002. **99**(19): p. 12031-6.

60. Mauro, V.P. and G.M. Edelman, The ribosome filter redux. Cell Cycle, 2007. **6**(18): p. 2246-51.

61. Cravatt, B.F., G.M. Simon, and J.R. Yates, 3rd, The biological impact of mass-spectrometry-based proteomics. Nature, 2007. **450**(7172): p. 991-1000.

62. Feng, X., et al., Mass spectrometry in systems biology: an overview. Mass Spectrom Rev, 2008. **27**(6): p. 635-60.

63. Aebersold, R. and M. Mann, Mass spectrometry-based proteomics. Nature, 2003. **422**(6928): p. 198-207.

64. Kelleher, N.L., Top-down proteomics. Anal Chem, 2004. **76**(11): p. 197A-203A.

65. Siuti, N. and N.L. Kelleher, Decoding protein modifications using top-down mass spectrometry. Nat Methods, 2007. **4**(10): p. 817-21.

66. Delahunty, C.M. and J.R. Yates, 3rd, MudPIT: multidimensional protein identification technology. Biotechniques, 2007. **43**(5): p. 563, 565, 567 passim.

67. Pappin, D.J., P. Hojrup, and A.J. Bleasby, Rapid identification of proteins by peptide-mass fingerprinting. Curr Biol, 1993. **3**(6): p. 327-32.

68. Mortz, E., et al., Identification of proteins in polyacrylamide gels by mass spectrometric peptide mapping combined with database search. Biol Mass Spectrom, 1994. **23**(5): p. 249-61.

69. Pappin, D.J., Peptide mass fingerprinting using MALDI-TOF mass spectrometry. Methods Mol Biol, 1997. **64**: p. 165-73.

70. Pappin, D.J., Peptide mass fingerprinting using MALDI-TOF mass spectrometry. Methods Mol Biol, 2003. **211**: p. 211-9.

71. Sommerer, N., D. Centeno, and M. Rossignol, Peptide mass fingerprinting: identification of proteins by MALDI-TOF. Methods Mol Biol, 2007. **355**: p. 219-34.

72. Cottrell, J.S., Protein identification by peptide mass fingerprinting. Pept Res, 1994. **7**(3): p. 115-24.

73. Hjerno, K., Protein identification by peptide mass fingerprinting. Methods Mol Biol, 2007. **367**: p. 61-75.

74. Henzel, W.J., C. Watanabe, and J.T. Stults, Protein identification: the origins of peptide mass fingerprinting. J Am Soc Mass Spectrom, 2003. **14**(9): p. 931-42.

75. Giddings, M.C., et al., Genome-based peptide fingerprint scanning. Proc Natl Acad Sci U S A, 2003. **100**(1): p. 20-5.

76. McLafferty, F.W., Tandem mass spectrometry. Science, 1981. **214**(4518): p. 280-7.

77. Coon, J.J., et al., Tandem mass spectrometry for peptide and protein sequence

analysis. Biotechniques, 2005. **38**(4): p. 519, 521, 523.

78. Wells, J.M. and S.A. McLuckey, Collision-induced dissociation (CID) of peptides and proteins. Methods Enzymol, 2005. **402**: p. 148-85.

79. Wiesner, J., T. Premsler, and A. Sickmann, Application of electron transfer dissociation (ETD) for the analysis of posttranslational modifications. Proteomics, 2008. **8**(21): p. 4466-83.

80. Little, D.P., et al., Infrared multiphoton dissociation of large multiply charged ions for biomolecule sequencing. Anal Chem, 1994. **66**(18): p. 2809-15.

81. Zubarev, R.A., et al., Electron capture dissociation for structural characterization of multiply charged protein cations. Anal Chem, 2000. **72**(3): p. 563-73.

82. Molina, H., et al., Comprehensive comparison of collision induced dissociation and electron transfer dissociation. Anal Chem, 2008. **80**(13): p. 4825-35.

83. Bertsch, A., et al., De novo peptide sequencing by tandem MS using complementary CID and electron transfer dissociation. Electrophoresis, 2009. **30**(21): p. 3736-47.

84. Tabb, D.L., et al., Statistical characterization of ion trap tandem mass spectra from doubly charged tryptic peptides. Anal Chem, 2003. **75**(5): p. 1155-63.

85. Zubarev, R.A., A.R. Zubarev, and M.M. Savitski, Electron capture/transfer versus collisionally activated/induced dissociations: solo or duet? J Am Soc Mass Spectrom, 2008. **19**(6): p. 753-61.

86. Datta, R. and M. Bern, Spectrum fusion: using multiple mass spectra for de novo peptide sequencing. J Comput Biol, 2009. **16**(8): p. 1169-82.

87. Han, X., et al., Extending top-down mass spectrometry to proteins with masses greater than 200 kilodaltons. Science, 2006. **314**(5796): p. 109-12.

88. Loo, J.A., C.G. Edmonds, and R.D. Smith, Primary sequence information from intact proteins by electrospray ionization tandem mass spectrometry. Science, 1990. **248**(4952): p. 201-4.

89. Reid, G.E. and S.A. McLuckey, 'Top down' protein characterization via tandem mass spectrometry. J Mass Spectrom, 2002. **37**(7): p. 663-75.

90. Andersen, J.S., B. Svensson, and P. Roepstorff, Electrospray ionization and matrix assisted laser desorption/ionization mass spectrometry: powerful analytical tools in recombinant protein chemistry. Nat Biotechnol, 1996. **14**(4): p. 449-57.

91. Bakhtiar, R. and R.W. Nelson, Electrospray ionization and matrix-assisted laser desorption ionization mass spectrometry. Emerging technologies in biomedical sciences. Biochem Pharmacol, 2000. **59**(8): p. 891-905.

92. Fenn, J.B., et al., Electrospray ionization for mass spectrometry of large biomolecules. Science, 1989. **246**(4926): p. 64-71.

93. Hillenkamp, F., et al., Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers. Anal Chem, 1991. **63**(24): p. 1193A-1203A.

94. Vestal, M.L., Modern MALDI time-of-flight mass spectrometry. J Mass Spectrom, 2009. **44**(3): p. 303-17.

95. Cotter, R.J., Time-of-flight mass spectrometry: an increasing role in the life sciences. Biomed Environ Mass Spectrom, 1989. **18**(8): p. 513-32.

96. Schwartz, J.C. and I. Jardine, Quadrupole ion trap mass spectrometry. Methods Enzymol, 1996. **270**: p. 552-86.

97. Chernushevich, I.V., A.V. Loboda, and B.A. Thomson, An introduction to quadrupole-time-of-flight mass spectrometry. J Mass Spectrom, 2001. **36**(8): p. 849-65.

98. Marshall, A.G., C.L. Hendrickson, and G.S. Jackson, Fourier transform ion cyclotron resonance mass spectrometry: a primer. Mass Spectrom Rev, 1998. **17**(1): p. 1-35.

99. Bergquist, J., FTICR mass spectrometry in proteomics. Curr Opin Mol Ther, 2003. **5**(3): p. 310-4.

100. Heeren, R.M., et al., A mini-review of mass spectrometry using high-performance FTICR-MS methods. Anal Bioanal Chem, 2004. **378**(4): p. 1048-58.

101. O'Farrell, P.H., High resolution two-dimensional electrophoresis of proteins. J Biol Chem, 1975. **250**(10): p. 4007-21.

102. Friedman, D.B., S. Hoving, and R. Westermeier, Isoelectric focusing and two-dimensional gel electrophoresis. Methods Enzymol, 2009. **463**: p. 515-40.

103. Bjellqvist, B., et al., Isoelectric focusing in immobilized pH gradients: principle, methodology and some applications. J Biochem Biophys Methods, 1982. **6**(4): p. 317-39.

104. Yan, J.X., et al., Fluorescence two-dimensional difference gel electrophoresis and mass spectrometry based proteomic analysis of *Escherichia coli*. Proteomics, 2002. **2**(12): p. 1682-98.

105. Regnier, F.E., High-performance liquid chromatography of biopolymers. Science, 1983. **222**(4621): p. 245-52.

106. Boysen, R.I. and M.T. Hearn, HPLC of peptides and proteins: standard operating conditions. Curr Protoc Mol Biol, 2001. **Chapter 10**: p. Unit 10 13.

107. Giddings, J.C., Two-dimensional separations: concept and promise. Anal Chem, 1984. **56**(12): p. 1258A-1260A, 1262A, 1264A passim.

108. Khatun, J.; Ramkissoon, K.; Giddings, M. C., Fragmentation characteristics of collision-induced dissociation in MALDI TOF/TOF mass spectrometry. Anal Chem **2007,** 79, (8), 3032-40.

Chapter 2

Integrated Top-down/Bottom-up Characterization of Protein Isoform Heterogeneity in *Escherichia coli* Ribosomal Protein

Abstract

Due to emerging evidence of a regulatory role for the eukaryotic ribosome, we examined the biochemical heterogeneity of *Escherichia coli* ribosomal proteins to ascertain the potential for variable post-translational modification in the modulation of prokaryotic ribosomal activity. Our combined top-down and bottom-up mass spectrometry analyses of a wild-type *E. coli* strain, and an antibiotic resistant and fitness-compensated derivative yielded insight into the range of modifications and isoforms permissible under typical laboratory growth conditions. Intact-mass measurements were used to predict putative modifications with peptide structural analysis by tandem mass spectrometry used to identify proteins and investigate specific modification scenarios. By integrating multiple mass spectrometric characterizations of both intact proteins and their proteolytic peptides we pinpointed numerous cases of N- terminal truncation, methylation, acetylation, one β-methylthiolation and an amino acid substitution. These analyses yielded evidence of heterogeneity in the chemical modifications present on a number of ribosomal proteins including L7/L12, L16, and L33. Our results indicate that there is potential for the ribosome to exist in over $3x10^4$ 'states' due to differential protein

modification, and provides evidence of possible complexity in the modification systems targeting ribosomes, with possible implications in modulating ribosomal assembly, translation and extraribosomal ribosomal protein activity.

2.1 Introduction

Phenotypic diversity within bacterial populations is a key contributor to bacteria's capacity to adapt to and survive variable and often adverse physiological conditions. While genetic diversity has long been thought of as the main contributor to bacterial phenotypic diversity, a range of mechanisms such as bistable gene expression[1] and variable protein post-translational modification (PTM) have indicated that phenotype can be switched by mechanisms other than genetic mutation. The output of such simple systems can in turn determine much larger aspects of cellular phenotype. From a bacterial perspective, it makes sense to employ mechanisms such as PTMs that can quickly switch a cell into or out of particular phenotypic states or modulate protein function, often without the commitments associated with genetic changes, and especially in response to conditions that are apt to change far more frequently than favorable mutations, or combinations thereof, arise in the genome.

Examples of distinct bistable population states include the switch to/from genetic competence in *B. subtilis* and activation of the lactose utilization pathway in *E. coli*[2, 3]. It is also becoming clear that changes in post-translational modifications (PTMs) on proteins can drive rapid changes in cellular state or phenotype. A well elucidated case is bacterial chemotaxis, where both protein phosphorylation and protein methylation play critical roles in signal transduction and receptor sensitivity to environmental ligands[4, 5]. In this work, we studied the diversity of protein PTMs in bacterial ribosomal proteins, since both prior work and our own have uncovered a significant number of PTMs whose functions largely remain unclear and whose importance may yet prove to be under-

appreciated in modulating ribosomal function, or perhaps in extraribosomal protein functions.

The bacterial ribosome is macromolecular machine at the heart of the cell: it is responsible for producing all proteins, and without its proper functioning, a cell quickly dies. In *E. coli*, the ribosome consists of 55 proteins and 3 RNA molecules in a bipartite structure which assembles on an mRNA molecule to begin translation of the encoded message into a protein. The ribosomes of both prokaryotes and eukaryotes are known to be extensively post-translationally modified, with both ribosomal RNA (rRNA) and ribosomal proteins being targets of modification enzymes, particularly methyltransferases[6]. But many open questions remain about the ribosome, such as: how does it respond to differing protein translation needs as the cellular environment changes, and what are the roles of the various chemical modifications present on ribosomal RNAs and proteins? Numerous rRNA post-transcriptional modifications have been shown to strongly impact the function and fidelity of the ribosome, some having been pinpointed as having roles in antibiotic susceptibility[7]. Less is known about the functions of ribosomal protein modification, however experimental evidence and high conservation point to important functional roles. A few cases have been well-studied, such as the methylation of *E. coli* ribosomal protein L3, suggested to play a role in ribosomal assembly[8] and the variable acetylation of the L7/12 N-terminus, which stabilizes its interaction with ribosomal protein L10 in a growth-stage dependent manner[9]. In spite of mounting evidence of important cellular functions and growth condition-dependent functions, a systematic study of the extent, variability and regulation the modifications has not previously been undertaken.

In *E. coli* and other prokaryotes mRNA synthesis and translation are usually synchronous. This synchronous coupling of transcription and translation has lent itself to the view of the ribosome as a machine with no regulatory propensity. However, multiple lines of evidence have been accumulating which indicate that the ribosome may play a direct regulatory role on protein expression in the cell. Ribosomal proteins have also gained renewed attention for their extraribosomal functions. Extaribosomal functions refer to cellular functions performed by ribosomal proteins beyond those executed as a component of the assembled holoribosome[10, 11].

The "ribosome filter" hypothesis put forward by Mauro and Edelman, proposes that ribosomes can regulate translation of specific mRNAs through interactions with rRNA and/or ribosomal proteins, resulting in either the enhancement or repression of translation initiation[12, 13]. A number of ribosomal proteins have already been shown to influence the translation of their own, as well as other translation component-encoding mRNAs through direct interactions[14]. Whether such influence is confined to the control of ribosome biogenesis or is a more general phenomenon in bacteria remains to be conclusively determined. While the filter hypothesis has slowly gained traction, a major challenge to its further elucidation has been determining the specific mechanisms by which the filter might operate. Because PTMs are used in many other aspects of cellular life to provide readily reversible control over protein or RNA binding and function, they are a highly likely suspect for involvement in/control of any ribosome-as-regulator functions in bacteria. By investigating whether, as well as how much, heterogeneity exists within ribosomal proteins, we sought to gain insight as to whether there are sufficient "levers and dials" that would allow the ribosome to be fine-tuned in the manner

required by the ribosome filter hypothesis. We applied a combination of top-down and bottom-up proteomic technologies to characterize PTM heterogeneity on the proteins in the ribosomes of the *E. coli* K-12 bacterium.

Specifically, we examined protein isoforms present in ribosomal proteins isolated from wild type (WT) *E. coli* K-12 strain MG1655 and two derivative strains, using the combination of mass spectrometry approaches summarized in figure 2.1. One strain was spontaneous mutant (SmR) exhibiting resistance to the ribosome-targeting antibiotic streptomycin, and the second, a further derivative of the resistant strain (SmRC), generated by serial passage, which exhibited a phenotype of improved growth compared to its parent under stress conditions such as late stationary phase. We produced extensive intact-mass and bottom-up peptide data, in repeated biological replicates on multiple instruments, that were combined to probe the PTM heterogeneity present. We focused on low- peak intensity PTM variants that were repeatably observed over multiple experiments.

The intact mass was useful to predict and constrain the space of possible modifications present. However, by itself, the top-down analysis had several limitations: even with highly accurate mass measurements, distinct combinations of possible PTMs were sometimes isobaric in mass. For example, considering two common types of ribosomal protein PTM: the presence of 3 methylations (42.08 Da) ~ = 1 acetylation (42.04 Da). Additionally PTMs could not be localized to particular regions or residues on a protein, and some proteins did not behave well enough during separation and mass spectrometry to be observed in the mass spectrometers.

Because of such limitations, we applied complimentary "bottom-up" or "shotgun"

approaches in parallel. Peptides, generated through proteolytic digestion of intact proteins, were analyzed by tandem mass spectrometry (MS/MS), producing spectra that were used to identify the sequence of the peptides and the location of any PTMs. These peptide-based approaches have greater detection sensitivity for complex protein mixtures than top-down methods. In a bottom-up approach, all it takes is a few well-behaved peptides from a protein to definitively identify its presence in the mix. However, the bottom-up methods had their own limitations particularly towards PTM and isoform analysis. Since not all peptides could be measured in any single experiment, only limited coverage across a protein was achieved, meaning that potential modifications could be missed. Also, when there were distinct isoforms present, we would observe particular peptides both with and without a given modification. However, knowing this could not indicate which combinations of PTMs were viable. For example, if we observed 3 peptides with distinct methylation sites from a single protein, this did not give any indication whether it was biologically viable for any given protein to have all 3 methylation sites occupied at once, or whether the sites were mutually exclusive with only one occupied in any given protein, or some case in between. The intact mass was vital to sorting this out, providing the big picture of how many methylations were present on any given protein isoform. To maximize peptide coverage of proteins, we also applied a directed strategy of first purifying and isolating intact proteins by liquid chromatography, then digesting the isolated proteins using trypsin and analyzing the resulting peptides by matrix assisted laser desorption ionization, tandem time-of-flight mass spectrometry (MALDI-TOF-TOF MS/MS) and by electrospray ionization quadrupole time of flight tandem mass spectrometry (ESI-Q-TOF MS/MS).

2.2 Materials and Methods

Materials. Salts, buffers, glacial acetic acid and sucrose were obtained from Fisher Scientific. RNase-free DNase 1, formic acid and dithiothreitol (DTT) were obtained from Sigma-Aldrich (St Louis, MO). RNase Away (Molecular BioProducts, San Diego, CA), was used to treat all glassware and RNase/Protease-free water was used to make all buffers to minimize RNase activity during the ribosomal extraction procedure.

Cell culture and ribosome isolation. Each of the three *E. coli* derivatives was grown aerobically in LB broth at 37^{o}C at ~250 rotations/min prior to harvesting. We purified ribosomes based on the previously described method of Strader *et al*.[15] Briefly, bacterial cell pellets were resuspended in lysis buffer (100mM NH4Cl/ 50mM MgCH3COO/ 20mM Tris-HCl pH 7.5/ 1mM DTT/ 0.5mM EDTA; 1:2 w:v). Cells were lysed by sonication and treated with RNase-free DNase I (20 min at 4°C) prior to clarification by centrifugation. Ribosomes and polysomes were isolated via centrifugation through a 1.1M sucrose-enriched lysis buffer cushion (1:1 v:v, 100,000 x g, 16 hrs., 4°C). The resultant pellet was resuspended in ~1ml ribosome buffer (50mM NH4Cl/ 6mM MgCH3COO/ 20mM Tris-HCl pH 7.5/ 1mM DTT/ 0.5mM EDTA) and fractionated through a 7-30% sucrose/ribosome buffer gradient (17,000 g, 4 hrs., 4^{o}C). Holoribosome-containing fractions were identified via maximal abs 280nm measurements, pooled and centrifuged to pellet holoribosomes (100,000 x g, 16 hrs., 4°C). The pellet was resuspended in 1 ml of ribosome buffer and protein separated from RNA by standard acid extraction and centrifugation. The clarified ribosome supernatant was dialyzed using 3500MW cutoff Slide-A-Lyzer cassettes (Pierce, Rockford, IL) against 8 liters ultrapure water for a total of 36 hours at 22°C. The clarified ribosome extract was concentrated by

lyophilization. Protein concentration was determined using BCA reagent (Pierce) and aliquots stored at -80°C until further analysis.

On-Line Intact Protein MS Analysis. Intact protein MS was achieved reversed-phase liquid-chromatography (RP-LC) coupled to electospray ionization (ESI) on two different instruments: a Bruker BIOTOF II ESI-TOF and a 9.4 Tesla LC-FTICR. Capillary HPLC-FT-ICR-MS was accomplished with a Dionex UltiMate HPLC interfaced directly to an Analytica electrospray source (Analytica of Branford, CT) on an IonSpec (Lake Forest, CA) 9.4-Tesla (Cryomagnetics Inc., Oak Ridge, TN) HiRes electrospray Fourier transform ion cyclotron resonance mass spectrometer. A C4 reversed-phase column (VYDAC 214MS5.325 C4 column 300μm id x 250mm, 300Å with 5μm particles, Grace-Vydac, Hesperia, CA) and a 75 min. linear gradient run from 100% solvent A (95% water, 5% acetonitrile, 0.1% formic acid, 50 mM hexafluorisopropanol) to 100% solvent B (95% acetonitrile, 5% water, 0.1% formic acid, 50 mM hexafluorisopropanol) were employed for all separations. Because the mass resolution was at least 50,000 for the intact protein measurements, the molecular masses of these proteins could be measured with isotopic resolution. As previously described by Strader *et al.*[15], the measured most abundant isotopic mass (MAIM) of each molecular ion region was used as an approximation of the protein's isotopically averaged molecular mass for further analysis as described below. The isotopically resolved molecular mass region of the suspected protein was calculated based on its sequence and compared to the measured data from the FTICR-MS experiment to provide a ppm accuracy and putative identification. In a complementary analysis, proteins were fractionated by RP-HPLC on a Supelco C5 column (2.1 x 100mm) using a gradient of 10% to 70% acetonitrile/0.1% formic acid at

0.1ml/min over 120 minutes in-line with a Bruker BioTOF II ESI-TOF MS (Bruker Daltonics, Billerica, MA). The eluent flow was split between the mass spectrometer for intact analysis and a 96-well plate fraction collector for subsequent offline peptide analysis. The resulting LC-MS data were analyzed using Data Analysis vs 3.2 (Bruker Daltonics). Mass spectra were averaged over distinct time ranges determined from the total ion chromatogram (TIC) intensity profile. Typical processing consisted of single cycle smoothing (Gaussian, Smoothing width - 2 and 0.15 respectively). Spectral deconvolution for intact mass assignments was carried out using the Maximum Entropy function (Instrument Resolving Power = 10000, Data point spacing = 0.1 m/z, Resolution = normal). Where appropriate, spectra were manually corroborated using the Charge State Ruler tool against the smoothed mass spectra which allowed visualization of which m/z peaks were considered by the deconvolution function in calculating an intact mass. Intact protein analysis on the BioTOF II, capable of 20,000 resolving power, yielded typical mass accuracies under 50 for the primary isoform detected for each ribosomal protein.

On-Line Peptide MS/MS Analyses. Ribosomal proteins were digested with sequencing grade trypsin (1:20 wt/wt, 37°C, 12 hrs.), desalted with an Omics 100 µl solid phase extraction pipette tip (Phenomenex, Torrance, CA) and stored at -80°C until LC-MS/MS analysis. One-dimensional (1D) LC-MS/MS experiments were performed with a Famos/ Switchos/Ultimate HPLC System (Dionex, Sunnyvale, CA) coupled to an LCQ-DECA XP Plus quadrupole ion trap mass spectrometer (Thermo Finnigan, San Jose, CA) equipped with a nanospray source. A 160-minute linear gradient from 100% solvent A (95% H2O/5% ACN/0.5% formic acid) to 100% solvent B (30% H2O/ 70% ACN/0.5%

formic acid) was employed. The LCQ was operated in the data dependent mode with dynamic exclusion enabled (repeat count 2), where the four most abundant peaks in every MS scan were subjected to MS/MS analysis. Data dependent LC-MS/MS was performed over a parent m/z range of 400-2000. The SEQUEST algorithm was used to identify MS/MS spectra with their counterparts predicted from a protein sequence database. For all database searches, an *E. coli* proteome database was used, which contained 4,312 proteins and 36 common external contaminants. All resultant output files from SEQUEST were filtered by DTASelect at the 1-peptide, 2-peptides and 3-peptides level with the following parameters: SEQUEST, delCN of at least 0.08 and cross-correlation scores (Xcorrs) of at least 1.8 (+1), 2.5 (+2) and 3.5 (+3). The DTASelect analysis was performed with searches for β-methythiolation of aspartic acid, single acetylations, and mono-, di-, and tri-methylated lysines and arginines. Spectra generated by peptides with potentially interesting modifications were each reviewed manually.

Off-Line Peptide MS/MS Analyses. Individual RP-HPLC fractions collected in 96-well plates via a post-column flow-splitting during on-line LC-MS ribosomal protein separations were digested with sequencing grade trypsin (1:20 wt/wt, 37°C, 12 hrs.), desalted using 10 μl ZipTip pipette tips (Millipore Corporation, Billerica, MA) and combined with alpha-cyano-4-hydroxycinnamic acid matrix for MS/MS analysis on a MALDI TOF/TOF 4700 proteomics analyzer (Applied Biosystems). The instrument was operated in data dependent mode with the10 highest intensity peaks in an MS spectra automatically selected for MS/MS analysis. The resulting spectra were analyzed using ABI's Data Explorer software, and signal to noise thresholds of 8 and 5 were enforced for creation of the MS and MS/MS mass lists respectively. The MS mass lists were also

filtered for trypsin autodigest fragments prior to submission to peptide mass fingerprinting (PMF) and characterization software. GFS was queried against a database comprised of the encoding gene sequences of both ribosomal and potential contaminant proteins, identified in the shotgun analysis as being present in our samples. Gene sequences were obtained from NCBI.

The sequence tag feature of GFS was used to identify post-translationally modified peptides. MS/MS spectra of peptides whose experimental precursor mass differed from the sequence prediction, and which had a predicted sequence tag of at least three amino acids were further examined manually to confirm the presence of a modification and to determine its position. A complementary analysis of these data was performed using Mascot (Matrix Science) as part of an automated pipeline to query against all proteins from *E. coli* contained in the MSDB database as of March 2006. Proteins with total PMF scores (including ion scores) greater than 95 were reported as significant ($p<0.05$). MS/MS mass lists in the '.pkl' format were submitted to Mascot's MS/MS ion search tool. The masses were queried against all *E. coli* proteins in the MSDB database. Both GFS and Mascot searches considered enzymatic digestion by trypsin and allowed a maximum of 2 missed cleavages. The most commonly considered variable modifications included acetylation and mono-, di- and tri-methylation of lysine and arginine. A peptide tolerance of 100ppm and an MS/MS tolerance of 0.3Da were employed for the Mascot searches.

Intact MS Analysis for isoform and PTM Prediction. To identify proteins and putative isoforms/modifications, intact-protein mass measurements were analyzed using both PROCLAME and a custom database of all known *E. coli* proteins along with common

modifications. Considered in the database search were commonly-known ribosomal modifications including N-terminal loss of methionine, methylation, acetylation, β-methythiolation, as well as potential byproducts of protein preparation and ionization such as oxidation of methionine or sodium adducts. PROCLAME was run in batch mode with a fuzzy cutoff score of 0.6 and a precision of 100ppm, analyzing average isotopic mass. In addition to the modifications considered by the database search, it considered possible N- or C- terminal truncations of arbitrary length. Previously observed ribosomal modifications were assigned higher frequency (FreqScore) and probability (P-score) scores, which in PROCLAME has the effect of focusing more of the search space on these modifications, while still allowing for novel modifications.

The PROCLAME predictions for each observed mass were manually examined in conjunction with the bottom-up data (described below) for those scenarios that had corroborating peptides and/or were logically compatible with prior PTM observations. Criteria for filtering included a high PROCLAME P-score, low error between the observed and calculated masses, and logical agreement with previously observed ribosomal modifications. Ribosomal protein sequences were obtained from the EcoCyc database (http://www.ecocyc.org). Spectra for which distinct masses representing putative alternative protein isoforms were identified were subjected to a second manual analysis to address any potential for false assignments by the deconvolution algorithm due to overlapping protein envelopes or peaks deemed to be due to contaminants. We searched both the deconvoluted and smoothed m/z spectra for closely spaced series of well-defined peaks that might represent isoform series, moving outwards in both increasing and decreasing mass from the most abundant peak. We focused on peaks at the

mass expected for the more common ribosomal PTMs (methylation, acetylation and N-terminal cleavage or lack thereof). We subsequently analyzed the bottom-up data for the presence of the predicted modification in detected peptides. In addition to repeated observations in the top-down data, we required that each isoform be present in at least three distinct charge states in the m/z spectrum.

Strain generation. The wild type, plasmid-free, streptomycin-sensitive strain of *Escherichia coli* MG1655 (a gift from Dr. Robert Bourret, Dept. of Microbiology and Immunology, UNC) was cultured on Luria-Bertani (LB) agar supplemented with streptomycin sulfate (50μg/ml, ~ 4 x MIC). A spontaneous mutant colony with resistance to streptomycin (SmR) was isolated and serially passaged in LB broth at 37oC on a 12 hr. cycle for five days (equaling approximately 256 generations) to select for highest fitness. One cfu used to generate a final stock population for the SmRC derivative. The original SmR derivative and the final passaged SmRC derivative were directly competed against the wild-type predecessor in minimal media into late stationary phase. At 24 and 48 hours, cell density of each competitor was determined by direct colony counting of diluted mixtures on LB agar plates. Differences in growth rates were assessed within a 95% confidence interval using five independent measures of competition in each case.

Gene Sequencing. Standard whole-cell PCR from single colonies was performed for *rpsL* gene sequencing to detect polymorphisms that would lead to amino acid substitutions located in the proteomic analyses. Primers were designed using the Primer3 software with default parameters (forward: CGTGTTTACGAAGCAAAAGC, reverse: GGCCTTACTTAACGGAGAACC). The thermocycler program consisted of an initial lysing/denaturing step (95^{o}C/5mins.), 30 cycles of denaturing (95^{o}C/45sec.), annealing

(55-60^{o}C gradient/5mins.) and extension (72^{o}C/45sec.); followed by a final 5min extension step (72^{o}C). PCR product size was assessed against a 100bp DNA ladder on a 2% agarose gel (1xTBE buffer at 75V) stained with ethidium bromide (0.5ug/ml, 30 mins.) and visualized on a Eagle Eye II system (Stratagene, La Jolla, CA). PCR products were purified from the gels using the Qiagen QIAquick PCR Purification kit (Qiagen, Valencia, CA) and submitted for sequencing to the UNC DNA Sequencing Facility. Contig assembly and sequence analysis was carried out using Sequencher (Gene Codes Corporation, Ann Arbor, MI).

Analysis of contaminant proteins. We used the bottom-up data to determine what contaminant proteins co-purified with the ribosomes. This was important to determine whether each assignment between a measured intact protein mass and a ribosomal protein entry from the database could be a false assignment due to a contaminant with nearly the same mass. We observed several proteins in the preparations from all three strains: bacterioferritin with 39-63% sequence coverage and 6-8 unique peptides (18,483.3 Da predicted mass according to GenBank sequence), GroEL with 23-42% sequence coverage and 9-16 unique peptides (55,123.8 Da predicted), and a Co-A linked acetaldehyde dehydrogenase with 35-39% sequence coverage and 25-27 unique peptides (96,066.7 Da predicted). In the peptide mass fingerprint data, we additionally identified a weak match to a putative glycosyltransferase (42,165.2 Da predicted, Q9S519_ECOLI). Bacterioferritin is the only one of these contaminants with a mass close to a ribosomal protein, L6, but the difference of ~420 Da was enough to distinguish the intact mass matches to L6 versus this contaminant.

To estimate the false-positive rate of our top-down database identifications two distracter database searches were performed by searching the experimental *E. coli* mass data against *Rhodopseudomonas palustris* and yeast *Saccharomyces cerevisiae* ribosomal databases. When using the *R. palustris* ribosomal database, five *E. coli* ribosomal measured masses were identified within 1.0 Da from the *R. palustris* database; these included L31, S17, S10, L36, and L28. Although, L31 was identified with an N-terminal methionine truncation for *E. coli*, that was not matched in *R. palustris* for this protein. For the searches against the yeast database, only three yeast proteins were identified within 1 Da using the measured *E. coli* masses, including the 60S L28, 60S L44, and 40S S21 proteins. The yeast 60S L44 protein has homology to the *E. coli* L12 protein that could provide a match within the yeast database.

Spectra Averaging. For some proteins where peptide MS/MS data was not available to support the multiple intact mass observations, we averaged and aligned multiple individual charge states from the raw mass spectra. The alignment of multiple charge states provided visual illustration of consistency in the spectral data, minimizing concerns of artifacts introduced by the deconvolution algorithm. Briefly, m/z ranges (~ ±500 Da) surrounding the predominant mass peak were exported from Data Analysis as mass/ intensity pair lists (absolute intensity threshold 5, S/N threshold 2) for three consecutive individual charge states in the smoothed raw spectra (e.g. charge states +6, +7, +8) . These were used to create an averaged spectrum using the most intense peaks as landmarks. SpecAlign vs 2.4 (PTCl, University of Oxford, Oxford, UK) was employed to generate the averaged spectra and to carry out an initial alignment using the Recursive alignment by fast Fourier Transform function. This averaged spectrum would best

represent the data by enhancing S/N ratios, through suppression of noise and enhancement of repeated signals across each charge state. To account for small differences in inter-peak distances due to PTMs, caused by the differences in charge states, we used the Mobility Shift function of our Shape Finder software[40] to generate the best fit of the three charge states. (N.B. This processing was carried out for data presentation purposes only.)

Combinatorics calculation. The possible number of states in which the ribosome could exist was calculated as the product of the contribution of each ribosomal protein to ribosomal heterogeneity. The contribution of a protein to heterogeneity was calculated using the combinatorial formula $C(n, r) = n!/(r!(n-r)!)$ where 'n' is the number of predicted isoforms of a protein (10 for L7/12, 5 for L11, 2 for L16, 3 for L33 and 5 for S6, 1 for all others), and 'r' is the number of occurrences of each protein in the ribosome (4 for L7/12, 1 for all others).

2.3 Results

A unique feature of this study was the high sampling rate achieved by analyzing multiple bacterial derivative samples on multiple mass spectrometers, comprising at least a six-fold coverage for the intact analyses. Intact protein MS analyses were carried out on a Bruker BioTOF II ESI-TOF MS and IonSpec 9.4-Tesla FTICR MS. Peptide MS and MS/MS analyses were undertaken using a Finnigan LCQ quadrupole ion trap mass spectrometer and an ABI 4700 MALDI TOF/TOF (MS/MS) (see Methods and Materials). Online reversed-phase liquid chromatography coupled to electrospray ionization mass spectrometry (RP-LC-ESI-MS) of intact ribosomal proteins combined with both online and offline RP-LC-MS analyses of trypsin-digested ribosomal protein

allowed identification of all 55 ribosomal proteins in wild-type MG1655, 54 in the streptomycin resistant derivative (SmR), and 52 in the compensated derivative (SmRC) (see Appendix). Our intact protein analyses yielded the identification of 52 of the 55 ribosomal proteins with S1, S22 and L34 unobserved by top-down methods. Each of these three proteins however were identified in our peptide MS and MS/MS analyses, resulting in 100% ribosomal protein identification by bottom-up methods.

The vast majority of ribosomal proteins were identified in more than one biological sample, with a few exceptions noted later. Protein identifications and PTM predictions were made through a combination of -

1. Comparisons of the experimentally observed intact mass measurements with a custom database of all known *E. coli* proteins, considering a number of common protein modifications.
2. PTM prediction analyses using our PROCLAME software[16] accomplished thorough comparisons of experimentally observed intact mass measurements with predicted masses generated via *in silico* translation of ribosomal protein-encoding genes, considering common ribosomal protein modifications as well as protein oxidation and sodium adducts.
3. Correlation of reversed-phase LC retention times in the top-down LC-MS experiments with the peptide mass fingerprint (PMF) identifications from the bottom-up MS and MS/MS analyses generated using our Genome-based peptide Fingerprint Scanning (GFS) software[17] and Mascot[18] (figure 2.2).

The data revealed a number of cases where multiple, closely spaced top-down masses were found in distinct biological replicate experiments, putatively corresponding

to multiple isoforms of ribosomal proteins. In most of these cases, we found that the isoform(s) observed in a single experimental analysis of a derivative were present in the other experiments as well. One isoform, corresponding to the mass most frequently observed in our data and in the literature, was usually of significant abundance, sometimes accompanied by one or more less intense peaks representing distinct isoforms. This characteristic isoform distribution is illustrated in figure 2.3 for ribosomal protein L33. Three isoforms of L33, determined to be due to variable methylation, were repeatedly observed across multiple biological samples with the primary monomethylated isoform dwarfing peaks due to unmethylated and dimethylated isoforms (figure 2.3).

Corroborating peptide evidence (peptide MS and/or MS/MS) was obtained in a number of these cases, which allowed confirmation of predictions and localization of the PTM(s). Cases where experimental masses were deemed to be solely due to the formation of adducts, and/or protein oxidation, were generally omitted from supplementary table 1. For reasons discussed later in the paper, one notable exception to this was ribosomal protein L16 which, exhibited a primary isoform predicted to be due to possible *in vivo* oxidation.

Since the goal of this study was to gauge the level of protein heterogeneity, the identification of low-abundance isoforms was critical. This presented the challenge that most of the peaks we were interested in had low signal to noise ratios (S/N) in both the raw and deconvoluted mass spectra. We applied several selection criteria to aid in the differentiation of low-abundance (low S/N) intact masses due real proteins versus those due to instrument/chemical noise or software deconvolution artifacts. First, we searched for an MS/MS spectrum in the bottom-up peptide data containing a confirmatory

modification. Failing that, we studied the MS masses of peptides in the bottom up data to find the predicted modifications in those.

In the absence of confirmatory peptide MS or MS/MS data supporting an isoform, we required observation of an intact mass and its associated PROCLAME PTM prediction in more than one biological sample. Mass shifts resulting in distinct isoforms that were consistent with previously identified modifications for the same protein, either in our work or in published literature, were considered as positive support for a prediction. For example, if we had already observed an intact mass showing a single methylation with a +14 Da mass shift, observance of a secondary peak in multiple biological replicates at ~ +28 Da was considered positive support for a multiply-methylated isoform. To address possible software deconvolution artifacts, we examined raw m/z spectra for the presence of multiple charge states in support of a deconvoluted mass. Overlaying multiple charge states (see figure 2.4) often demonstrated strong consistency, indicative of an underlying feature rather than noise.

2.3.1 Proteins Exhibiting Isoform Variability.

One contributor to the number of isoforms present was variable N-terminal protein processing. We observed 33 of the ribosomal proteins with N-terminal methionine truncation, resulting in an intact-protein mass shift of -131.04 Da, in accord with previous results[19]. Of these, 2 additionally exhibited intact masses corresponding to isoforms both with and without methionine loss: L11 and L7/L12. Cleavage of N-terminal methionine is strongly dependent on the length of the side chain of the amino acid in position 2, with those having a length less than 2.55 Å (glycine, alanine, proline, serine, and threonine) strongly favoring methionine cleavage, and those with a longer

side chains like lysine and arginine favoring methionine retention[20]. Both L7/12 (serine2) and L33 (alanine2) follow this rule, with N-terminal methionine loss represented in the most abundant isoform. Interestingly S18 also has an alanine in position two, predicting that the loss of N- terminal methionine should predominate, as was previously observed19. Our lone intact mass observation for S18 however found a predominant peak with intact mass of 8,987 Da in the WT strain, indicating the retention of the N-terminal methionine. While S18 went unobserved in the other top-down biological replicates, our observation indicates there could be a secondary mechanism acting to regulate N-terminal processing for S18. In general all other ribosomal proteins observed held true to this cleavage 'rule'. We also observed a number of intact protein masses for L7/12 that were predicted in PROCLAME to have the second N-terminal amino acid cleaved. These predictions were based on top-down mass only and lacked bottom-up support. The *E. coli* methionine aminopeptidase responsible for N-terminal methionine cleavage is highly specific and if additional cleavage is indeed occurring it is probably due to a second aminopeptidase[21]. PROCLAME only considers cleavage at one end of the polypeptide chain at a time and our analysis was weighted towards N-terminal cleavage as this is where the majority of biological cleavage activity occurs. It remains possible that some of these masses were due to C-terminally truncated polypeptide products. Ribosomal protein S1, discussed later, was not observed in our intact protein analyses however our on-line LC-MS/MS analyses identified multiple peptides that indicated variable N-terminal processing of S1 also occurs.

Our intact mass analyses using PROCLAME indicated that the largest contributor to ribosomal protein heterogeneity was variable methylation. We identified 3 proteins

(L7/12, L11 and L33) as having possible states due to differences in methylation. The primary isoform identified for L33 was ~ 6254 Da, corresponding to the loss of N-terminal methionine and a single methylation. We additionally identified an isoform with a mass of ~ 6268 Da, assigned in PROCLAME to be the result a second methylation event. This mass was consistently observed in the L33-containing fraction for all three strains analyzed. L11 has been previously observed undergo N-terminal methionine cleavage and three trimethylations at the newly formed N-terminus and two internal lysines[19, 22]. We observed a primary intact mass peak of ~14870 Da, corresponding to a loss of N-terminal methionine and the addition of 9 methylations. We observed additional mass peaks of that were assigned as corresponding to isoforms containing 2, 3, 7 and 12 methylations. Each was of significantly lower intensity compared to the primary isoform. With the exception of the 12-times methylated isoform prediction, the PROCLAME predictions of methylation were accompanied by other modifications such as oxidation. A mass of ~ 14911.5 Da observed in the WT and SmRC strains was predicted to be due to a loss of N-terminal methionine and the addition of 12 methylations. A precedent for 12 methylations was established in the L11 homolog in the hyperthermophile *Thermus thermophilus*24. The function of the extensive methylation of L11 remains unclear, with the methylating enzyme PrmA having been shown to be nonessential for culture growth in both *E. coli* and *T. thermophilus*[23, 24].

The protein L7/12 offered an interesting case, with the number of predicted methylations ranging from 0 to 5. PTM assignments were supported by repeated intact protein mass measurements as well as the identification of a number of peptides exhibiting either ~ +14 Da (mono-methylation) or ~ +28 Da (di-methylation) mass shifts

in addition to the unmodified sequence mass. Amongst these were three repeatedly observed masses at ~ 1044.6 Da, 1058.6 Da and 1072.6 Da, forming a mass series corresponding to sequential methylation (figure 2.4). MS/MS analysis identified that variable methylation of the tryptic peptide 75-GATGLGLKEAK-85 was occurring with un-, mono- and di-methylated isoforms identified. Lysine 82 was identified as the location of methylation in each case (figure 2.5). Our bottom-up analyses identified a number of additional peaks at m/z corresponding to mono-methylated tryptic L7/12 peptides. These had not been selected for fragmentation analysis in the data-dependent MS/MS runs due to their low intensities, so further analysis is underway to definitively identify the peptides and confirm their modification status. The top-down data nonetheless pointed to the presence of methylation sites other than those identified at Lys82.

Ribosomal protein L16 exhibited two major isoforms with intact masses of ~15311.5 Da and 15327.4 Da, predicted as being due to methylation plus a single and double oxidation respectively (see figure 2.4). The N-terminus of L16 has previously been reported to be methylated[25]. In their analysis of *E. coli* ribosomes Arnold *et al.*19 observed L16 at a mass of 15326.2 Da and partially attributed this mass to an unknown modification of Arg81, while discussing the possibility of protein hydroxylation. Our MS/MS peptide analyses identified oxidation at Arg81 and Pro69. Sometimes considered an experimental artifact in MS analysis, protein oxidation also commonly occurs *in vivo*, typically in response to oxidative stress, altering protein stability and function[26]. From the conservation of the observed intact mass between Arnold's and our work, and its localization to a proline residue, it is likely that the ~ +16 and +32 Da mass shifts

observed were due to *in vivo* enzymatic hydroxylation. Proline is the most frequently hydroxylated amino acid (from http://dbptm.mbc.nctu.edu.tw). This raises the possibility that some isoforms we excluded from this analysis as being due to experimentally-induced oxidation could instead be due to *in vivo* modification of other ribosomal proteins.

S1, the largest of the ribosomal proteins with predicted sequence mass of ~ 61 kDa, has proven difficult to isolate and analyze in both the previous top-down MS studies and our own. S1 was efficiently purified with the ribosome complex using our extraction methods, indicated by its identification in the bottom-up analyses where we achieved a total of over 67% protein sequence coverage. While a lack of intact mass measurement of ribosomal protein S1 precluded obtaining a big picture of its protein isoforms, the bottom-up analyses indicated that S1 is a source of protein heterogeneity in the *E. coli* ribosome. Two isoforms of the protein N-terminus peptides, one at m/z 1529.4 indicating cleavage of the N-terminal methionine (2-TESFAQLFEESLK-14) and one at m/z 1660.8 (1-MTESFAQLFEESLK-14) indicating its retention, were identified in the on-line peptide LC-MS/MS analysis. Both peptides had MS/MS ion coverage above 70%. N-terminal cleavage statistics indicate the methionine-cleaved isoform should predominate as proteins with threonine at amino acid position 2 are cleaved approximately 90% of the time[27]. Our data reflected this with the cleaved (1529.4 m/z) peptide observed 8 times as frequently as the uncleaved (1660.8 m/z) peptide. We also found a third peptide at 1571.2 m/z (78% MS/MS ion coverage) indicating putative acetylation of the cleaved N-terminus. Taken together, these data indicated the presence of three S1 isoforms due to variable N-terminal processing alone.

Ribosomal protein S6 is known to be post-translationally modified by the enzyme RimK, responsible for the addition of up to 4 C-terminal glutamic acid residues[28]. Intact S6 was observed once in our top-down analysis of the SmR derivative at a mass of ~15445.6 Da, corresponding to the addition of two glutamic acid residues.

Ribosomal protein S12 presented an instance of protein variation between the different strains. Through the combination of top down and bottom up analyses we were able to identify a single amino acid substitution K42T in the streptomycin resistant SmR and SmRC strains (figure 2.6A/B & table 2.1). The assignment was supported by two lines of evidence: a measured intact mass of within -1.9 ppm mass error of the calculated value for the substitution and MS/MS analyses of peptides in the bottom-up data indicating an amino acid substitution at position 42. Further confirmation came from a single nucleotide substitution detected by sequencing the encoding *rpsL* gene. Single amino acid substitutions at Lys 42 in S12 were previously associated with streptomycin resistance in a number of bacteria, with some imparting a fitness cost on the bacteria via a reduction in translational efficiency manifested as slower growth[29, 30]. We had observed both streptomycin resistance and slow growth phenotypes in the SmR and SmRC strains, likely associated with this mutation.

We also identified a ß-methylthiolation of aspartate 88 (D88) consistently in all three strains, a PTM originally discovered by Kowalak *et al.*[31]. Our observation of its conservation across the WT and streptomycin resistant strains is in accordance with those of Carr *et al.*[32] whose work indicated that the ß-methylthiolation of S12 is not affected by the acquisition of streptomycin resistance.

We detected only one isoform for each of the remaining ribosomal proteins in the

replicate intact analyses for the growth conditions examined. Where observed, the intact mass measurements were consistent across all three strains analyzed. For example, in ribosomal protein S5 we identified an N-terminal methionine loss and subsequent N-terminal acetylation in all three strains, confirmed in both intact MS and bottom up data, corroborating prior observations 19. Unlike the variably N-terminal acetylation of L7/12, no mass corresponding to an unacetylated S5 was observed, indicating a high efficiency of the cognate acetyltransferase, possibly through a tight coupling with translation. Similarly in S11, six distinct top-down measurements indicated N-terminal loss of methionine plus a single methylation, with the methylated peptide identified in the bottom up data for the SmR and SmRC strains.

Ribosomal proteins S1, S22 and L34 were the only proteins not identified in our intact mass analyses. The small ribosomal proteins S22 (~5 kDa predicted) and L34 (~5.3 kDa predicted) were identified by bottom-up analyses with ~ 71% and ~ 58% protein sequence coverage respectively. All matched peptides for S22 and L34 were void of PTMs.

2.4 Discussion

While generally considered to be highly conserved and static, our analysis indicates that under just a single growth condition, variability exists in at least three modification systems targeting ribosomal proteins – acetylation, methylation and N- terminal processing. While some of the isoforms predicted solely from intact mass observations in this study require further validation, our TD/BU MS analysis adds to a growing field of evidence in support of the ribosome being a complex cellular machine with many possible distinct states partially modulated by PTMs on ribosomal proteins.

Several groups have previously studied ribosomal proteins from various prokaryotes using either, or a combination of, top-down and bottom-up MS approaches[15, 19, 33]. While no single study reported the level of heterogeneity uncovered in this work, support for said heterogeneity can be gleaned from some of the results. In an MS analysis of purified *Rhodopseudomonas palustris* ribosomal proteins, Strader *et al.* 15 observed a number of intact mass species that potentially altered forms of ribosomal proteins. In an MS analysis of the ribosomes of *Caulobacter cresentus*, Running *et al.* 33 observed what appeared to be a secondary, less abundant isoform of ribosomal protein L7/12.

The integration of top-down and bottom-up MS analyses, combining independent biological replicates, multiple *E. coli* strains, and multiple analysis approaches, uncovered a significantly higher level of protein isoform heterogeneity than observed in previous studies. When these results are considered in combination with the previously reported results, it starts to appear that variable and heterogeneous post translational modification of ribosomal proteins is common across prokaryotes.

Early biochemical surveys of ribosomal protein methylation by Chang pointed to some variability in ribosomal protein L7/12 methylation under different growth temperatures[34]. Additional non-MS studies provided evidence of growth stage-dependent and growth condition-dependent ribosome composition and PTM variation, suggesting that a wide variety of growth conditions would need to be sampled to obtain a complete picture of the full diversity of possible states[35, 36].

Many of the protein isoforms we observed were due to variable methylation. Methylation is a common means of regulating the function of nucleic-acid binding proteins[37] and frames a potential role for variable PTMs as a mechanistic component of

the ribosome filter hypothesis - possibly with different species of ribosomes existing in a cell and modulating the translation of different mRNA species. While the N-terminal acetylation of L7/12 has been shown to affect its protein-protein interaction with L10, direct evidence for a function of ribosomal protein methylation on protein-nucleic acid interactions is lacking in bacteria.

However, in other classes of organisms, there is evidence that ribosomal protein methylation has an effect on mRNA selection for translation. In the slime mold *D. discoideum,* ribosomal protein S24 was found to be required for selective binding of certain ribosomal mRNA species, whereas the methylation of S31 resulted in the rapid destabilization of ribosomal protein-encoding mRNAs[38, 39]. The interaction with, and modulation of expression of translation-associated mRNAs by individual bacterial ribosomal proteins could represent function modulated by variable protein methylation. If we include our verified results plus the previously reported isoforms of S6, a combinatoric calculation indicates the ribosome may exist in over 3x104 different states. This is indeed good evidence that the ribosome has a system of "levers and dials" that can be modulated to alter its function or interactions, though the nature of those alterations remains opaque. The full extent of the contribution towards ribosomal heterogeneity by proteins unobserved in our intact mass analysis may yet be found to be greater than that indicated by the limited information afforded by the bottom-up data.

While the combinatorial estimate is large, it is likely to be a significant underestimate. Several of the modifications are known to be modulated by growth conditions, meaning that to find all the heterogeneity we would have to sample all the growth conditions, a somewhat arduous task. Also, we noted some weaker and less

consistent evidence for other possible isoforms, but did not include them in the results or calculations here because they did not meet the inclusion criteria. If even a portion of those are later proven to be real variants, it will expand the number of combinatorial states considerably.

It is not clear whether the observed heterogeneity is due to small population differences, i.e. each cell exhibits each protein in primarily only a single PTM state, or if the full repertoire of heterogeneity exists within individual cells. These two scenarios have significantly different biological implications. If the source is population heterogeneity, it implies that the variety of isoforms may play a role in adapting to environmental stress by readying different cells to respond to different conditions as they arise similar to bistability in genetic networks, though in this case the distinct states would be represented by different modification patterns rather than distinct modes of gene expression. However, if the isoform diversity is at the cellular level, then it may have more to do with stochastic fluctuations within the various modification pathways, the modulation of which, by intra- or extra-cellular conditions, could provide a mechanism for the regulation of mRNA translation via the ribosome.

Table 2.1

Precursor Mass	Theoretical Mass	Mass Difference (Da)	Sequence Start	Sequence End	Amino Acid Sequence	Tags with N-term Mass Matching	Tags with C-term Mass Matching	Tags with both Masses Matching
2982.64	2982.52	0.12	168	249	LTNGFEVTSYIGGEGHNLQEHSVILIR			GHNLQEHS
928.55	882.52	46.03	258	282	VKDLPGVR		LPGV	
1547.89	1574.87	-26.98	108	150	VYTTTPKKPNSALR	TTTP	KPNSAL	

Table 2.1. **MS identification of PTM and single amino acid substitution in ribosomal protein S12 by GFS**. Bottom up peptide MS analysis using GFS identified two distinct peptides with delta mass shifts of approximately +46Da and -27Da corresponding to a β-methylthiolation and K–>T substitution respectively. Sequence tag analysis of MS/MS fragmentation data by GFS pinpointed the D88 and K42 as the location of the respective modification and substitution respectively. Matched sequence tags are indicated in red.

Figure 2.1

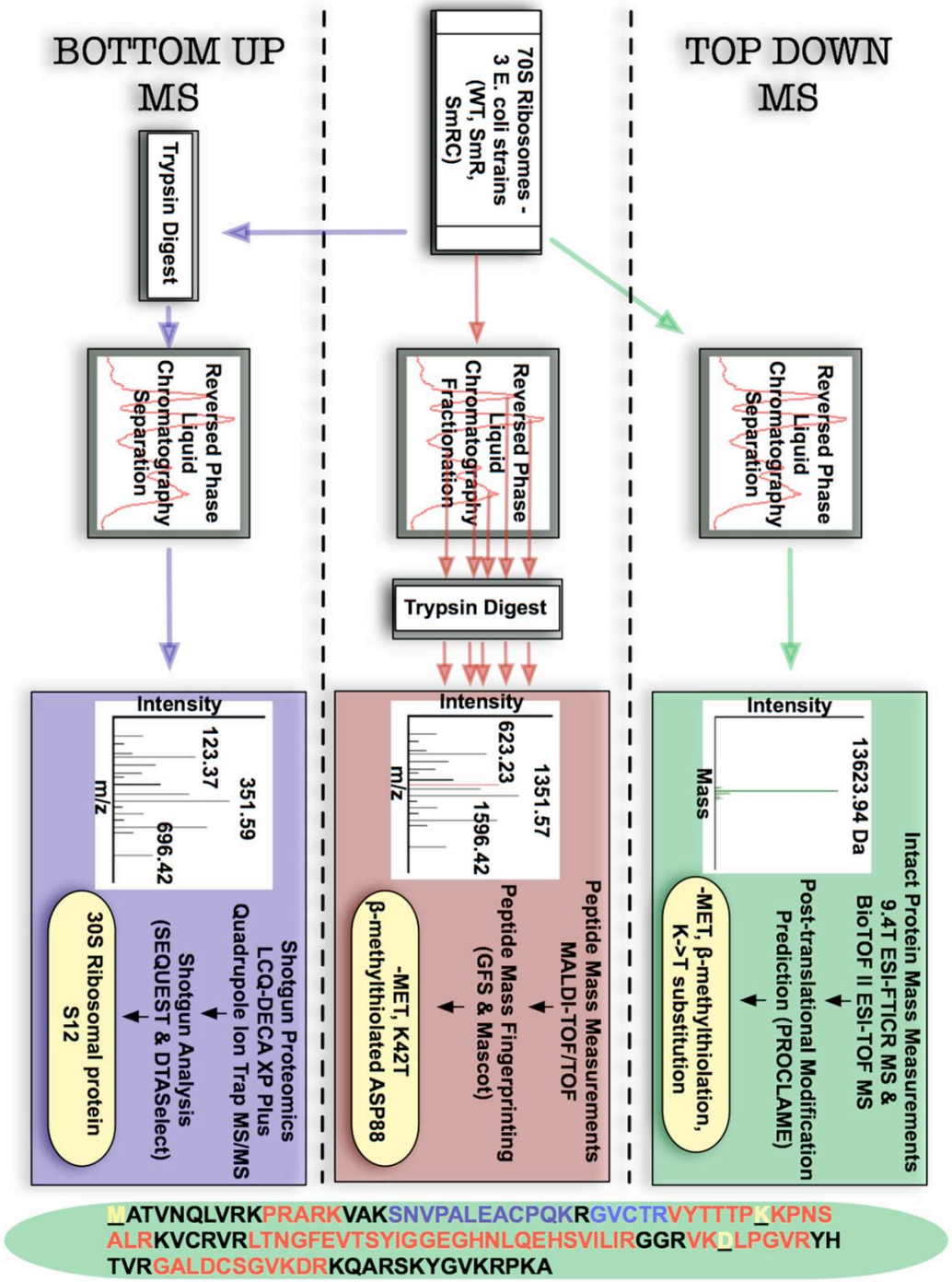
BOTTOM UP MS
TOP DOWN MS
70S Ribosomes - 3 E. coli strains (WT, SmR, SmRC)
Trypsin Digest
Reversed Phase Liquid Chromatography Separation
Reversed Phase Liquid Chromatography Fractionation
Reversed Phase Liquid Chromatography Separation
Trypsin Digest
Intensity
m/z
123.37
351.59
696.42
Shotgun Proteomics
LCQ-DECA XP Plus
Quadrupole Ion Trap MS/MS
Shotgun Analysis
(SEQUEST & DTASelect)
30S Ribosomal protein S12
Intensity
m/z
623.23
1351.57
1596.42
Peptide Mass Measurements
MALDI-TOF/TOF
Peptide Mass Fingerprinting
(GFS & Mascot)
-MET, K42T
β-methylthiolated ASP88
Intensity
Mass
13623.94 Da
Intact Protein Mass Measurements
9.4T ESI-FTICR MS &
BioTOF II ESI-TOF MS
Post-translational Modification
Prediction (PROCLAME)
-MET, β-methylthiolation,
K->T substitution
MATVNQLVRKPRARKVAKSNVPALEACPQKRGVCTRVYTTTPKKPNS
ALRKVCRVRLTNGFEVTSYIGGEGHNLQEHSVILIRGGRVKDLPGVRYH
TVRGALDCSGVKDRKQARSKYGVKRPKA

Figure 2.1. **Protein characterization using a combined Top-down/Bottom-up (TD/BU) mass spectrometry approach.** LC separated ribosomal proteins were analyzed via top- down and bottom-up LC-MS on multiple instruments. Bottom-up MS (shotgun) analysis enabled identification of proteins in the sample. Top-down MS (Intact & PMF) analysis enabled protein identification and prediction of post-translational modifications (and amino acid substitutions) from intact mass measurements. PMF and Tandem MS (MS/ MS) applied in both analyses enabled localization of the post-translational modifications to specific peptides and/or amino acid positions.

Figure 2.2

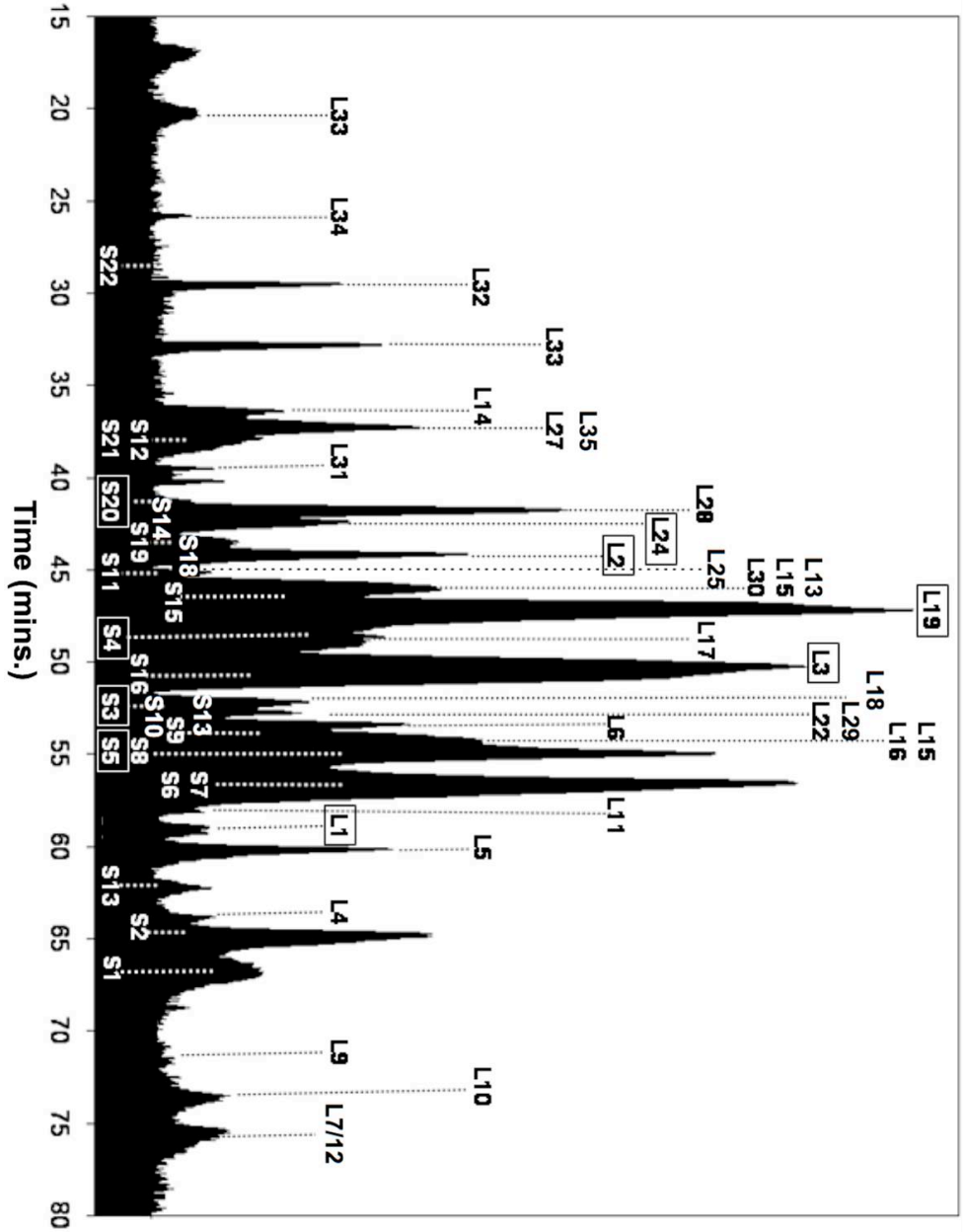
Time (mins.)
15
20
25
30
35
40
45
50
55
60
65
70
75
80
L33
L34
S22
L32
L33
L14
L27
L35
S12
S21
L31
S20
S14
S19
S18
S11
S15
S4
S16
S3
S10
S9
S13
S8
S5
S7
S6
L28
L24
L2
L25
L30
L15
L13
L19
L17
L3
L18
L29
L22
L6
L16
L15
L11
L1
L5
S13
L4
S2
S1
L9
L10
L7/12

Figure 2.2. Reversed phase liquid chromatography (RP-LC) separation UV trace of ribosomal proteins isolated from WT *E. coli*. Ribosomal proteins were separated by a single round of RP-LC with fractions collected at 1 minute intervals in a 96-well plate. Following lyophilization and resuspension, individual fractions were enzymatically digested with trypsin and analyzed by MALDI MS/MS. For creation of the elution map the constituent proteins of each fraction were identified via peptide mass fingerprinting (PMF) using GFS and Mascot.

Figure 2.3

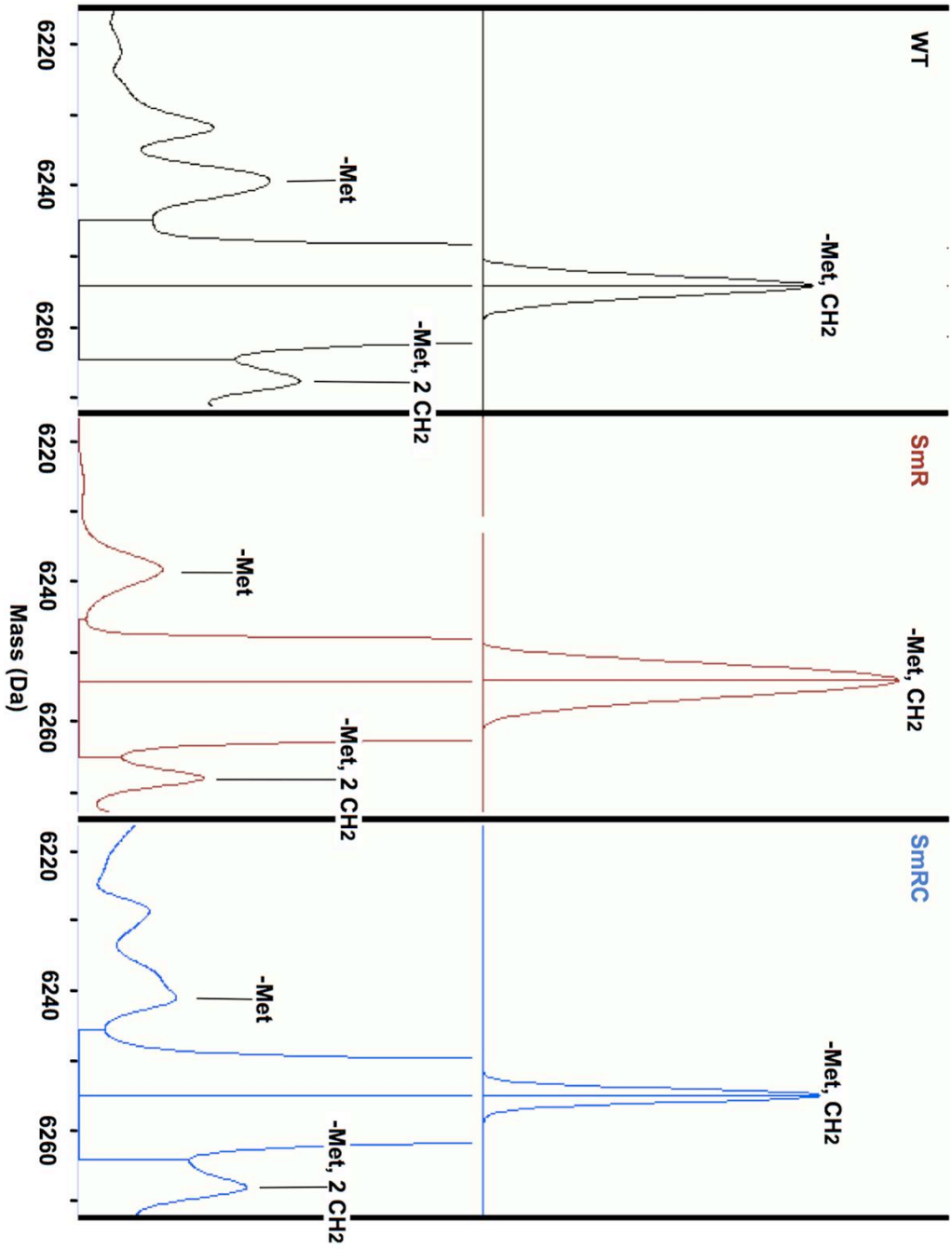

Figure 2.3. Intact isoforms of ribosomal proteins L33 identified in deconvoluted spectra. Three isoforms of L33 were observed in ribosomal protein derived from WT (black), SmR (red) and SmRC (blue) derivatives. In each case a dominant monomethylated isoform at ~ 6254 Da was bounded by two secondary isoforms at ~ 6240 Da and ~ 6268 Da predicted by PROCLAME to be due to the loss of the N-terminal methionine and variable protein methylation (CH2). The intensity axis for each spectra were scaled to highlight the very low-intensity 6240 Da and 6268 Da isoforms along with the primary 6254 Da isoform peaks.

Figure 2.4

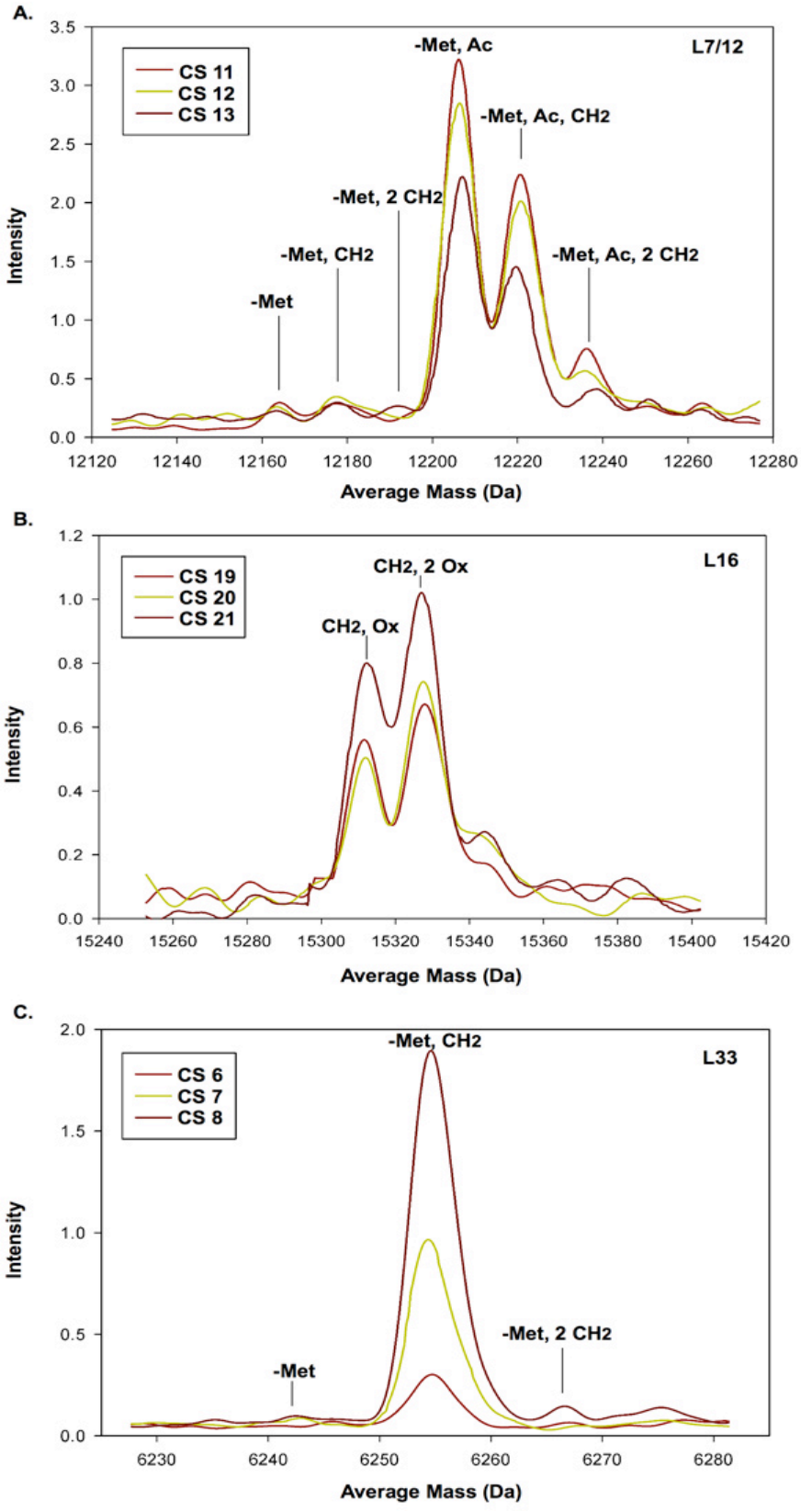
A.
L7/12
CS 11
CS 12
CS 13
-Met, Ac
-Met, Ac, CH2
-Met, 2 CH2
-Met, CH2
-Met
-Met, Ac, 2 CH2
Intensity
Average Mass (Da)
B.
L16
CS 19
CS 20
CS 21
CH2, 2 Ox
CH2, Ox
Intensity
Average Mass (Da)
C.
L33
CS 6
CS 7
CS 8
-Met, CH2
-Met, 2 CH2
-Met
Intensity
Average Mass (Da)

Figure 2.4. **Overlays of multiple charge states in smoothed m/z spectra supports the presence of multiple protein isoforms due to differential modification in the m/z spectra.** For low-abundance isoforms, a key challenge was to determine whether these were real, noise or software deconvolution artifacts. Multiple regions of m/z spectra representing individual protein charge states (CS) were aligned and overlaid using SpecAlign and Shape Finder (see methods). The x-axis indicates the resulting deconvoluted intact mass. The averaging of spectra is a good representation of the dataset as the suppression of noise and enhancement of repeated signals across each charge state are simultaneously achieved. (A) Multiple isoforms of L7/12 were identified due to combinations of N- terminal methionine cleavage (-Met), acetylation (Ac) and methylation (CH2). (B) Two isoforms of L16 are shown predicted by PROCLAME to be due to methylation and either one or two oxidations (Ox). (C) L33 also exhibited variable methylation, with masses observed corresponding to methionine loss and either 0, 1 or 2 methylations.

Figure 2.5

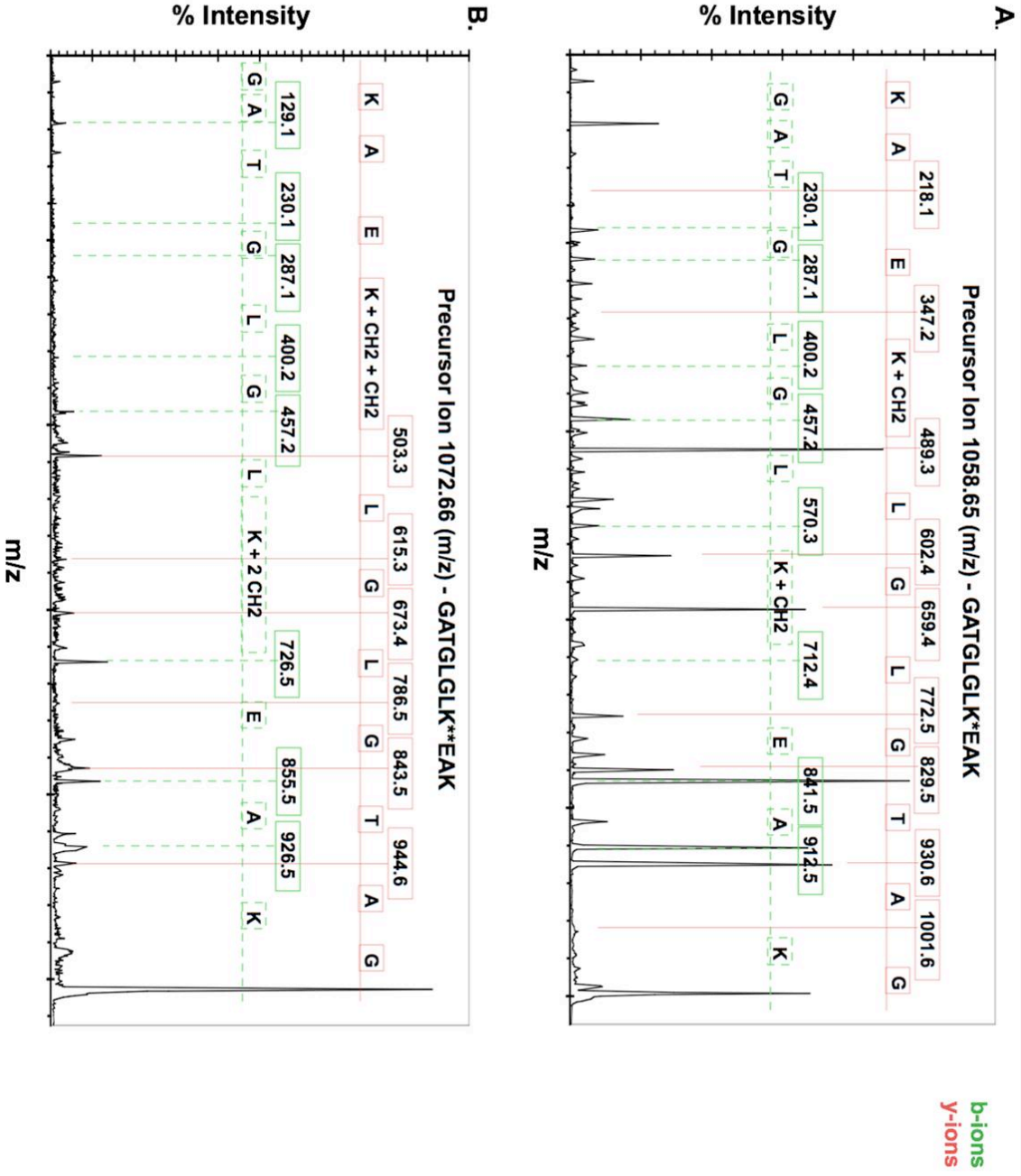
A.
Precursor Ion 1058.65 (m/z) - GATGLGLK*EAK
% Intensity
m/z
b-ions
y-ions
B.
Precursor Ion 1072.66 (m/z) - GATGLGLK**EAK
% Intensity
m/z

Figure 2.5. Multiple modification states of a single L7/12 peptide. The peptide GATGLGLKEAK corresponding to amino acids 75-85 in ribosomal protein L7/L12 was observed by Mascot analysis of MALDI TOF/TOF MS/MS data to be variably modified, existing in either (A) an ~ 1058.65 Da monomethylated state or (B) an ~ 1072.66 Da dimethylated state. Lys82 was determined to be the site of modification in both peptide isoforms. The same peptide was also identified in an unmodified state (data not shown). b-ion (green) and y-ion (red) sequences are illustrated.

Figure 2.6

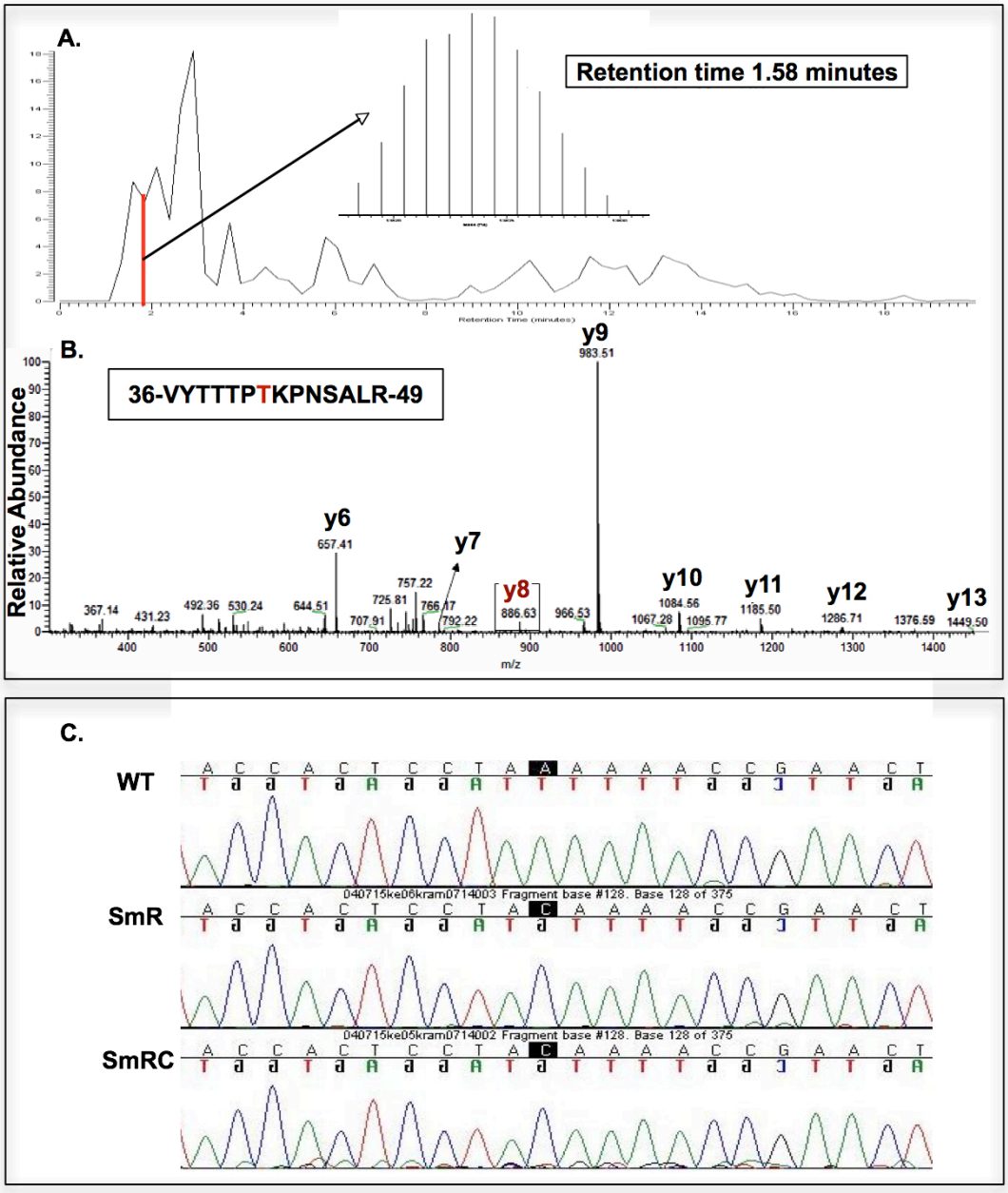
A.
Retention time 1.58 minutes
Retention Time (minutes)
B.
36-VYTTTPTKPNSALR-49
Relative Abundance
m/z
y6
657.41
y7
y8
886.63
y9
983.51
y10
1084.56
y11
1185.50
y12
1286.71
y13
1449.50
C.
WT
SmR
SmRC
040715ke06kram0714003 Fragment base #128. Base 128 of 375
040715ke05kram0714002 Fragment base #128. Base 128 of 375

Figure 2.6. Top-down/Bottom-up MS identified an amino acid substitution in protein S12 between strains, a likely determinant of phenotypic differences. (A) Total ion chromatogram (TIC) from the LC-FTICR analysis and deconvoluted spectrum of the peak at 1.58 mins. (B) Shotgun (MS/MS) analysis of SmR-derived ribosomes identified a peptide from protein S12 containing an amino acid substitution (K42T) previously associated with streptomycin resistance. (C) DNA sequencing of the SmR *rpsL* locus identified a point mutation in codon 43 in both SmR and SmRC.

References

1. Veening, J. W.; Smits, W. K.; Kuipers, O. P., Bistability, epigenetics, and bet-hedging in bacteria. *Annu Rev Microbiol* **2008,** 62, 193-210.

2. Leisner, M.; Stingl, K.; Frey, E.; Maier, B., Stochastic switching to competence. *Curr Opin Microbiol* **2008,** 11, (6), 553-9.

3. Ozbudak, E. M.; Thattai, M.; Lim, H. N.; Shraiman, B. I.; Van Oudenaarden, A., Multistability in the lactose utilization network of Escherichia coli. *Nature* **2004,** 427, (6976), 737-40.

4. Hess, J. F.; Oosawa, K.; Kaplan, N.; Simon, M. I., Phosphorylation of three proteins in the signaling pathway of bacterial chemotaxis. *Cell* **1988,** 53, (1), 79-87.

5. Silversmith, R. E.; Bourret, R. B., Throwing the switch in bacterial chemotaxis. *Trends Microbiol* **1999,** 7, (1), 16-22.

6. Polevoda, B.; Sherman, F., Methylation of proteins involved in translation. *Mol Microbiol* **2007,** 65, (3), 590-606.

7. Chow, C. S.; Lamichhane, T. N.; Mahto, S. K., Expanding the nucleotide repertoire of the ribosome with post-transcriptional modifications. *ACS Chem Biol* **2007,** 2, (9), 610-9.

8. Lhoest, J.; Colson, C., Cold-sensitive ribosome assembly in an Escherichia coli mutant lacking a single methyl group in ribosomal protein L3. *Eur J Biochem* **1981,** 121, (1), 33-7.

9. Gordiyenko, Y.; Deroo, S.; Zhou, M.; Videler, H.; Robinson, C. V., Acetylation of L12 increases interactions in the Escherichia coli ribosomal stalk complex. *J Mol Biol* **2008,** 380, (2), 404-14.

10. Wool, I. G., Extraribosomal functions of ribosomal proteins. *Trends Biochem Sci* **1996,** 21, (5), 164-5.

11. Warner, J. R.; McIntosh, K. B., How common are extraribosomal functions of ribosomal proteins? *Mol Cell* **2009,** 34, (1), 3-11.

12. Mauro, V. P.; Edelman, G. M., The ribosome filter hypothesis. *Proc Natl Acad Sci U S A* **2002,** 99, (19), 12031-6.

13. Mauro, V. P.; Edelman, G. M., The ribosome filter redux. *Cell Cycle* **2007,** 6, (18), 2246-51.

14. Nomura, M.; Gourse, R.; Baughman, G., Regulation of the synthesis of ribosomes and ribosomal components. *Annu Rev Biochem* **1984,** 53, 75-117.

15. Strader, M. B.; Verberkmoes, N. C.; Tabb, D. L.; Connelly, H. M.; Barton, J. W.; Bruce, B. D.; Pelletier, D. A.; Davison, B. H.; Hettich, R. L.; Larimer, F. W.; Hurst, G. B., Characterization of the 70S Ribosome from Rhodopseudomonas palustris using an integrated "top-down" and "bottom-up" mass spectrometric approach. *J Proteome Res* **2004,** 3, (5), 965-78.

16. Holmes, M. R.; Giddings, M. C., Prediction of posttranslational modifications using intact-protein mass spectrometric data. *Anal Chem* **2004,** 76, (2), 276-82.

17. Giddings, M. C.; Shah, A. A.; Gesteland, R.; Moore, B., Genome-based peptide fingerprint scanning. *Proc Natl Acad Sci U S A* **2003,** 100, (1), 20-5.

18. Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S., Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **1999,** 20, (18), 3551-67.

19. Arnold, R. J.; Reilly, J. P., Observation of Escherichia coli ribosomal proteins and their posttranslational modifications by mass spectrometry. *Anal Biochem* **1999,** 269, (1), 105-12.

20. Hirel, P. H.; Schmitter, M. J.; Dessen, P.; Fayat, G.; Blanquet, S., Extent of N-terminal methionine excision from Escherichia coli proteins is governed by the side-chain length of the penultimate amino acid. *Proc Natl Acad Sci U S A* **1989,** 86, (21), 8247-51.

21. Ben-Bassat, A.; Bauer, K.; Chang, S. Y.; Myambo, K.; Boosman, A.; Chang, S., Processing of the initiation methionine from proteins: properties of the Escherichia coli methionine aminopeptidase and its gene structure. *J Bacteriol* **1987,** 169, (2), 751-7.

22. Demirci, H.; Gregory, S. T.; Dahlberg, A. E.; Jogl, G., Multiple-site trimethylation of ribosomal protein L11 by the PrmA methyltransferase. *Structure* **2008,** 16, (7), 1059-66.

23. Vanet, A.; Plumbridge, J. A.; Guerin, M. F.; Alix, J. H., Ribosomal protein methylation in Escherichia coli: the gene prmA, encoding the ribosomal protein L11 methyltransferase, is dispensable. *Mol Microbiol* **1994,** 14, (5), 947-58.

24. Cameron, D. M.; Gregory, S. T.; Thompson, J.; Suh, M. J.; Limbach, P. A.; Dahlberg, A. E., Thermus thermophilus L11 methyltransferase, PrmA, is dispensable for growth and preferentially modifies free ribosomal protein L11 prior to ribosome assembly. *J Bacteriol* **2004,** 186, (17), 5819-25.

25. Brosius, J.; Chen, R., The primary structure of protein L16 located at the peptidyltransferase center of Escherichia coli ribosomes. *FEBS Lett* **1976,** 68, (1), 105-9.

26. Shacter, E., Quantification and significance of protein oxidation in biological samples. *Drug Metab Rev* **2000,** 32, (3-4), 307-26.

27. Frottin, F.; Martinez, A.; Peynot, P.; Mitra, S.; Holz, R. C.; Giglione, C.; Meinnel, T., The proteomics of N-terminal methionine cleavage. *Mol Cell Proteomics* **2006,** 5, (12), 2336-49.

28. Kang, W. K.; Icho, T.; Isono, S.; Kitakawa, M.; Isono, K., Characterization of the gene rimK responsible for the addition of glutamic acid residues to the C-terminus of ribosomal protein S6 in Escherichia coli K12. *Mol Gen Genet* **1989,** 217, (2-3), 281-8.

29. Paulander, W.; Maisnier-Patin, S.; Andersson, D. I., The Fitness Cost of Streptomycin Resistance Depends on rpsL Mutation, Carbon Source and RpoS ({sigma}S). *Genetics* **2009**.

30. Maisnier-Patin, S.; Berg, O. G.; Liljas, L.; Andersson, D. I., Compensatory adaptation to the deleterious effect of antibiotic resistance in Salmonella typhimurium. *Mol Microbiol* **2002,** 46, (2), 355-66.

31. Kowalak, J. A.; Walsh, K. A., Beta-methylthio-aspartic acid: identification of a novel posttranslational modification in ribosomal protein S12 from Escherichia coli. *Protein Sci* **1996,** 5, (8), 1625-32.

32. Carr, J. F.; Hamburg, D. M.; Gregory, S. T.; Limbach, P. A.; Dahlberg, A. E., Effects of streptomycin resistance mutations on posttranslational modification of ribosomal protein S12. *J Bacteriol* **2006,** 188, (5), 2020-3.

33. Running, W. E.; Ravipaty, S.; Karty, J. A.; Reilly, J. P., A top-down/bottom-up study of the ribosomal proteins of Caulobacter crescentus. *J Proteome Res* **2007,** 6, (1), 337-47.

34. Chang, F. N., Temperature-dependent variation in the extent of methylation of ribosomal proteins L7 and L12 in Escherichia coli. *J Bacteriol* **1978,** 135, (3), 1165-6.

35. Deusser, E.; Wittmann, H. G., Ribosomal proteins: variation of the protein composition in Escherichia coli ribosomes as function of growth rate. *Nature* **1972,** 238, (5362), 269-70.

36. Deusser, E., Heterogeneity of ribosomal populations in Escherichia coli cells grown in different media. *Mol Gen Genet* **1972,** 119, (3), 249-58.

37. Gary, J. D.; Clarke, S., RNA and protein interactions modulated by protein arginine methylation. *Prog Nucleic Acid Res Mol Biol* **1998,** 61, 65-131.

38. Mangiarotti, G., Synthesis of ribosomal proteins in developing Dictyostelium discoideum cells is controlled by the methylation of proteins S24 and S31. *Biochem Cell Biol* **2002,** 80, (2), 261-70.

39. Mangiarotti, G., Two Dictyostelium ribosomal proteins act as RNases for specific classes of mRNAs. *Biochem J* **2003,** 370, (Pt 2), 713-7.

40. Vasa, S. M.; Guex, N.; Wilkinson, K. A.; Weeks, K. M.; Giddings, M. C., ShapeFinder: a software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis. *Rna* **2008,** 14, (10), 1979-90.

Chapter 3

Mass Spectrometry Analysis of Variable Methylation of *Escherichia coli* Ribosomal Protein L7/12 Reveals Novel Methylation Sites and Insights Into Regulation

Abstract

Escherichia coli ribosomal protein L7/L12 is known to be modified by both N-terminal acetylation and the monomethylation of Lys82. When purified and studied by intact-mass spectrometry, additional peaks are often observed in a spectrum, presumably due to other modification isoforms. We investigated these additional isoforms through a study that combined chromatographic purification and isolation of L7/L12 with analysis by both intact-mass spectrometry and bottom-up tandem mass spectrometry. We also examined a panel of 28 gene disruption mutants in putative and characterized methyltransferases to determine whether any of them disrupt methylation of the protein. Finally, we followed a previously noted effect of growth temperature on methylation of the protein, by examining specifically how temperature affects each identified site of methylation. Our results indicate three new methylation sites, each located in the C-terminal domain of the protein where it interacts with initiation factor 2 (IF-2), elongation factor G (EF-G) and elongation factor Tu (EF-Tu). We observed that growth temperature primarily modulated methylation at Lys82 rather than affecting methylation patterns at the other sites. It was somewhat surprising that our panel of methyltransferase mutants did not identify any effects on L7/L12 methylation, leaving it a remaining mystery. The function of the newly uncovered sites of variable methylation in L7/12 remains unclear,

but the localization of the modified amino acid residues to the C- terminal region suggests a role in modulating the interactions of L7/12 with other accessory components of the translation machinery.

3.1 Introduction

Though the ribosome is a sophisticated molecular machine that plays a vital role in maintaining nearly all cellular life, its complexity has challenged efforts to fully elucidate the mechanism by which it is translational behavior is regulated. One aspect of the ribosome's complexity is the extensive enzymatic co- or post-translational modification that occurs on both its RNA and protein components. Both its RNA and protein modifications remain to be fully elucidated, and the role or function of protein modifications that have been identified is not clear in most instances.

Methylation is by far the most common of the prokaryotic ribosomal modifications identified. The methylating enzymes - methyltransferases (MTases) - target ribosomal RNAs (rRNA), ribosomal proteins, and transfer RNAs (tRNAs). The MTases also target various translation factors, indicating a potentially important role in regulating or promoting translation. Little is known of the functional significance of the methylation of ribosomal proteins, which in *Escherichia coli* includes ribosomal proteins S11, L3, L7/12, L11, L16 and L33[1]. This high frequency of protein methylation comes at a significant metabolic cost. It also exhibits a high level of evolutionary conservation across bacterial species. These facts suggest that methylation plays important, albeit as yet unknown roles in the cell[2].

Most of the methylated proteins have important functions in various aspects of translation, and some of them have been the focus of significant research efforts. L3 is a monomethylated ribosomal protein that was shown to be important in ribosome assembly[3]. S11, also monomethylated, is located on the platform of the 30S subunit and

forms part of the Shine-Delgarno cleft of the 70S ribosome and was shown to be important in the control of translational fidelity[4]. L11, the most extensively methylated ribosomal protein, forms part of the GTPase-associated center of the 50S subunit and is important in the processes of decoding, tRNA accommodation and translocation[5, 6]. Few ribosomal proteins though have garnered as much interest from the scientific research community as L7/12.

Encoded by the *rplL* gene in *E. coli* and differing only in the presence (L7) or absence (L12) of an N-terminal acetylation, ribsomal protein L7/12 is the only multi-copy protein present in the bacterial ribosome. It is found in the ribosome as a pair of heterodimers. It is the main component of the pentameric ribosomal stalk complex and associates with the ribosome thorough its interactions with the other stalk component, ribosomal protein L10. L7/12 can essentially be divided into three structural regions - an N-terminal region (residues 2-38; methionine-1 is typically cleaved) which mediates the protein's interactions with L10, a flexible hinge region (residues 39-50) which facilitates the numerous conformations that have been observed and links the N-terminal region to the C-terminal region (residues 51-121), the latter which plays key roles in translation factor recruitment. Extensive studies have enabled elucidation of the structure and functional significance of the ribosomal stalk[7, 8].

L7/12 has been shown to directly and specifically interact with initiation factor 2 (IF2), elongation factors Tu (EF-Tu) and G (EF-G), and release factor 3 (RF3) via a conserved region of its C-terminal domain[9], and in so doing enhance the GTPase activity

of both elongation factors[10]. Yet despite extensive research, little is known about the significance of the PTMs harbored by L7/L12, particularly its methylation.

The previously characterized isoforms of L7/12 are due to a combination of multiple post-translational modification mechanisms, including N-terminal methionine cleavage (methionine-1), N-acetylation of the resulting N-terminal serine-2 (to form L7), and internal monomethylation of lysine-82. The enzymes responsible for the N-terminal processing of L7/12 have been identified as methionine aminopeptidase MAP[11] and serine-acetyltransferase RimL, resulting in N-terminal cleavage and acetylation respectively[12]. It undergoes cleavage of Met1 in accordance with the general rule that nearly all proteins with small side chains on the residue at position two undergo cleavage (approximately 90% cleavage for Ser2)[13, 14].

Though protein N-acetylation is comparatively rare in *E. coli,* surveys of protein N-acetylation in both eukaryotes and prokaryotes suggest that N-acetylation at Ser residues has one of the highest likelihoods, explaining the N-terminal cleavage of L7/12[15]. Yet a significant fraction of the cellular L7/12 pool remains unacetylated (L12). A number of studies have uncovered evidence of biological regulation of L7/12 acetylation. Deusser *et al*[16] and Ramagopaul *et al*[17] observed that the ratio of L7 (acetylated) : L12 (unacetyalted) varied as a function of growth stage. More recently Gordiyenko *et al* showed that this increase served to enhance L7/12's interaction with L10, likely as part of a cellular strategy to increase the stability of the stalk under stressful conditions[18].

Early biochemical evidence of L7/12 monomethylation was uncovered by Chang[20, 21]. An increase in the amount of monomethylated L7/12 in cells grown at lower temperatures was also noted[19]. This indicated that there may also be some regulation of L7/12 methylation, though much less is understood about it compared to the N-terminal processing. These unknowns include both the function significance of the modification and the identity of the cognate methyltransferase.

Arnold and Reilly identified Lys82 as a site of monomethylation in L7/12 by mass spectrometry (MS)[22]. In a recent MS analysis of protein isoform heterogeneity in *E. coli* ribosomal proteins carried out in our lab, we observed a number of low-abundance isoforms of L7/12 that hinted at the presence of novel alternate PTM isoforms. Many of these exhibited intact masses that were predicted to be due to additional methylations, indicating that secondary methylation sites may yet remain to be uncovered. We investigated this possibility via tandem mass spectrometry (MS/MS) analysis of L7/12 peptides wherein tryptic peptides were fragmented in the mass spectrometer to yield amino acid sequence information. Those experiments provided the discovery of a new dimethylated isoform, dimethylated at the same residue previously known to be harbor a single methylation.

In the aforementioned experiments, MS/MS was carried out in an automated, data-dependent fashion, with only the top 10 most intense peaks in the MS spectra being selected for collision-induced dissociation (CID). Hence, some lower-abundance peptides were skipped for MS/MS, including some whose precursor masses indicate the potential for holding additional methylation sites. This evidence of additional heterogeneity in

L7/L12 was too tempting to overlook. We performed further experiments to uncover the types and locations of additional modifications present on L7/L12, the nature of methylation regulation, and to attempt to identify the enzyme(s) responsible for its methylation.

To identify any additional sites of methylation we carried out independent biological replicate experiments in which we purified L7/12 from *E. coli* ribosomal protein extracts using either reversed-phase liquid chromatography (RP-LC) or one-dimensional polyacrylamide gel electrophoresis (1-D PAGE). We then performed both electrospray ionization-based (ESI) tandem mass spectrometry (MS/MS) and matrix assisted laser desorption/ionization-based (MALDI) MS/MS analyses of L7/12 peptides generated via tryptic digestion. We manually targeted peptides for fragmentation analysis that were shifted by mass multiples of ~+14 Da, indicative of methylation events.

We sought insight into the nature and dynamics of the regulation of L7/12 methylation by conducting additional MS-based analyses of both intact and tryptic L7/12 protein/peptides derived from cells cultured under two distinct growth temperatures, specifically comparing the levels of protein monomethylation under each growth condition. To examine whether L7/12 protein may be methylated pre-assembly, we constructed and expressed a N-terminally tagged version of the protein in *E. coli*, adding a 6x His-tag at the N-terminus to block L7/12 - L10 interaction, and carried out intact MS measurements to determine whether the free protein could be methylated. We additionally undertook a MS-based screen of a panel of single-gene knockout mutants of

both characterized and putative methyltransferases in an attempt to identify the L7/12 methyltransferase.

3.2 Materials and Methods

Bacterial strains and plasmids. *E. coli* K12 MG1655 was a kind gift from Dr. R. Bourret (UNC Chapel Hill). *E. coli* strain BW25113 and the single gene knockouts used in the screen for the L7/12 methyltransferase were obtained from the Keio Collection - National Institute of Genetics, Japan, either directly or via the Coli Genetic Stock Center (CGSC) at Yale University. Plasmid pQE-30, used in the expression of the His-tagged L7/12, was commercially obtained from Qiagen (Valencia, CA).

Materials. Salts, buffers, glacial acetic acid and sucrose were obtained from Fisher Scientific. RNase-free DNase 1, formic acid and dithiothreitol (DTT) were obtained from Sigma-Aldrich (St Louis, MO). RNase Away (Molecular BioProducts, San Diego, CA), was used to treat all glassware and RNase/Protease-free water was used to make all buffers to minimize RNase activity during ribosome isolation.

Cell culture. Wild-type *E. coli* was cultured in LB media and harvested after a 12 hrs. of growth. All strains were routinely cultured under aeration at temperatures of either 37°C or 28°C as indicated. For *E. coli* mutants obtained from the CGSC and Keio Collection, LB growth media was supplemented with 25ug/ml kanamycin. For the expression of His-tagged L7/12, cells were first cultured to an OD_{600} of 0.3 before expression was induced by the addition of 10mM isopropyl-beta-D-thiogalactopyranoside (IPTG). Expression was carried out for 3 hours before cells were harvested.

DNA manipulations. Restriction enzymes were obtained from New England Biolabs (Ipswich, MA). Primers and sequencing services were obtained through MWG-Biotech (Huntsville, AL). Digestions, ligations, PCR and cloning were performed using manufacturer recommended and standard procedures. Expression of L7-12 was carried out in XL1-Blue competent cells transformed with the pQE-30-L7/12 expression vector. Briefly a wild type copy of the *rplL* gene was amplified by standard whole-cell PCR (primers: forward- ATG TCT ATC ACT AAA GAT CAA ATC ATT, reverse- TTA TTT AAC TTC AAC TTC AGC GC.) PCR products were purified using QIAquick PCR Purification Kit (Qiagen, Valencia, CA) prior to ligation into the pre-linearized pQE-30 vector using the QIAexpress UA cloning kit (Qiagen) following manufacturer recommended protocols. XL1-Blue competent cells were transformed with the His-tagged L7/12 expression vector using standard heat shock methods. Transformants were selected on LB agar plates supplemented with 0.1mg/ml ampicillin.

Purification of ribosomes. *E. coli* cultures were harvested by centrifugation, washed, and the cells lysed by sonication. Ribosomes were purified as previously described in Strader et al. Briefly ribosomes were isolated by pelletting through a 1.1M sucrose/lysis buffer cushion and further purified from other large protein complexes by fractionation through a 7%-30% sucrose gradient. Ribosome-containing fractions were determined by measuring absorbance at 260nm and pooled for protein isolation. RNA was precipitated from the ribosome preparation using standard acid extraction techniques and removed by centrifugation. The clarified ribosome supernatant was dialyzed using 3500MW cutoff slide-a-lyzer cassettes (Pierce, Rockford, IL) against a total of 10 liters of ultra-pure 18M∧ water (Millipore Corp., Billerica, MA) over 36 hours at 4°C. The ribosomal

protein extract was concentrated by lyophilization and the concentration determined by measuring absorbance at 280nm using a Nanodrop spectrophotometer (Thermo Scientific, Wilmington, DE).

L7/12 isolation and mass spectrometry analysis. L7/12 protein was isolated from total ribosomal protein using either reversed-phase HPLC (RP-LC) or 1-D SDS PAGE. For RP-LC approximately 400ug of protein was loaded onto a C5 reversed-phased column (Discovery BIO Wide Pore C5, 10cm x 2.1mm, 3μm, Supelco, Bellefonte, PA) and separated using a gradient of 10%-70% acetonitrile/0.1% formic acid at a flow rate of 0.1ml/min. Fractions containing L7/12 were determined from the UV chromatogram and confirmed using intact protein mass spectrometry and peptide mass fingerprinting (PMF). Isolation of His-tagged L7/12 protein was carried out using the HisLink protein purification resin (Promega Corp., Madison, WI). Briefly, clarified supernatant from sonicated cell cultures was passed through the resin column to bind the His-tagged protein. Protein binding was followed by subsequent wash and elution steps using increasing concentrations of imidazloe buffer. Off-line intact mass analysis of RP-LC separated L7/12 protein were carried out on either a Bruker BioTOF II ESI-TOF MS (Bruker Daltonics, Billerica, MA) or an Applied Biosystems PE Sciex Q-Star Pulsar quadrupole time-of-flight (TOF) MS (Applied Biosystems, Foster City, CA). L7/12 proteins were additionally isolated using 1-D SDS PAGE prior to tryptic digestion and MS analysis for modification site determination. Briefly total ribosomal protein were separated through 12% precast ReadyGel polyacrylamide gels (Bio-Rad Laboratories, Hercules CA) along with reference molecular weight ladder (Bio-Rad Laboratories.) The gels were fixed (40% MeOH, 10% HAc) for 30 mins and protein bands visualized by

staining with GelCode Blue Stain Reagent (Thermo Scientific, Rockford, IL). Following destaining (10% MeOH, 7% HAc) and wash steps, protein bands corresponding to L7/12 were excised and subjected to in-gel digestion with sequencing grade trypsin (Promega Corp.) The resulting peptides were cleaned using 10 μl ZipTip pipette tips (Millipore Corp.) and combined with alpha-cyano-4-hydroxycinnamic acid matrix for MS/MS analysis on a MALDI TOF/TOF 4700 proteomics analyzer (Applied Biosystems).

Data analysis. Mass spectra generated on the various mass spectrometers utilized in this study were processed using instrument-specific software. Intact protein mass spectra obtained on the BioTOF II were analyzed using Bruker Daltonics' Data Analysis software (ver. 3.2.121). Typically spectra were subjected to one round of smoothing followed by deconvolution using the maximum entropy function (Instrument Resolving Power = 10000, Data point spacing = 0.1 m/z, Resolution = normal). Spectra generated using the Q-Star Pulsar were analyzed using Applied Biosystems' Analyst (ver. 1.1.0) software package. Peptide MS and MS/MS spectra generated by MALDI analysis were processed and analyzed using Applied Biosystems' Data Explorer (ver. 4.3) software. Intact mass spectra generated on the Q-Star Pulsar were analyzed in Analyst QS (ver. 1.1) (Applied Biosystems) with protein intact masses generated using the Bayesian Protein Reconstruct function. Peptide mass fingerprinting analysis for protein identification and PTM characterization was carried out using Mascot (Matrix Science, Boston, MA) and InsPecT (University of California, San Diego), searching against the *E. coli* K12 MSDB and NCBI protein databases respectively. For database searches a maximum of 2 missed cleavages, a peptide tolerance of 1 Da and a MS/MS tolerance of 0.5 Da were considered.

The most commonly considered variable modifications were acetylation, mono-, di- and tri-methylation.

3.3 Results

3.3.1 Tandem Mass Spectrometry Identifies Novel Sites of L7/12 Methylation

In a previous MS analysis of intact ribosomal proteins we identified masses predicted to be the result of multiply methylated ribosomal protein L7/12. To confirm our predictions L7/12 was purified from *E. coli* ribosomal proteins extracts as described in the methods and analyzed by tandem MS to characterize any methylated peptides. L7/12 protein was isolated from ribosomes using 1-D SDS PAGE, excised from the gel and subjected to in-gel trypsin digestion. MALDI TOF/TOF MS/MS analyses of the resulting tryptic L7/12 peptides were carried out to identify any novel sites of L7/12 methylation. Peptide mass fingerprinting (PMF) analysis of the resulting data identified 7 peptides, encompassing a minimum of 4 unique modification sites, that exhibited isoforms with masses shifted by ~+14 Da, indicative of methylation (figure 3.1). The majority of our observations were corroborated in subsequent experimental and independent biological repeats. Excluding Lys82, the MS/MS analysis identified three novel sites of protein methylation (table 3.1 & figure 3.2). PMF was carried out using Mascot (MS/MS Ion Search mode)[23] and our Genome Fingerprint Scanning (GFS) software[24], searching against the MSDB *E. coli* protein database and the *E. coli* K12 genome respectively, considering variable peptide mono- and di-methylation.

The low signal to noise ratios of some peaks in the MS/MS scans, possibly due to the low abundance of the peptides themselves, translated into some ambiguity in the

assignment of the specific methylated residue in the modified peptides. For example, the tryptic peptide GATGLGLKEAK (residues 75-85) (m/z 1058.6) containing the previously observed monomethylated Lys82 was identified in Mascot analysis as being possibly methylated at either Glu81, Lys82 or Lys85 with similar scores, all above Mascot's threshold score for an identity match. Mascot seeks the minimal set of proteins that can account for the peptide matches found (the Principal of Parsimony) which could potentially result in some bias in the assignment of an MS/MS spectrum to a particular protein due to the presence of other peptides matching the same protein. We attempted to addressed the possibility that the matching of any low-scoring MS/MS spectrum to L7/12 may have been biased by the presence of more abundant L7/12 peptides by performing a complementary data analysis using InsPecT[25]. InsPecT, an MS/MS database search tool, allows identification of post-translationally modified peptides from a single MS/MS spectrum. InsPecT searches were similarly carried out against the entire *E. coli* protein database and largely mirrored the Mascot results, matching each of the modified peptides to L7/12 and occasionally indicating alternative methylation sites within a peptide. The sites of modification indicated in table 3.1 are the best consensus of both manual and software-based spectra analysis. The highest scoring Mascot and InsPecT result from software analysis of biological and experimental repeats are indicated where the software result agreed with manual analysis. MS/MS fragmentation ion mass lists used in these runs are provided as supplementary data.

Interestingly, a comparison of the relative intensity ratios of the methylated : unmethylated versions of the observed peptides suggest that methylation at the newly discovered sites is quite frequent, often indicating a near 1:1 ratio (table 3.1). However

intact L7/12 is primarily observed in either unmethylated or monomethylated states suggesting that, while not strictly mutually exclusive, monomethylation at one site may discourage methylation at another. If this were not the case one would expect multiply-methylated intact isoforms of L7/12 to be strongly represented in intact MS data. A small possibility exists that a fraction of observed protein isoforms attributed to L7 in our intact mass assignments could in fact have been due to, and gone undistinguished from, trimethylated L12. But if this were the case one would expect to also observe trimethylated isoforms of L7. For reasons yet unclear, previous experiments have observed that L12 may be monomethylated more readily than L7[26]. Whether the mechanisms underlying this disparity could be discouraging higher methylated states of L7 remains an interesting area for future analysis.

In addition to the discovery of novel monomethylated residues, MS/MS analyses of the peptide comprising amino acid residues 76-82 GATGLGLK, and the slightly longer 76-85 GATGLGLKEAK, the result of a single missed tryptic cleavage, yielded the only case of dimethylation observed in our studies. Both peptides were observed at the +14 Da monomethylated mass at m/z 730.41 Da and 1058.61 Da respectively, however only the shorter peptide, was observed in its unmethylated state (716.39 Da) (figure 3.3). The longer peptide was the only one additionally observed at a mass shifted by ~ +28 Da, corresponding to a dimethylation event. Previous MS/MS analysis had identified Lys82 as the site of both methylation events. The lack of observation of a corroborating dimethylated GATGLGLK peptide at an expected m/z of 744 is likely due to a steric hindrance of trypsin cleavage posed by dimethylation at Lys82. The dimethylated peptide was on average more than 25 times lower in intensity compared to

its monomethylated isoform (based on 4 observations), suggesting that dimethylation is a relatively rare event. The dimethylated isoform may simply not be common due to the efficiency of the yet unidentified cognate L7/12 methyltransferase. Alternatively dimethylation may be dependent on growth conditions or growth stage. EF-Tu, which specifically interacts with the C-terminal domain of L7/12, exhibits just such a growth phase dependent variation in methylation with Lys56 being monomethylated during logarithmic growth and shifting to being primarily dimethylated during stationary phase[27].

3.3.2 Temperature-dependent Variation in L7/12 Methylation

In light of the earlier work by Chang and our discovery of novel methylation sites in L7/12, we sought to determine how L7/12 methylation varied in response to differing growth temperatures. A comparative MS analysis of L7/12 protein derived from stationary phase *E. coli* cultures grown at 28°C and 37°C was undertaken to determine whether the variable methylation observed in our intact mass measurements was influenced by temperature and whether the previously observed increase in methylation is due to increased methylation at Lys82 or one of the secondary sites identified in our studies. Ribosomal protein from *E. coli* cultures grown in LB media at 28°C and 35°C were isolated and purified as described in the methods. Both L7 and L12 protein isolated from cells grown at each temperature were observed as primarily unmethylated and monomethylated isoforms. We estimated the relative abundances of unmethylated and monomethylated L7 and L12 protein under each growth temperature by calculating the area under the peaks, one means of estimating ratios when absolute quantitation is not

possible[37]. This comparison indicated an approximate 5-fold increase in the ratio of monomethylated L7/12 protein derived from cells cultured at 28°C, corroborating the previous observation (figure 3.4). A similar analysis of Lys82-containing peptides derived from proteins at both temperatures indicated a slightly greater than 3-fold increase in monomethylation at 28°C (see table 3.2).

3.3.3 L7/12 Methyltransferase Screen.

Unlike the L7/12 N-acetyltransferase, the cognate methyltransferase has not been identified. Both our and previous studies suggest that Lys82 is the most prevalent site of monomethylation in the protein. Identification of the cognate methyltransferase would facilitate studies aimed at elucidation of the functional importance of L7/12 methylation. In an attempt to identify potential L7/12 methyltransferase(s) we conducted a rapid MS-based screen of intact L7/12 protein isolated from a panel of single-gene knockout *E. coli* mutants lacking either putative or previously confirmed methyltransferases looking specifically for significant defects in protein monomethylation. To identify putative methyltransferases we used NCBI's PSI-BLAST algorithm, as well as HHpred[28], a structure and homology search/prediction tool, to probe the protein databank (PDB) and NCBI's non-redundant protein data bank for *E. coli* proteins with structural homology to two characterized ribosomal protein methyltransferases - PrmA and PrmB. This list consisted of uncharacterized and putative methytransferases such as SmtA, YafE and YhiQ, as well as a number previously characterized methyltransferases including PrmA and PrmB. A list of mutants examined can be found in table 3.3. Cells harboring the selected gene deletions were sourced from the Japanese National Institute of Genetics'

Keio Collection, a collection of systematic single-gene *E. coli* knockouts[29]. We observed neither a complete loss of, nor severe defects in L7/12 protein monomethylation in any of the mutants tested.

With the identity of the L7/12 methyltransferase unidentified, we sought some insight into the timing of L7/12 methylation. To accomplish this we constructed and expressed a tagged version of L7/12 using the pQE-30 UA vector system, containing a 6x His-tag at the N-terminus. The N-terminus of L7/12 interacts with ribosomal protein L10 and the tag was chosen in an attempt to disrupt L7/12 incorporation into the ribosome. MS analysis of ribosomes derived from cells carrying the expressing the tagged L7/12 indicated that the tagged version of the protein was precluded from the ribosome. The tagged L7/12 was isolated from the cytoplasmic lysate by affinity purification using a Ni-resin column (Promega HisLink Resin.) Intact MS analysis of the tagged protein indicated that ribosome-free L7/12 can be monomethylated. While this observation gives little insight into the actual function of the modification, pools of variably methylated L7/12 protein could be a possible mechanism through which any extraribosomal functions of the protein, such as regulating the translation of its own mRNA, may be modulated.

3.4 Discussion

The goal of our study was to confirm predictions that variable protein methylation was responsible for multiple protein isoforms of L7/12 observed in a previous MS analysis of ribosomal proteins. Our results show that the modification of L7/12 is indeed more complex than has previously been thought, with multiple isoforms due to variable

methylation being present. Using a combination of intact protein and peptide MS analyses we identified at least four distinct sites of methylation, three of which had not been previously reported. All of the methylation events we observed were isolated to the C-terminal region of the protein. Interestingly the novel methylation sites identified were all Glu residues. The methylation of internal Lys, Gln and Arg residues have been previously identified in ribosomal proteins of various organisms (reviewed by Polevoda and Sherman[1]). Our analysis offers the first indication of Glu methylation in ribosomal proteins to our knowledge. Protein methylation plays a key role in bacterial chemotaxis, where the controlled methylation and demethylation of Glu residues of the methyl-accepting chemotaxis proteins (MCPs) are known to play a critical role in regulating the adaptation to chemotactic stimuli[30]. Whether methylation of sites within L7/12 is permanent or may be similarly removed is unknown, but the methylation of multiple Glu residues, a polar, charged residue, in the C-terminal region of L7/12 could have a very strong impact on the protein's various interactions, and represents a viable mechanism through which L7/12 function could be modulated.

The growth temperature-associated changes in the ratio of monomethylated protein observed in our work and that of others indicates a need for the cell to regulate the modification, which in turn suggests that its functional significance may likely be conditional, similar to the N-terminal variable acetylation. Comparing the increases in the ratio of intact monomethylated protein (~5-fold) and Lys82-containing peptides (~3.4-fold), it is apparent that the increase in protein monomethylation observed at the lower growth temperature may not be solely due to an increase in monomethylation of Lys82, and that monomethylation at secondary sites may increase as well. The mechanism

behind the observed temperature-dependent increase remains unknown. It is possible that increased methyltransferase activity is responsible, possibly due to increased expression of the enzyme under the low temperature conditions. Alternatively differences in the conformation of the highly flexible L7/12 protein at lower temperatures could enhance the methyltransferase affinity or accessibility to Lys82 or other sites of methylation. Although our results indicate that free L7/12 can be methylated independent of ribosome incorporation the exact timing of methylation is not conclusively known. If methylation occurs pre-assembly as suggested by our results, slowed ribosomal assembly kinetics at lower temperatures could allow more time for the as yet unidentified cognate methyltransferase(s) to act on the pool of L7/12 protein thereby reducing the pool of unmethylated isoforms available for inclusion into mature ribosomes, thereby explaining the increase observed at lower growth temperatures.

Intriguingly, some bacteria have been shown to continue to express L7/12 during the adaptation to cold shock[36]. The expression of ribosomal components is highly regulated and that cells would continue the metabolically expensive process of producing one of the ribosomal proteins while decreasing global translation rates suggests that there may be yet undiscovered functions of L7/12 beyond translation. Typically cold-induced proteins (CIPs) function as RNA chaperones, binding to mRNA and preventing the formation of secondary structures. Whether L7/12 functions in a similar manner remains to be investigated, but it would be interesting to analyze whether any significant changes to the methylation state of L7/12 occur under cold shock conditions. An increase in methylation levels might enhance the protein's ability to interact with mRNA and hence function as a chaperone. A decrease in methylation may also be possible if expression of

the yet unidentified cognate methyltransferase is diminished or halted all together under cold shock conditions.

The *E. coli* genes *prmA* and *prmB* have been shown to encode the methyltransferases responsible for modifying ribosomal proteins L11 and L3 respectively. L11, the most extensively methylated ribosomal protein, is tri-methylated at three different residues – two internal lysines and an N-terminal alanine. PrmA has been shown to be responsible for methylating all three positions and is currently the only known lysine methyltransferase in *E. coli*. PrmB is responsible for the mono-methylation of ribosomal protein L3 on a glutamine residue. PrmA and PrmB belong to the class I family of S-adenosyl-L-methionine (AdoMet/SAM)-dependent methyltransferases (MTases) whose varied substrates include small molecules, nucleic acids and proteins. Despite this diverse list of substrates the structure of the catalytic domain of the class I family of methyltransferases is highly conserved. The amino acid sequence similarity among class I MTases however can be as low as 10%, possibly reflecting the high level of diversity in the structure required for the high specificity of methyltransferases across such diverse spectrum of substrates. SAM-dependent lysine methyltransferases have highly degenerate sequences. We attempted to identify the gene(s) encoding the L7/12 methyltransferase(s) by screening a panel of *E. coli* mutants harboring single-gene knockouts of putative methyltransferases for defects in L7/12 methylation. The panel of mutants examined thus far is not comprehensive, making it possible that we have yet to hit the right gene target. Yet our results rule out a number of putative *E. coli* methyltransferases, raising the probability that the cognate L7/12 methyltransferase may be an already characterized enzyme performing a dual role in the cell. However,

contradicting this theory is that protein methyltransferases are generally highly specific for their substrates.

Whether L7/12 is targeted by one or multiple methyltransferases remains to be determined. Ribosomal protein L11 is methylated at different sites and on different residues by a single methyltransferase, setting the precedent that a single protein could indeed be responsible for the variable methylation of L7/12 observed in our analyses. In such a case, the enzyme, while being specific to L7/12, could have differing affinities or physical access to the various methylation sites. The high number of isoforms due to variable methylation and large differences in the frequency at which Lys and Glu residues were observed to be methylated however also offer support for a scenario involving multiple, residue-specific enzymes.

Redundancy in the methylation system, or compensation of the loss of methylation at Lys82 by the upregulation of methylation at a secondary site would not be distinguished in our intact protein screen. The employment of either system though would make L7/12 unique amongst the methylated ribosomal proteins and suggest a heightened importance for L7/12 methylation.

The functional importance of ribosomal protein methylation in general remains somewhat murky with the range of importance ascribed to the methylation of various ribosomal proteins being quite disparate. Undoubtedly this is in no small part due to the complex cooperativity among proteins and various nucleic acid species (rRNA, mRNA and tRNA) involved in translation. This high level of intermolecular crosstalk can make drawing conclusions on the importance of a modification from the study of a single protein non-trivial - akin to studying just one cog in a complex machine. Indeed the

interplay between modifications as well as the conditions under which experiments are conducted may yet prove to be critical in teasing apart the roles of ribosomal protein PTM.

Vanet *et al.* observed little effect on phenotype upon disrupting *E. coli's* three trimethylations of ribosomal protein L11[31]. Such extensive modification of a single protein suggests a functional significance to the cell that remains to be uncovered. A likely scenario is that we have yet to discover the specific physiological conditions under which the function of the modifications are most apparent. Condition-dependent phenotypes associated with protein methylation can be found in work carried out by Lhoest and Colsen which indicated a role for the lysine methylation of *E. coli* ribosomal protein L3 in ribosomal assembly[3]. There, a mutant lacking the single L3 methylation exhibited a cold–sensitive growth phenotype and aberrant ribosome assembly, a phenotype only discernible at lower growth temperatures. A similar temperature-dependent phenotype was also apparent when the acetylation of ribosomal protein S5 was abolished[32]. Growth media-dependent phenotypes have also been observed when the methylation of release factors 1 (RF1) and 2 (RF2) was abolished[33]. The prevailing evidence suggests that temperature is an important player in L7/12 methylation. Our work suggests that the increase in methylation observed in lower growth temperatures is mainly due to increased monomethylation at Lys82 and not a shift in methylation patterns to a secondary site. With evidence that unincorporated L7/12 can be methylated, the increased methylation at the lower temperature could be a result of slower kinetics of ribosomal assembly allowing a larger window in which methyltransferase(s) can act on the pool of L7/12 protein.

Evidence of L7/12 polymethylation can also be found in other bacteria. In a MS analysis of the ribosomes of *Caulobacter cresentus*, Running *et al.* observed a likely dimethylatyed isoform of L7/12 in addition to the primary monomethylated isoform[34]. They were however unable to determine the site of the additional methylation. Strader *et al.* observed intact masses that suggested the L7/12 homolog in *Rhodopseudomonas palustri*s may possibly be trimethylated[35]. Based on relative abundance measurements in our studies and previous MS studies of L7/12 by other labs, Lys82 appears to be the primary site of methylation under typical laboratory growth conditions. Whether there are growth conditions under which the 'secondary' sites identified herein constitute the predominant isoform is unclear. All of the possible methylation sites identified in our analysis were located within the C-terminal domain of L7/12 (figure 3.5), which has a helix-turn-helix motif comprised of residues 69-87, strikingly similar to those found in many DNA-binding regulatory proteins, and may constitute an RNA binding site whose interactions could be modulated by the variable methylations. Lys82 is located within a region of the C-terminal domain that is highly conserved in bacteria (figure 3.6). Helgstrand *et al.* have shown that EF-Tu, EF-G, IF2 and RF3 all interact with this same region of L7/12. How L7/12 methylation affects these interactions remains an interesting question for future research. It is highly possible that the methylations modulate L7/12's interactions with the G-proteins which in turn could have a significant effect on GTP hydrolysis and translation accuracy.

Table 3.1

Peptide (Residue Nos.)	Predicted Sequence Mass (Da)	Experimental Mass (Da)	Error (ppm)	InsPecT PTM / Mascot Ion Score	Relative Intensity Ratio
K.FGVSAAAAVAVAAGPVEAAEE*KT (31-52)	2029.06	2029.06	0.00	3.61 / -	1.19
K.TE*FDVILKA (53-60)	978.55	978.55	0.00	3.61 / 46	1.08
R.GATGLGLK*E (75-82)	730.45	730.41	54.76	- / 59	1.01
R.GATGLGLK*EAK.D (75-85)	1058.62	1058.61	9.45	- / 69	ND
R.GATGLGLK*EAK.D (75-85)	1072.64	1072.60	37.29	0.70 / -	0.05[a]
K.EGVSKDDAE*ALKKA (97-109)	1403.74	1403.72	14.25	3.32 / -	0.21
K.DDAE*ALKKA (102-109)	903.48	903.45	33.20	0.79 / -	1.27

Table 3.1. MALDI MS/MS identification of L7/12 methylation sites. Peptides identified in MS/MS analyses corresponding to masses due to the unmodified, monomethylated (~ +14 Da) and dimethylated (~ +28 Da) sequence. Predicted sequence and experimental masses, along with the error (ppm) are listed. Sites of methylation as identified via manual and software-based spectra analysis are indicated in bold with '*' and '**' indicating monomethylation and dimethylation respectively. Corresponding Mascot MS/MS and InSpecT PTM scores are listed. The relative intensity ratios represent the ratio of intensities of the modified to the unmodified version of the peptide where observed. (ND - not detected; ^ ratio calculated with respect to the monomethylated peptide isoform.)

Table 3.2

Culture Temperature		Detected Peptides	Experimental Mass (Da)	Peak Area	Methylated/Unmethylated Peptide Ratio
28°C	Methylated	GATGLGLK	729.57	179.0	7.2
		GATGLGLKEAK	1057.81	1868.7	
	Unmethylated	GATGLGLK	715.52	120.5	
		AVRGATGLGLK	1041.71	164.0	
35°C	Methylated	GATGLGLK	729.58	520.4	2.1
		GATGLGLKEAK	1057.81	3013.5	
	Unmethylated	GATGLGLK	715.57	1706.1	

Table 3.2. Monomethylated to unmethylated ratios of L7/12 peptides under different growth temperatures. MS analysis of observed L7/12 peptides encompassing Lys82 derived from cells cultured at 28°C and 35°C indicated a more than 3-fold increase (7.2 ÷ 2.1) in the ratio of methylated to unmethylated protein in cells cultured at the lower growth temperature.

Table 3.3

Gene	Keio ID	Gene Product Description
prmA	JW3227-2	Ribosomal protein L11 methyltransferase
prmB	JW5841-1	Ribosomal protein L3 N(5)-glutamine methyltransferase
smtA	JW0904-5	Predicted S-adenosylmethionine dependent methyltransferase
yafE	JW0200-1	Predicted S-adenosylmethionine dependent methyltransferase
yafS	JW0203-1	Predicted S-adenosylmethionine dependent methyltransferase
cmoA/yecO	JW1859-1	Predicted methyltransferase
cmoB/yecP	JW1860-1	Predicted S-adenosylmethionine dependent methyltransferase
yccW/rlmI	JW5898-1	23S rRNA methyltransferase monomer
yfiC	JW2559-1	Predicted S-adenosylmethionine dependent methyltransferase
yfiF	JW2565-1	Predicted methyltransferase
yhiQ	JW5672-1	Predicted S-adenosylmethionine dependent methyltransferase
tehB	JW1426-2	Tellurite resistance protein
rsmD/yhhF	JW3430-4	16S rRNA methyltransferase
rsmF/yebU	JW5301-1	16S rRNA methyltransferase
rlmL/ycbY	JW0931-1	23S rRNA methyltransferase
rlmG/ygjO	JW5513-1	23S rRNA methyltransferase
rlmB/yjfH	JW4138-1	23S rRNA methyltransferase monomer
ygdE	JW2777-1	Predicted methyltransferase
ymfD	JW1123-1	Predicted SAM-dependent methyltransferase - e14 prophage
ygdL	JW2783-2	Conserved protein
mraW/yabC	JW0080-1	S-adenosyl-dependent methyltrnasferase
yjhP	JW4268-1	Predicted methyltransferase, KpLE2 phage-like element
bioC	JW0760-2	Predicted methyltransferase, enzyme of biosynthesis
ybcY	JW0551-1	Predicted SAM-dependent methyltransferase
rumA/rlmD	JW2756-1	23S rRNA methyltransferase
gatD	JW2075	Galacticol-1-phosphate dehydrogenase
ybeY	JW0656	Conserved protein - shows similarity to metal-dependent hydrolases
ksgA/rsmA	JW0050	16S rRNA dimethyltransferase

Table 3.3. Putative and characterized methyltransferases identified in structural homology searches to ribosomal protein methyltransferases PrmA and PrmB. Intact L7/12 protein isolated from single-gene *E. coli* knockout mutants were examined by MS for loss of methylation in an attempt to identify the cognate methyltransferase.

Figure 3.1

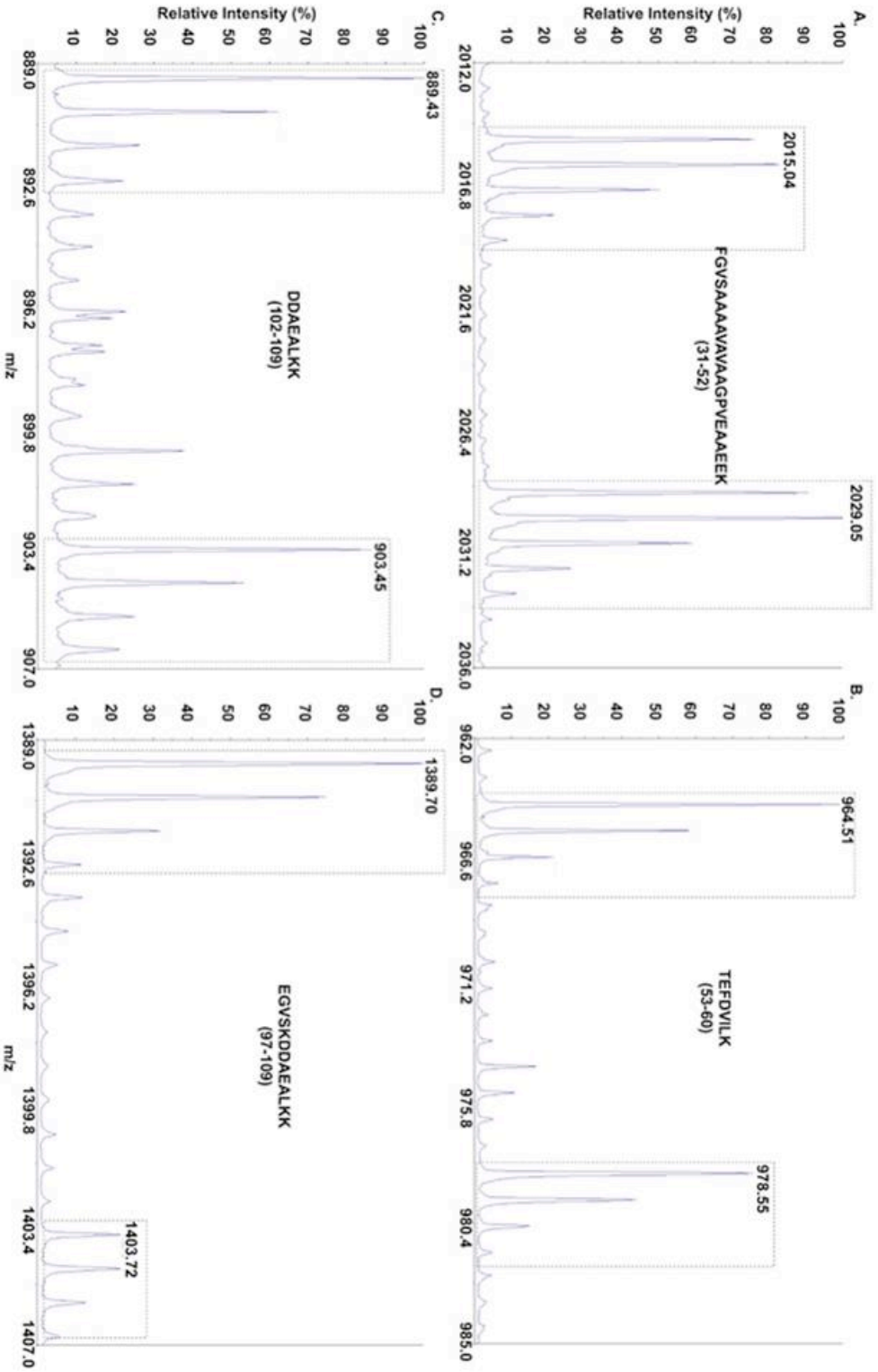
A.
Relative Intensity (%)
2015.04
2029.05
FGVSAAAAVAVAAGPVEAAEEK
(31-52)
2012.0
2016.8
2021.6
2026.4
2031.2
2036.0
B.
964.51
978.55
TEFDVILK
(53-60)
962.0
966.6
971.2
975.8
980.4
985.0
C.
889.43
903.45
DDAEALKK
(102-109)
889.0
892.6
896.2
899.8
903.4
907.0
m/z
D.
1389.70
1403.72
EGVSKDDAEALKK
(97-109)
1389.0
1392.6
1396.2
1399.8
1403.4
1407.0
m/z

Figure 3.1 A-D. Raw MALDI-TOF MS spectra of L7/12 peptides. The isotopic distribution of unmethylated and monomethylated states of 4 peptides, encompassing a minimum of 3 distinct methylation sites, are illustrated in panels A, B, C and D. The mass of the monoisotopic peak of both the unmethylated and monomethylated isoforms are labelled. Peptide amino acid sequences (residue nos.) were identified through MS/MS analysis of the indicated peptide peaks.

Figure 3.2

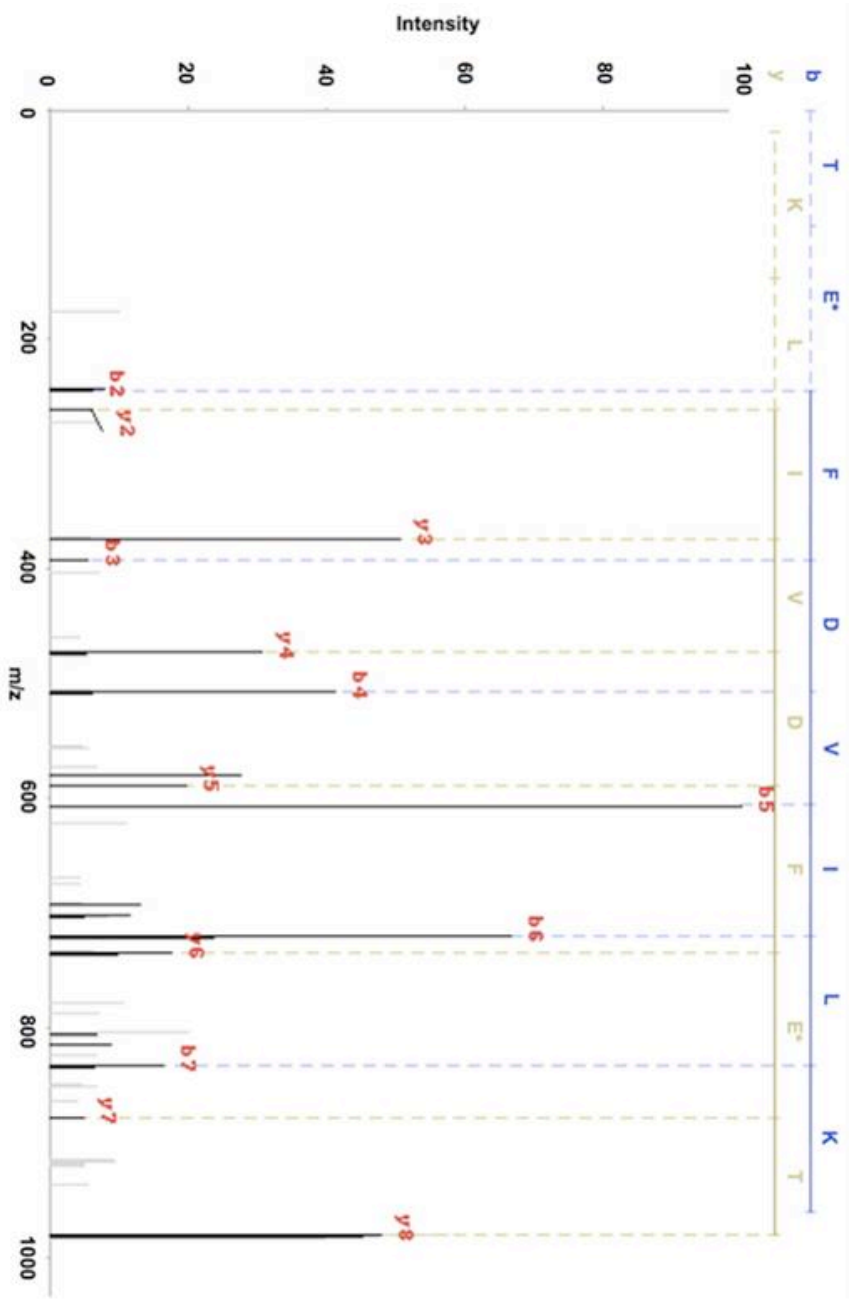
Intensity
0
20
40
60
80
100
m/z
0
200
400
600
800
1000
b
y
T
K
L
I
V
D
F
E*
L
K
T
b2
y2
y3
b3
y4
b4
y5
b5
b6
y6
b7
y7
y8

Figure 3.2. Novel Glu methylation of ribosomal protein L7/12. InsPecT PTM analysis of MALDI TOF/TOF MS/MS spectra derived from the analysis of the L7/12 tryptic peptide 53-TEFDVILK-60, observed to be shifted by ~+14 Da in multiple MS spectra, identifies Glu-54 (denoted "E*") as a novel site of monomethylation in L7/12.

Figure 3.3

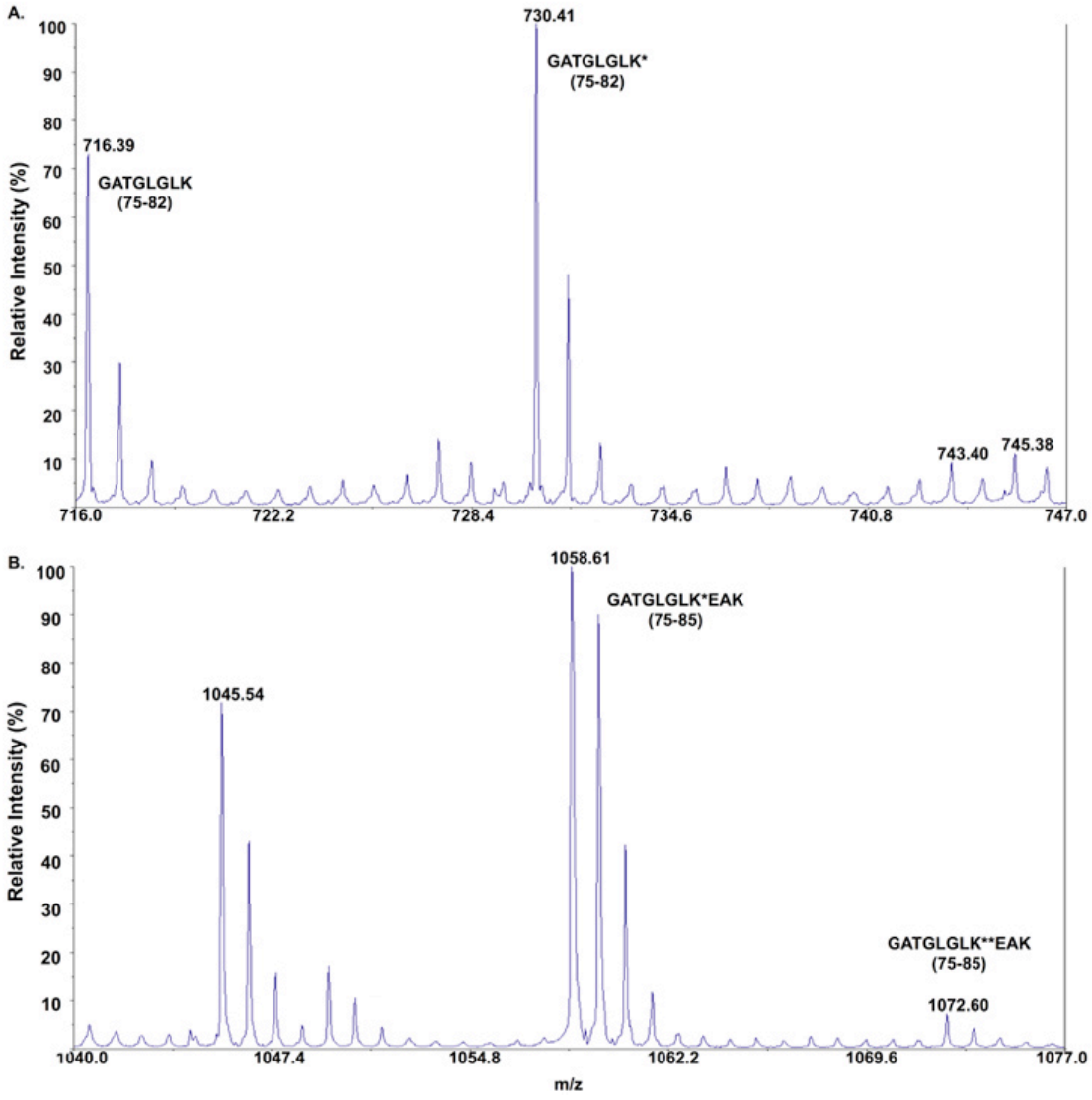
A.
730.41
GATGLGLK*
(75-82)
716.39
GATGLGLK
(75-82)
743.40
745.38
Relative Intensity (%)
716.0
722.2
728.4
734.6
740.8
747.0
B.
1058.61
GATGLGLK*EAK
(75-85)
1045.54
GATGLGLK**EAK
(75-85)
1072.60
Relative Intensity (%)
1040.0
1047.4
1054.8
1062.2
1069.6
1077.0
m/z

Figure 3.3. MALDI-TOF MS spectra of peptides containing the variably methylated Lys82. A. The tryptic peptide GATGLGLK (76-82) was observed in both a unmethylated (m/z 716.39) and monomethylated (m/z 730.41) states but not as being dimethylated. B. The peptide GATGLGLKEAK (76-85), the result of a single missed tryptic cleavage, was additionally observed to be dimethylated (m/z 1072.6). Previous MS/MS analysis identified Lys82 as the site of the dimethylation event. The lack of observation of a dimethylated GATGLGLK (76-82) peptide is likely due to inhibition of trypsin cleavage at Lys82 when the peptide is dimethylated.

Figure 3.4

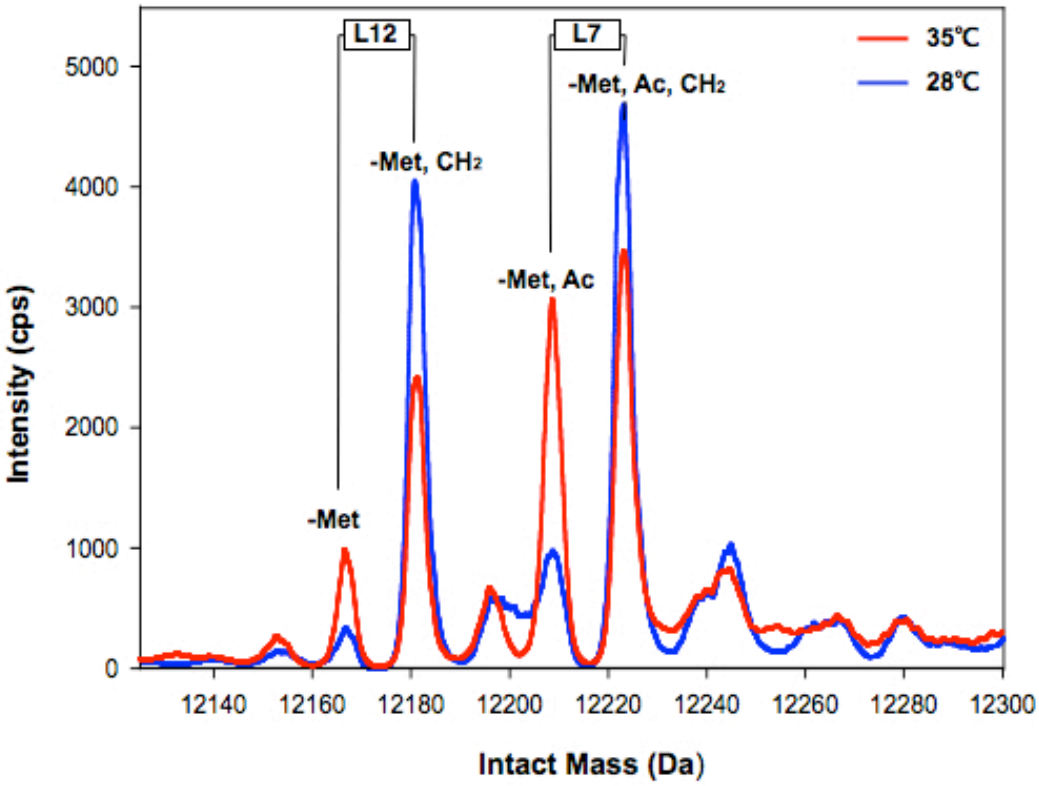
L12
L7
35°C
28°C
-Met, Ac, CH2
-Met, CH2
-Met, Ac
-Met
Intensity (cps)
5000
4000
3000
2000
1000
0
12140
12160
12180
12200
12220
12240
12260
12280
12300
Intact Mass (Da)

Figure 3.4. Deconvoluted intact mass spectra of L7/12 from cells cultured at 35°C and 28°C. Intact protein MS analysis of L7/12 showed an approximate 5-fold increase in the ratio of monomethylated protein isoforms of L7/12 at the lower growth temperature of 28°C (blue) compared to cells cultured at 35°C (red).

Figure 3.5

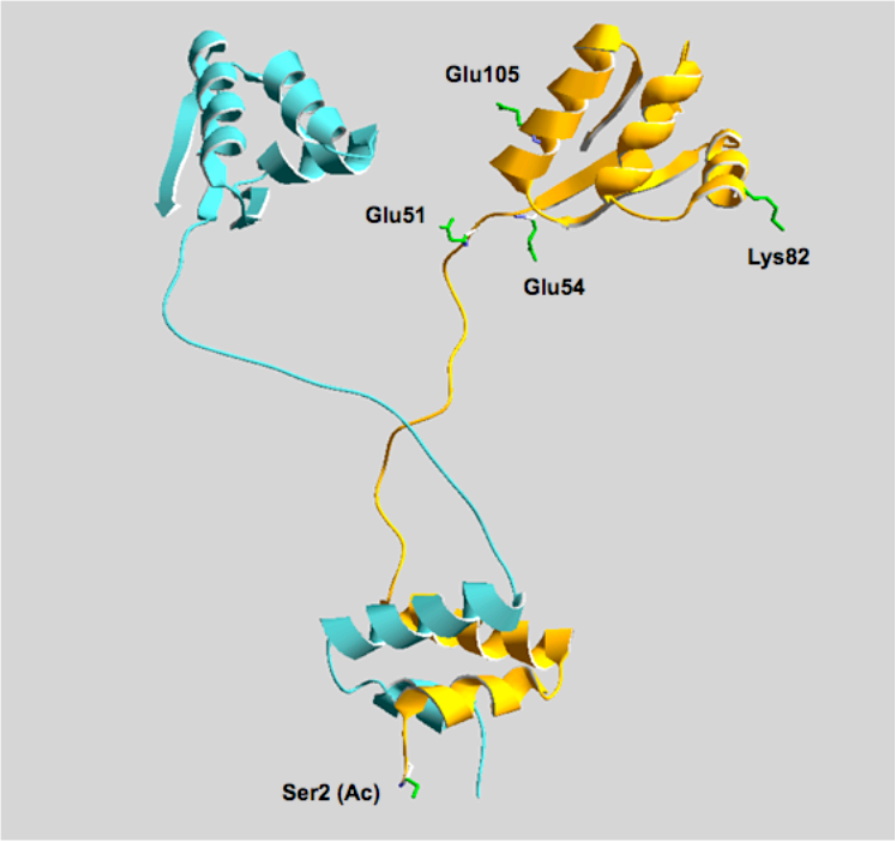

Figure 3.5. Sites of modification mapped onto the structure of the L7/L12 dimer. The The DeepView / Swiss-PdbViewer (v 4.01)[34] was used to display of the structure for the L7/12 dimer (PDB: 1RQU, J Biol Chem 2004). The dimmer is comprised of two distinct, organized domains connected by an extensive linker or "hinge" region. The N-terminal domain is responsible for L7/L12 dimerization and for anchoring the protein to the ribosome, through interactions with L10. The C-terminal domain functions as a binding site for translation factors. The residues indicated in green were identified as being either mono-methylated (K82, Glu51, Glu54, Glu105) or di-methylated (K82), with the exception of Ser2 which is acetylated (Ac) in L7 but not in L12.

Figure 3.6

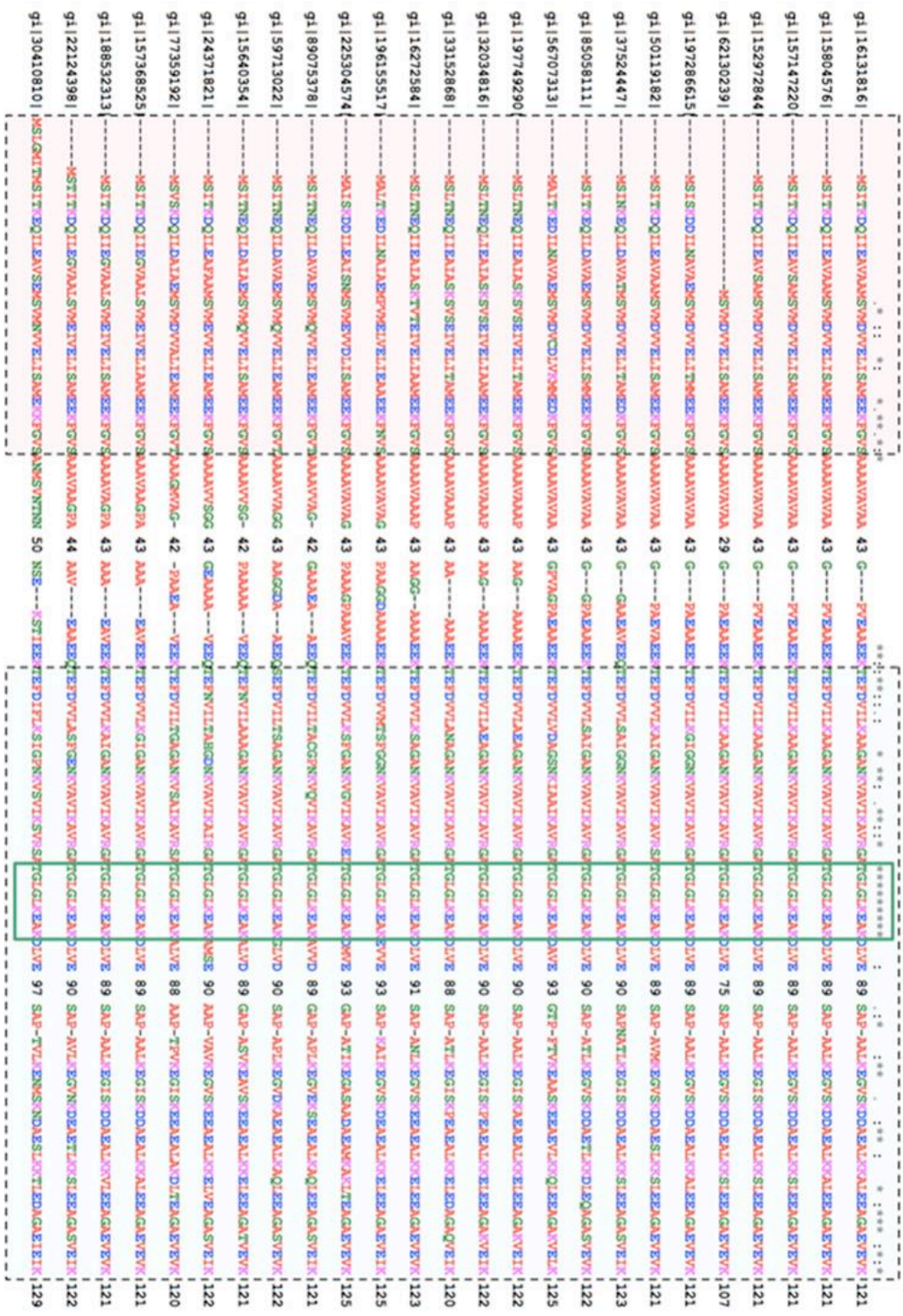

Figure 3.6. Conservation of bacterial L7/12 protein sequence. The L7/12 protein sequence is highly conserved in bacteria. All identified sites of methylation in our study were located within the C-terminal domain of L7/12. Lys82, the primary site of methylation and the only observed site of di-methylation in our study, is contained within a highly conserved stretch of amino acids in the C-terminal domain (green box) that has been shown to interact with multiple initiation and elongation factors.

References

1. Polevoda, B.; Sherman, F., Methylation of proteins involved in translation. *Mol Microbiol* **2007,** 65, (3), 590-606.

2. Gary, J. D.; Clarke, S., RNA and protein interactions modulated by protein arginine methylation. *Prog Nucleic Acid Res Mol Biol* **1998,** 61, 65-131.

3. Lhoest, J.; Colson, C., Genetics of ribosomal protein methylation in *Escherichia coli*. II. A mutant lacking a new type of methylated amino acid, N5-methylglutamine, in protein L3. *Mol Gen Genet* **1977,** 154, (2), 175-80.

4. Robert, F.; Brakier-Gingras, L., A functional interaction between ribosomal proteins S7 and S11 within the bacterial ribosome. *J Biol Chem* **2003,** 278, (45), 44913-20.

5. Agrawal, R. K.; Linde, J.; Sengupta, J.; Nierhaus, K. H.; Frank, J., Localization of L11 protein on the ribosome and elucidation of its involvement in EF-G-dependent translocation. *J Mol Biol* **2001,** 311, (4), 777-87.

6. Li, W.; Sengupta, J.; Rath, B. K.; Frank, J., Functional conformations of the L11-ribosomal RNA complex revealed by correlative analysis of cryo-EM and molecular dynamics simulations. *Rna* **2006,** 12, (7), 1240-53.

7. Stark, H.; Rodnina, M. V.; Rinke-Appel, J.; Brimacombe, R.; Wintermeyer, W.; van Heel, M., Visualization of elongation factor Tu on the *Escherichia coli* ribosome. *Nature* **1997,** 389, (6649), 403-6.

8. Diaconu, M.; Kothe, U.; Schlunzen, F.; Fischer, N.; Harms, J. M.; Tonevitsky, A. G.; Stark, H.; Rodnina, M. V.; Wahl, M. C., Structural basis for the function of the ribosomal L7/12 stalk in factor binding and GTPase activation. *Cell* **2005,** 121, (7), 991-1004.

9. Helgstrand, M.; Mandava, C. S.; Mulder, F. A.; Liljas, A.; Sanyal, S.; Akke, M., The ribosomal stalk binds to translation factors IF2, EF-Tu, EF-G and RF3 via a conserved region of the L12 C-terminal domain. *J Mol Biol* **2007,** 365, (2), 468-79.

10. Wahl, M. C.; Moller, W., Structure and function of the acidic ribosomal stalk proteins. *Curr Protein Pept Sci* **2002,** 3, (1), 93-106.

11. Ben-Bassat, A.; Bauer, K.; Chang, S. Y.; Myambo, K.; Boosman, A.; Chang, S., Processing of the initiation methionine from proteins: properties of the *Escherichia coli* methionine aminopeptidase and its gene structure. *J Bacteriol* **1987,** 169, (2), 751-7.

12. Brot, N.; Marcel, R.; Cupp, L.; Weissbach, H., The enzymatic acetylation of ribosomal bound protein L 12. *Arch Biochem Biophys* **1973,** 155, (2), 475-7.

13. Frottin, F.; Martinez, A.; Peynot, P.; Mitra, S.; Holz, R. C.; Giglione, C.; Meinnel, T., The proteomics of N-terminal methionine cleavage. *Mol Cell Proteomics* **2006,** 5, (12), 2336-49.

14. Hirel, P. H.; Schmitter, M. J.; Dessen, P.; Fayat, G.; Blanquet, S., Extent of N-terminal methionine excision from *Escherichia coli* proteins is governed by the side-chain length of the penultimate amino acid. *Proc Natl Acad Sci U S A* **1989,** 86, (21), 8247-51.

15. Polevoda, B.; Sherman, F., N-terminal acetyltransferases and sequence requirements for N-terminal acetylation of eukaryotic proteins. *J Mol Biol* **2003,** 325, (4), 595-622.

16. Deusser, E.; Wittmann, H. G., Ribosomal proteins: variation of the protein composition in *Escherichia coli* ribosomes as function of growth rate. *Nature* **1972,** 238, (5362), 269-70.

17. Ramagopal, S.; Subramanian, A. R., Alteration in the acetylation level of ribosomal protein L12 during growth cycle of *Escherichia coli*. *Proc Natl Acad Sci U S A* **1974,** 71, (5), 2136-40.

18. Gordiyenko, Y.; Deroo, S.; Zhou, M.; Videler, H.; Robinson, C. V., Acetylation of L12 increases interactions in the *Escherichia coli* ribosomal stalk complex. *J Mol Biol* **2008,** 380, (2), 404-14.

19. Chang, F. N., Temperature-dependent variation in the extent of methylation of ribosomal proteins L7 and L12 in *Escherichia coli*. *J Bacteriol* **1978,** 135, (3), 1165-6.

20. Chang, F. N.; Chang, C. N.; Paik, W. K., Methylation of ribosomal proteins in *Escherichia coli*. *J Bacteriol* **1974,** 120, (2), 651-6.

21. Chang, C. N.; Chang, N., Methylation of the ribosomal proteins in *Escherichia coli*. Nature and stoichiometry of the methylated amino acids in 50S ribosomal proteins. *Biochemistry* **1975,** 14, (3), 468-77.

22. Arnold, R. J.; Reilly, J. P., Analysis of methylation and acetylation in *E. coli* ribosomal proteins. *Methods Mol Biol* **2002,** 194, 205-10.

23. Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S., Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **1999,** 20, (18), 3551-67.

24. Giddings, M. C.; Shah, A. A.; Gesteland, R.; Moore, B., Genome-based peptide fingerprint scanning. *Proc Natl Acad Sci U S A* **2003,** 100, (1), 20-5.

25. Tanner, S.; Shu, H.; Frank, A.; Wang, L. C.; Zandi, E.; Mumby, M.; Pevzner, P. A.; Bafna, V., InsPecT: identification of posttranslationally modified peptides from tandem mass spectra. *Anal Chem* **2005,** 77, (14), 4626-39.

26. Liu, X.; Reilly, J. P., Correlating the chemical modification of *Escherichia coli* ribosomal proteins with crystal structure data. *J Proteome Res* **2009,** 8, (10), 4466-78.

27. Van Noort, J. M.; Kraal, B.; Sinjorgo, K. M.; Persoon, N. L.; Johanns, E. S.; Bosch, L., Methylation in vivo of elongation factor EF-Tu at lysine-56 decreases the rate of tRNA-dependent GTP hydrolysis. *Eur J Biochem* **1986,** 160, (3), 557-61.

28. Soding, J., Protein homology detection by HMM-HMM comparison. *Bioinformatics* **2005,** 21, (7), 951-60.

29. Baba, T.; Ara, T.; Hasegawa, M.; Takai, Y.; Okumura, Y.; Baba, M.; Datsenko, K. A.; Tomita, M.; Wanner, B. L.; Mori, H., Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2006,** 2, 2006 0008.

30. Springer, M. S.; Zanolari, B.; Pierzchala, P. A., Ordered methylation of the methyl-accepting chemotaxis proteins of *Escherichia coli*. *J Biol Chem* **1982,** 257, (12), 6861-6.

31. Vanet, A.; Plumbridge, J. A.; Guerin, M. F.; Alix, J. H., Ribosomal protein methylation in *Escherichia coli*: the gene prmA, encoding the ribosomal protein L11 methyltransferase, is dispensable. *Mol Microbiol* **1994,** 14, (5), 947-58.

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

33. Mora, L.; Heurgue-Hamard, V.; de Zamaroczy, M.; Kervestin, S.; Buckingham, R. H., Methylation of bacterial release factors RF1 and RF2 is required for normal translation termination in vivo. *J Biol Chem* **2007,** 282, (49), 35638-45.

34. Running, W. E.; Ravipaty, S.; Karty, J. A.; Reilly, J. P., A top-down/bottom-up study of the ribosomal proteins of Caulobacter crescentus. *J Proteome Res* **2007,** 6, (1), 337-47.

35. Strader, M. B.; Verberkmoes, N. C.; Tabb, D. L.; Connelly, H. M.; Barton, J. W.; Bruce, B. D.; Pelletier, D. A.; Davison, B. H.; Hettich, R. L.; Larimer, F. W.; Hurst, G. B., Characterization of the 70S Ribosome from *Rhodopseudomonas palustris* using an integrated "top-down" and "bottom-up" mass spectrometric approach. *J Proteome Res* **2004,** 3, (5), 965-78.

36. Graumann, P.; Schroder, K.; Schmid, R.; Marahiel, M. A., Cold shock stress-induced proteins in Bacillus subtilis. J Bacteriol **1996,** 178, (15), 4611-9.

37. Old, W.M.; Meyer-Arendt, K.; Aveline-Wolf, L.; Pierce, K. G.; Mendoza, A.; Sevinsky, J. R.; Resing, K. A.; Ahn, N. G., Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Mol Cell Proteomics* **2005,** 4, (10), 1487-502.

Chapter 4

Conclusions

4.1 Summary

Amongst the ultimate goals of proteomics research is the complete understanding of cells through the characterization of the complement of proteins expressed, their dynamics, interactions and functions. In bacteria, protein post-translational modification is a near ubiquitous process. Mass spectrometry (MS) has become one of the most important tools within the proteomics community because of its ability to achieve high accuracy, sensitivity and resolution protein analysis all in a high-throughput fashion. It is a versatile analytical tool that, when utilized to its full potential, can be employed not only to identify proteins, but can also aid in their quantitation and structural elucidation, identify interaction partners and characterization any post-translational modifications (PTMs). Bottom-up MS, the analysis of peptides by MS and MS/MS, allows for unambiguous protein identification and can provide structural information on amino acid composition as well as the numbers, types and locations of any modifications present. While bottom-up MS analysis coupled with peptide mass fingerprinting (PMF) is currently the standard forbearer of protein identification in proteomics it has significant limitations towards full protein characterization. These stem particularly from the lack of information gained on the presence of protein isoforms and the possibility of missed

PTMs in unobserved peptides. Top-down MS analysis, in which mass measurements are directly made of intact proteins, is emerging as the preferred method for protein characterization. While high accuracy mass measurements of intact proteins is sometimes sufficient to infer identity and discern protein isoforms due to PTMs, the technique is not ideal for unambiguous protein identification, or for that matter isoform characterization, especially when dealing with complex mixtures. This may be simply due to the potential presence of proteins isobaric in mass, or the huge combinatorics of post-translational modifications that can result in any observed experimental mass as PTMs can either increase or decrease the mass of the mature protein.

The work described in this dissertation illustrates that the most critical limitations of each method towards protein characterization can be readily addressed when bottom-up peptide analyses are combined with complementary MS analyses of the intact proteins using available instrumentation and informatics tools. Our research was primarily focused on leveraging the strengths of the two main schools of mass spectrometry - top-down MS and bottom-up MS - in a synergistic workflow towards the comprehensive analysis of protein isoform heterogeneity in *Escherichia coli* ribosomal proteins. Our analysis put us in the company of only a few other research groups that are integrating top-down and bottom-up MS towards the characterization of proteins and protein complexes. In so doing we have clearly illustrated that currently available MS strategies, when combined with an efficient protein separation, can function to identify and characterize protein isoforms in complex samples due to PTMs and significantly, indicate potentially important genomic mutations through the identification of amino acid substitutions.

In a global characterization of ribosomal protein heterogeneity discussed in chapter 2, ribosomal protein derived from three *E. coli* derivatives were analyzed by top-down and bottom-up MS methods. Significant impetus for the global analysis of protein heterogeneity stemmed from the thought that variable PTM of ribosomal proteins in bacteria was a possible mechanism through which the ribosome filter hypothesis might be realized. MS studies, using either one or both of top-down and bottom-up MS methods, have been previously carried out on bacterial ribosomes[1-8]. While this study is not the first, the comprehensive approach to protein characterization that was undertaken and discussed herein enabled the identification of a level of heterogeneity that has previously gone unreported. This study was unique for the 6-fold coverage achieved for most proteins which facilitated identification of conserved isoforms between the strains and previously unidentified PTMs. Our observations show that variable PTM of proteins is an avenue through which ribosomal protein function could be modulated. While the specific functions of the PTMs identified in our studies and those of others remains unknown, the potential impact of the multitude of PTMs and isoforms observed in this study, particularly on the function of the finely tuned translation machinery, cannot be overlooked. The variable modifications represent a mechanism wherein the ribosome filter hypothesis could be realized and could also impact any extra-ribosomal functions of the proteins. That such heterogeneity was observed under a single growth condition in bacteria with the ability to rapidly adapt to environmental changes suggest that more growth conditions need to be examined to get a full picture of the states in which the ribosome may exist.

This study clearly illustrates the power of such combined approaches to identify

not only PTMs, but potentially important genomic mutations. While genome sequencing is no longer a time-limiting factor in bacterial characterization, not all mutations result in the expression of altered proteins. MS has the advantage over sequencing of simultaneously providing information on potentially critical parameters of protein expression changes and potential changes to protein interactions and structure without the need for the application of additional techniques such as mRNA expression analyses, which in and of itself is not always indicative of protein levels. Before such uses of MS become widely accepted, advances are necessary that improve the typical protein sequence coverage, which in a typical large-scale bottom-up experiment is less than 60%. Improvements in separation technologies will also serve to advance the application of top-down proteomics. While noteworthy in its characterization of different protein isoforms both within and amongst the bacterial strains analyzed, our heterogeneity analysis also underlined a major hurdle facing protein characterization using MS. Whilst this study was completed using currently available informatics tools, our data analysis pipeline illustrated a desperate need for the development of informatics tools that can integrate and interpret top-down and bottom-up datasets in an automated fashion. Informatics tools for the processing and interpretation of bottom-up MS data have largely kept pace with the rapid developments in instrumentation an increasing scales and complexity of experiments. However, the development of informatics tools for the interpretation of top-down MS data have lagged significantly. Our in-house developed PROCLAME software provided all of the PTM predictions carried out in this study and were manually checked against the bottom-up MS data. The Protein Inference Engine (PIE), developed in the Giddings lab, and PTMSearchPlus[9] are the only tools to

our knowledge that currently attempt to overcome the daunting challenge of integrating bottom-up and top-down MS data. The PIE uses Markov chain Monte Carlo (MCMC) methods to integrate top-down and bottom-up mass spectrometry data, drawing on multiple sources of biological information (such as residue-specific modification frequencies) to predict PTMs and characterize proteins from experimental MS data.

One of the more significant challenges that arose during our analyses was that of how to correctly distinguish low-abundant protein isoforms, which typically produced mass spectra low signal-to-noise ratios (S/N), from peaks due to instrument and chemical noise. While hard S/N cutoffs can result in significantly cleaner and by extension, easier to interpret spectral data, it was evident from this study that there is significant risk for meaningful data to be lost. While we sought to overcome this hurdle by instituting multiple data filter criteria, the need for more efficient and less subjective methods became quickly evident. Statistical analysis tools that could be applied across multiple analyses to calculate likelyhood metrics on whether a given peak is due to random noise or genuine signal would be extremely useful as proteomics matures from a field focused more on protein identification to one more focused on protein characterization.

In chapter 3 we demonstrated that L7/12 is a target for variable protein methylation. In a focused analysis of the variable methylation of ribosomal protein L7/12 that was undertaken subsequent to the identification of L7/12 as one of the more varible proteins in our global heterogeneity analysis. We identified a number of novel sites of methylation inclusive of a novel dimethylation event. The localization of all of the methylations sites we identified to the C-terminus of the protein suggest a possible role for methylation in regulation of L7/12's interactions with either one or multiple

translation factors, which in turn could have impacts on translation initiation, translation fidelity, and ribosome translocation rates. The identifications of these sites directly illustrate the aforementioned benefits that better front-end separations and increased MS sensitivity would achieve as many of the sites of modification were not characterized in the global analysis simply due to the lack of observation of the modified peptides in the digests of the complex mixture of ribosomal proteins. Similar focused analyses of other ribosomal proteins remain necessary to confirm and identify many modification sites predicted in our global top-down/bottom-up analyses.

Protein methylation has also been shown to modulate protein-nucleic acid interactions raising questions as to whether the purpose of the methylated residues could be to regulate mRNA translation - either enhancing expression of certain mRNA species by stabilizing them or hindering their expression thorough binding at ribosome entry sites. In fact our observation that incorporation into ribosomes is not required for L7/12 methylation suggests that free pools of variably modified L7/12 are free to perform extra-ribosomal functions in the cell. Our experiments probing the regulation of L7/12 methylation indicated that the levels of methylated protein increased at lower growth temperatures. This is possibly due to the slower kinetics of ribosome assembly at lower temperatures. While this analysis was focused on L7/12 derived from active ribosomes we expect from the aforementioned results that free L7/12 would exhibit a similar increase.

Improved mass spectrometry and separation techniques are continually being developed. It bodes well for the future of high-throughput protein characterization that the development of informatics tools to speed the integration and interpretation of top-

down and bottom-up datasets is underway in a few labs including ours. As these advancements become mainstream the new bottleneck to bacterial characterization may indeed become functional elucidation of protein isoforms and modifications identified in large scale proteomics experiments. However, as the successes of the genomics revolution gave rise to the proteomics era, continued improvements in high-throughput protein characterization will no doubt drive innovation in realm of functional proteomics.

References

1. Benjamin, D. R.; Robinson, C. V.; Hendrick, J. P.; Hartl, F. U.; Dobson, C. M., Mass spectrometry of ribosomes and ribosomal subunits. Proc Natl Acad Sci U S A 1998, 95, (13), 7391-5.

2. Arnold, R. J.; Reilly, J. P., Observation of *Escherichia coli* ribosomal proteins and their posttranslational modifications by mass spectrometry. Anal Biochem 1999, 269, (1), 105-12.

3. Arnold, R. J.; Reilly, J. P., Analysis of methylation and acetylation in *E. coli* ribosomal proteins. Methods Mol Biol 2002, 194, 205-10.

4. Strader, M. B.; Verberkmoes, N. C.; Tabb, D. L.; Connelly, H. M.; Barton, J. W. Bruce, B. D.; Pelletier, D. A.; Davison, B. H.; Hettich, R. L.; Larimer, F. W.; Hurst, G. B., Characterization of the 70S Ribosome from *Rhodopseudomonas palustris* using an integrated "top-down" and "bottom-up" mass spectrometric approach. J Proteome Res 2004, 3, (5), 965-78.

5. Running, W. E.; Ravipaty, S.; Karty, J. A.; Reilly, J. P., A top-down/bottom-up study of the ribosomal proteins of Caulobacter crescentus. J Proteome Res 2007, 6, (1), 337-47.

6. Iost, I.; Charollais, J.; Vinh, J.; Pflieger, D., Characterization of *E. coli* ribosomal particles : combined analysis of whole proteins by mass spectrometry and of proteolytic digests by liquid chromatography-tandem mass spectrometry. Methods Mol Biol 2008, 432, 321-41.

7. Lauber, M. A.; Running, W. E.; Reilly, J. P., *B. subtilis* ribosomal proteins: structural homology and post-translational modifications. J Proteome Res 2009, 8, (9), 4193-206.

8. Liu, X.; Reilly, J. P., Correlating the chemical modification of *Escherichia coli* ribosomal proteins with crystal structure data. J Proteome Res 2009, 8, (10), 4466-78.

9. Kertesz, V.; Connelly, H. M.; Erickson, B. K.; Hettich, R. L., PTMSearchPlus:

software tool for automated protein identification and post-translational modification characterization by integrating accurate intact protein mass and bottom-up mass spectrometric data searches. Anal Chem 2009, 81, (20), 8387-95.

Appendix A - Summary of Top-down/Bottom-up MS Protein Analyses

Subunit & Sequence-predicted Ave. Mass	Calculated Ave. Mass	Derivative	Experimental Mass	Source	Predicted PTM	Calculated Mass (plus PTM)	Prediction Error (ppm)	*Unique Peptides	^Mascot MS & MS/MS Score
RpsA S1	61158.18	WT	Not Observed					24	853
61158.18	61158.18	SmR	Not Observed					25	781
	61158.18	SmRC	Not Observed					25	770
RpsB S2	26743.73	WT	26612.51	ESI-FT	-MET	26612.53	0.95	9	805
26743.73	26743.73		26613.26	ESI-TOF	-MET	26612.53	-27.24		
	26743.73	SmR	26614.09	ESI-TOF	-MET	26612.53	-58.47	14	319
	26743.73	SmRC	Not Observed					11	705
RpsC S3	25983.30	WT	25852.81	ESI-FT	-MET	25852.10	-27.47	13	692
25983.30	25983.30		25852.41	ESI-TOF	-MET	25852.10	-12.18		
	25983.30	SmR	25852.10	ESI-FT	-MET	25852.10	-0.31	13	582
	25983.30		25852.33	ESI-TOF	-MET	25852.10	-9.05		
	25983.30	SmRC	25852.03	ESI-FT	-MET	25852.10	2.65	10	937
	25983.30		25855.91	ESI-TOF	-MET	25852.10	-147.42		
RpsD S4	23469.15	WT	23337.99	ESI-FT	-MET	23337.95	-1.48	9	714
23469.15	23469.15		23337.80	ESI-TOF	-MET	23337.95	6.43		
	23469.15	SmR	23337.95	ESI-FT	-MET	23337.95	0.25	12	688
	23469.15		23338.41	ESI-TOF	-MET	23337.95	-19.32		

Subunit & Sequence-predicted Ave. Mass	Calculated Ave. Mass	Derivative	Experimental Mass	Source	Predicted PTM	Calculated Mass (plus PTM)	Prediction Error (ppm)	*Unique Peptides	^Mascot MS & MS/MS Score
	23469.15	SmRC	23337.41	ESI-FT	-MET	23337.95	23.27	9	672
RpsE S5	17603.44	WT	17514.22	ESI-FT	-MET, Ac	17514.28	3.12	12	389
17603.44	17603.44		17515.36	ESI-TOF	-MET, Ac	17514.28	-61.70		
	17603.44	SmR	17514.31	ESI-FT	-MET, Ac	17514.28	-1.62	14	281
	17603.44		17514.21	ESI-TOF	-MET, Ac	17514.28	3.96		
	17603.44		17515.55	ESI-TOF	-MET, Ac	17514.28	-72.55		
	17603.44	SmRC	17514.32	ESI-FT	-MET, Ac	17514.28	-2.12	10	408
	17603.44		17516.81	ESI-TOF	-MET, Ac	17514.28	-144.68		
RpsF S6	15187.09	WT	Not Observed					3	89
15187.09	15187.09	SmR	15445.63	ESI-TOF	C-terminal EE	15445.29	-22.21	4	310
	15187.09	SmRC	Not Observed					3	307
RpsG S7	20019.14	WT	19889.29	ESI-TOF	-MET	19887.95	-67.63	13	465
20019.14	20019.14	SmR	19888.05	ESI-TOF	-MET	19887.95	-5.28	15	503
	20019.14		19887.97	ESI-FT	-MET	19887.95	-1.49		
	20019.14	SmRC	19887.90	ESI-FT	-MET	19887.95	2.10	12	492
RpsH S8	14126.61	WT	13995.44	ESI-FT	-MET	13995.41	-2.07	3	237
14126.61	14126.61		13995.98	ESI-TOF	-MET	13995.41	-40.94		

Subunit & Sequence-predicted Ave. Mass	Calculated Ave. Mass	Derivative	Experimental Mass	Source	Predicted PTM	Calculated Mass (plus PTM)	Prediction Error (ppm)	*Unique Peptides	^Mascot MS & MS/MS Score
	14126.61	SmR	13995.31	ESI-FT	-MET	13995.41	7.28	2	146
	14126.61		13995.46	ESI-TOF	-MET	13995.41	-3.64		
	14126.61	SmRC	13995.39	ESI-FT	-MET	13995.41	1.55	4	156
	14126.61		13996.72	ESI-TOF	-MET	13995.41	-93.53		
RpsI S9	14856.25	WT	14725.03	ESI-TOF	-MET	14725.05	1.09	2	237
14856.25	14856.25	SmR	14725.06	ESI-TOF	-MET	14725.05	-0.95	5	196
	14856.25	SmRC	14724.26	ESI-FT	-MET	14725.05	53.63	3	160
RpsJ S10	11735.62	WT	11735.46	ESI-FT	None	11735.62	13.35	6	221
11735.62	11735.62	SmR	11735.50	ESI-FT	None	11735.62	10.37	1	211
	11735.62	SmRC	11735.62	ESI-FT	None	11735.62	-0.51	4	68
RpsK S11	13844.97	WT	13727.34	ESI-TOF	-MET, CH2	13727.80	33.21	3	561
13844.97	13844.97	SmR	13727.75	ESI-TOF	-MET, CH2	13727.80	3.34	6	290
	13844.97	SmRC	13727.66	ESI-FT	-MET, CH2	13727.80	9.90	27	467
	13844.97		13727.42	ESI-TOF	-MET, CH2	13727.80	27.09		
	13844.97		13727.73	ESI-FT	-MET, CH2	13727.80	5.11		
RpsL S12	13737.11	WT	13652.15	ESI-FT	-MET, β-methylthiolation	13652.00	-11.19	1	278
13737.11	13737.11		13649.01	ESI-TOF	-MET, β-methylthiolation	13652.00	219.13		

Subunit & Sequence-predicted Ave. Mass	Calculated Ave. Mass	Derivative	Experimental Mass	Source	Predicted PTM	Calculated Mass (plus PTM)	Prediction Error (ppm)	*Unique Peptides	^Mascot MS & MS/MS Score
	13737.11	SmR	13624.85	ESI-FT	-MET, β-methylthiolation, K->T	13624.93	5.92	2	286
	13737.11		13625.11	ESI-TOF	-MET, β-methylthiolation, K->T	13624.93	-13.39		
	13737.11	SmRC	13624.78	ESI-FT	-MET, β-methylthiolation, K->T	13624.93	10.89	2	206
	13737.11		13623.94	ESI-TOF	-MET, β-methylthiolation, K->T	13624.93	72.56		
RpsM S13	13099.43	WT	Not Observed					10	103
13099.43	13099.43	SmR	12968.32	ESI-TOF	-MET	12968.23	-6.79	10	161
	13099.43	SmRC	Not Observed					9	118
RpsN S14	11580.52	WT	11448.97	ESI-FT	-MET	11449.33	31.09	1	211
11580.52	11580.52	SmR	Not Observed					1	146
	11580.52	SmRC	11449.16	ESI-FT	-MET	11449.33	14.40	1	184
RpsO S15	10268.79	WT	10137.62	ESI-TOF	-MET	10137.59	-2.47	2	61
10268.79	10268.79	SmR	10137.89	ESI-FT	-MET	10137.59	-29.60	1	-
	10268.79		10137.48	ESI-TOF	-MET	10137.59	11.44		

Subunit & Sequence-predicted Ave. Mass	Calculated Ave. Mass	Derivative	Experimental Mass	Source	Predicted PTM	Calculated Mass (plus PTM)	Prediction Error (ppm)	*Unique Peptides	^Mascot MS & MS/MS Score
	10268.79	SmRC	10137.59	ESI-FT	-MET	10137.59	0.75	1	177
	10268.79		10138.26	ESI-TOF	-MET	10137.59	-65.60		
RpsP S16	9190.58	WT	9190.60	ESI-FT	None	9190.58	-2.93	3	366
9190.58	9190.58		9190.38	ESI-TOF	None	9190.58	21.76		
	9190.58	SmR	9190.20	ESI-FT	None	9190.58	40.41	1	394
	9190.58		9190.67	ESI-TOF	None	9190.58	-10.55		
	9190.58	SmRC	9190.20	ESI-FT	None	9190.58	40.41	3	142
	9190.58		9191.89	ESI-TOF	None	9190.58	-143.41		
RpsQ S17	9704.48	WT	9573.26	ESI-FT	-MET	9573.28	2.04	1	47
9704.48	9704.48		9573.04	ESI-TOF	-MET	9573.28	25.17		
	9704.48	SmR	9573.58	ESI-FT	-MET	9573.28	-30.92	1	-
	9704.48		9572.49	ESI-TOF	-MET	9573.28	82.52		
	9704.48	SmRC	9573.12	ESI-FT	-MET	9573.28	16.43	1	83
	9704.48		9572.68	ESI-TOF	-MET	9573.28	62.57		
RpsR S18	8986.46	WT	8987.45	ESI-TOF	None	8986.46	-110.39	2	-
8986.46	8986.46	SmR	Not Observed					1	-
	8986.46	SmRC	Not Observed					2	56
RpsS S19	10430.32	WT	10298.43	ESI-FT	-MET	10299.12	66.77	3	274

Subunit & Sequence-predicted Ave. Mass	Calculated Ave. Mass	Derivative	Experimental Mass	Source	Predicted PTM	Calculated Mass (plus PTM)	Prediction Error (ppm)	*Unique Peptides	^Mascot MS & MS/MS Score
10430.32	10430.32		10299.46	ESI-TOF	-MET	10299.12	-33.21		
	10430.32	SmR	10299.17	ESI-TOF	-MET	10299.12	-5.05	3	263
	10430.32		10299.02	ESI-FT	-MET	10299.12	8.96		
	10430.32	SmRC	10298.69	ESI-FT	-MET	10299.12	41.41	2	270
	10430.32		10299.66	ESI-TOF	-MET	10299.12	-52.82		
RpsT S20/L26	9684.42	WT	9553.37	ESI-FT	-MET	9553.23	-14.99	1	137
9684.42	9684.42		9553.40	ESI-TOF	-MET	9553.23	-17.80		
	9684.42	SmR	9552.87	ESI-FT	-MET	9553.23	37.54	4	82
	9684.42		9553.27	ESI-TOF	-MET	9553.23	-4.92		
	9684.42	SmRC	9553.34	ESI-FT	-MET	9553.23	-12.14	1	116
	9684.42		9553.45	ESI-TOF	-MET	9553.23	-23.87		
RpsU S21	8499.98	WT	8368.73	ESI-FT	-MET	8368.78	6.11	1	391
8499.98	8499.98	SmR	8368.72	ESI-FT	-MET	8368.78	6.90	2	263
	8499.98		8368.78	ESI-TOF	-MET	8368.78	-0.12	2	263
	8499.98	SmRC	8368.85	ESI-FT	-MET	8368.78	-8.37	1	258
RpsV S22	5095.83	WT	Not Observed					0	393
5095.83	5095.83	SmR	Not Observed					0	25
	5095.83	SmRC	Not Observed					0	-

Subunit & Sequence-predicted Ave. Mass	Calculated Ave. Mass	Derivative	Experimental Mass	Source	Predicted PTM	Calculated Mass (plus PTM)	Prediction Error (ppm)	*Unique Peptides	^Mascot MS & MS/MS Score
RplA L1	24729.71	WT	24598.41	ESI-FT	-MET	24598.51	3.90	16	492
24729.71	24729.71		24598.80	ESI-TOF	-MET	24598.51	-11.83		
	24729.71	SmR	24598.32	ESI-FT	-MET	24598.51	7.63	20	458
	24729.71		24598.74	ESI-TOF	-MET	24598.51	-9.39		
	24729.71	SmRC	24598.81	ESI-FT	-MET	24598.51	-12.40	16	544
	24729.71		24603.52	ESI-TOF	-MET	24598.51	-203.54		
RplB L2	29860.53	WT	29728.46	ESI-FT	-MET	29729.33	29.14	9	496
29860.53	29860.53		29728.77	ESI-TOF	-MET	29729.33	18.84		
	29860.53	SmR	29729.79	ESI-TOF	-MET	29729.33	-15.51	12	235
	29860.53	SmRC	29729.05	ESI-FT	-MET	29729.33	9.24	9	214
	29860.53		29729.57	ESI-TOF	-MET	29729.33	-8.24		
RplC L3	22243.56	WT	22257.42	ESI-FT	CH2	22257.59	7.66	8	302
22243.56	22243.56		22258.99	ESI-TOF	CH2	22257.59	-62.95		
	22243.56	SmR	22257.60	ESI-FT	CH2	22257.59	-0.73	10	282
	22243.56		22258.30	ESI-TOF	CH2	22257.59	-32.13		
	22243.56	SmRC	22257.66	ESI-FT	CH2	22257.59	-3.14	8	474
RplD L4	22086.59	WT	22087.77	ESI-TOF	None	22086.59	-53.47	4	528
22086.59	22086.59	SmR	22086.53	ESI-TOF	None	22086.59	2.54	10	590

Subunit & Sequence-predicted Ave. Mass	Calculated Ave. Mass	Derivative	Experimental Mass	Source	Predicted PTM	Calculated Mass (plus PTM)	Prediction Error (ppm)	*Unique Peptides	^Mascot MS & MS/MS Score
	22086.59		22085.85	ESI-FT	None	22086.59	33.19		
	22086.59	SmRC	22087.34	ESI-TOF	None	22086.59	-34.32	4	776
RplE L5	20301.65	WT	20170.54	ESI-FT	-MET	20170.45	-4.62	12	426
20301.65	20301.65		20171.94	ESI-TOF	-MET	20170.45	-73.97		
	20301.65	SmR	20170.28	ESI-FT	-MET	20170.45	8.09	14	430
	20301.65		20171.36	ESI-TOF	-MET	20170.45	-45.17		
	20301.65	SmRC	20170.46	ESI-FT	-MET	20170.45	-0.83	13	403
RplF L6	18903.82	WT	18773.23	ESI-TOF	-MET	18772.62	-32.23	10	452
18903.82	18903.82	SmR	18772.58	ESI-FT	-MET	18772.62	2.22	11	428
	18903.82		18773.32	ESI-TOF	-MET	18772.62	-37.18		
	18903.82	SmRC	18773.17	ESI-FT	-MET	18772.62	-28.97	7	344
	18903.82		18773.79	ESI-TOF	-MET	18772.62	-61.90		
RplL L7/L12	12295.23	WT	12206.05	ESI-FT	-MET, Ac	12206.06	1.11	10	778
12295.23	12295.23		12133.78	ESI-TOF	-MET, -SER, Ac, CH2	12133.01	-62.97		
	12295.23		12149.91	ESI-TOF	-MET, -SER, Ac, 2-Ox	12150.98	88.13		
	12295.23		12206.89	ESI-TOF	-MET, Ac	12206.06	-67.32		
	12295.23		12221.71	ESI-TOF	-MET, Ac, CH2	12220.09	-132.55		

Subunit & Sequence-predicted Ave. Mass	Calculated Ave. Mass	Derivative	Experimental Mass	Source	Predicted PTM	Calculated Mass (plus PTM)	Prediction Error (ppm)	*Unique Peptides	^Mascot MS & MS/MS Score
	12295.23		12235.00	ESI-TOF	-MET, Ac, 2-CH2	12234.12	-72.09		
	12295.23		12264.81	ESI-TOF	-MET, Ac, 3-CH2, Ox	12264.14	-54.35		
	12295.23		12274.02	ESI-TOF	-MET, Ac, CH2, 2-Ox, Na	12274.07	4.56		
	12295.23		12296.25	ESI-TOF	None	12295.23	-83.37		
	12295.23	SmR	12206.53	ESI-FT	-MET, Ac	12206.06	-38.09	9	-
	12295.23		12134.33	ESI-TOF	-MET, -SER, Ac, Ox	12134.99	53.85		
	12295.23		12150.05	ESI-TOF	-MET, -SER, Ac, 2-Ox	12150.98	77.19		
	12295.23		12164.00	ESI-TOF	-MET	12164.03	2.22		
	12295.23		12207.35	ESI-TOF	-MET, Ac	12206.06	-105.50		
	12295.23		12221.14	ESI-TOF	-MET, Ac, CH2	12220.09	-85.99		
	12295.23		12247.89	ESI-TOF	-MET, Ac, 3-CH2	12248.15	21.06		
	12295.23		12264.60	ESI-TOF	-MET, Ac, 3-CH2, Ox	12264.14	-37.15		
	12295.23		12273.26	ESI-TOF	-Met, Ac, CH2, 2-Ox, Na	12274.07	66.24		
	12295.23		12285.28	ESI-TOF	-MET, Ac, 3-CH2,Ox, Na	12286.08	65.36		
	12295.23		12297.85	ESI-TOF	-Met, Ac, 2-CH2,4-Ox	12298.12	21.79		
	12295.23	SmRC	12206.84	ESI-FT	-MET, Ac	12206.06	-63.75	10	738

Subunit & Sequence-predicted Ave. Mass	Calculated Ave. Mass	Derivative	Experimental Mass	Source	Predicted PTM	Calculated Mass (plus PTM)	Prediction Error (ppm)	*Unique Peptides	^Mascot MS & MS/MS Score
	12295.23		12220.10	ESI-FT	-MET, Ac, CH2	12220.09	-0.90		
	12295.23		12134.51	ESI-TOF	-MET, -SER, Ac, Ox	12134.99	38.85		
	12295.23		12152.55	ESI-TOF	-MET, -SER, CH2, 2-Ox, Na	12151.96	-48.96		
	12295.23		12220.35	ESI-TOF	-MET, Ac, CH2	12220.09	-20.93		
	12295.23		12298.48	ESI-TOF	-Met, Ac, 2-CH2,4-Ox	12298.12	-29.27		
RplI L9	15769.09	WT	15769.28	ESI-TOF	None	15769.09	-12.18	9	189
15769.09	15769.09	SmR	15768.20	ESI-TOF	None	15769.09	56.12	9	-
	15769.09	SmRC	15770.68	ESI-TOF	None	15769.09	-101.34	9	290
RplJ L10	17711.65	WT	17579.97	ESI-FT	-MET	17580.45	27.06	5	403
17711.65	17711.65		17581.35	ESI-TOF	-MET	17580.45	-51.14		
	17711.65	SmR	17580.72	ESI-FT	-MET	17580.45	-15.45	9	404
	17711.65		17580.60	ESI-TOF	-MET	17580.45	-8.59		
	17711.65	SmRC	17580.00	ESI-FT	-MET	17580.45	25.69	5	277
RplK L11	14875.44	WT	14870.39	ESI-FT	-MET, 9-CH2	14870.48	6.21	6	166
14875.44	14875.44		14870.96	ESI-TOF	-MET, 9-CH2	14870.48	-32.27		

Subunit & Sequence-predicted Ave. Mass	Calculated Ave. Mass	Derivative	Experimental Mass	Source	Predicted PTM	Calculated Mass (plus PTM)	Prediction Error (ppm)	*Unique Peptides	^Mascot MS & MS/MS Score
	14875.44		14840.02	ESI-TOF	-MET, 3-CH2, 2-Ox, Na	14840.26	15.84		
	14875.44		14889.26	ESI-TOF	-MET, 7-CH2, 3-Ox	14890.38	75.28		
	14875.44		14911.41	ESI-TOF	-MET, 12-CH2	14912.56	77.10		
	14875.44		14936.35	ESI-TOF	2-CH2, 2-Ox	14935.49	-57.41		
	14875.44	SmR	14870.39	ESI-FT	-MET, 9-CH2	14870.48	6.21	3	92
	14875.44		14870.72	ESI-TOF	-MET, 9-CH2	14870.48	-16.33		
	14875.44	SmRC	14870.62	ESI-FT	-MET, 9-CH2	14870.48	-9.45	5	115
	14875.44		14840.96	ESI-TOF	-MET, 3-CH2, 2-Ox, Na	14840.26	-47.51		
	14875.44		14870.41	ESI-TOF	-MET, 9-CH2	14870.48	4.98		
	14875.44		14889.95	ESI-TOF	-MET, 7-CH2, 3-Ox	14890.38	29.21		
	14875.44		14911.79	ESI-TOF	-MET, 12-CH2	14912.56	51.49		
	14875.44		14937.76	ESI-TOF	2-CH2, 2-Ox	14935.49	-152.21		
RplM L13	16018.58	WT	16018.03	ESI-FT	None	16018.58	34.59	4	201
16018.58	16018.58		16018.55	ESI-TOF	None	16018.58	2.37		
	16018.58	SmR	16017.87	ESI-FT	None	16018.58	44.36	5	-
	16018.58		16018.97	ESI-TOF	None	16018.58	-24.35		
	16018.58	SmRC	16018.50	ESI-FT	None	16018.58	5.46	4	310
	16018.58		16018.83	ESI-TOF	None	16018.58	-15.05		

Subunit & Sequence-predicted Ave. Mass	Calculated Ave. Mass	Derivative	Experimental Mass	Source	Predicted PTM	Calculated Mass (plus PTM)	Prediction Error (ppm)	*Unique Peptides	^Mascot MS & MS/MS Score
RplN L14	13541.08	WT	13540.55	ESI-FT	None	13541.08	39.52	3	170
13541.08	13541.08		13539.80	ESI-TOF	None	13541.08	94.60		
	13541.08	SmR	13541.00	ESI-FT	None	13541.08	5.85	4	178
	13541.08		13538.87	ESI-TOF	None	13541.08	163.28		
	13541.08		13541.58	ESI-TOF	None	13541.08	-36.85		
	13541.08	SmRC	13540.93	ESI-FT	None	13541.08	11.17	4	185
	13541.08		13541.41	ESI-TOF	None	13541.08	-24.07		
RplO L15	14980.46	WT	14980.42	ESI-FT	None	14980.46	2.32	5	164
14980.46	14980.46		14981.92	ESI-TOF	None	14980.46	-97.66		
	14980.46	SmR	14980.63	ESI-FT	None	14980.46	-11.83	8	312
	14980.46		14980.41	ESI-TOF	None	14980.46	3.14		
	14980.46	SmRC	14980.31	ESI-FT	None	14980.46	9.49	4	257
	14980.46		14982.98	ESI-TOF	None	14980.46	-168.29		
RplP L16	15281.26	WT	15327.07	ESI-TOF	CH2, 2-Ox	15327.28	13.88	4	126
15281.26	15281.26		15311.60	ESI-TOF	CH2, Ox	15311.28	-20.88		
	15281.26	SmR	15327.92	ESI-TOF	CH2, 2-Ox	15327.28	-41.25	6	148
	15281.26		15311.31	ESI-TOF	CH2, Ox	15311.28	-1.68		
	15281.26	SmRC	15327.19	ESI-FT	CH2, 2-Ox	15327.28	5.76	4	221
	15281.26		15327.34	ESI-TOF	CH2, 2-Ox	15327.28	-3.80		

Subunit & Sequence-predicted Ave. Mass	Calculated Ave. Mass	Derivative	Experimental Mass	Source	Predicted PTM	Calculated Mass (plus PTM)	Prediction Error (ppm)	*Unique Peptides	^Mascot MS & MS/MS Score
	15281.26		15311.72	ESI-TOF	CH2, Ox	15311.28	-28.33		
RplQ L17	14364.64	WT	14364.05	ESI-FT	None	14364.64	40.56	5	176
14364.64	14364.64	SmR	14365.07	ESI-TOF	None	14364.64	-30.42	6	189
	14364.64	SmRC	14365.26	ESI-FT	None	14364.64	-43.11	5	202
RplR L18	12769.66	WT	12769.87	ESI-FT	None	12769.66	-17.02	2	369
12769.66	12769.66		12769.72	ESI-TOF	None	12769.66	-5.09		
	12769.66	SmR	12770.35	ESI-FT	None	12769.66	-54.70	2	-
	12769.66		12769.65	ESI-TOF	None	12769.66	0.39		
	12769.66	SmRC	12770.35	ESI-FT	None	12769.66	-54.70	2	99
	12769.66		12771.53	ESI-TOF	None	12769.66	-146.44		
RplS L19	13133.26	WT	13001.53	ESI-FT	-MET	13002.06	40.78	4	276
13133.26	13133.26		13002.02	ESI-TOF	-MET	13002.06	3.38		
	13133.26	SmR	13002.41	ESI-FT	-MET	13002.06	-26.56	4	79
	13133.26		13002.16	ESI-TOF	-MET	13002.06	-7.38		
	13133.26	SmRC	13002.20	ESI-FT	-MET	13002.06	-10.53	4	257
	13133.26		13002.96	ESI-TOF	-MET	13002.06	-68.76		
RplT L20	13496.99	WT	13365.42	ESI-TOF	-MET	13365.79	27.38	6	73
13496.99	13496.99	SmR	13365.48	ESI-FT	-MET	13365.79	22.71	4	53

Subunit & Sequence-predicted Ave. Mass	Calculated Ave. Mass	Derivative	Experimental Mass	Source	Predicted PTM	Calculated Mass (plus PTM)	Prediction Error (ppm)	*Unique Peptides	^Mascot MS & MS/MS Score
	13496.99		13366.01	ESI-TOF	-MET	13365.79	-16.91		
	13496.99	SmRC	13365.53	ESI-FT	-MET	13365.79	19.08	4	-
	13496.99		13367.60	ESI-TOF	-MET	13365.79	-135.79		
RplU L21	11564.37	WT	11564.37	ESI-FT	None	11564.37	0.51	5	-
11564.37	11564.37		11564.74	ESI-TOF	None	11564.37	-31.56		
	11564.37	SmR	11562.90	ESI-TOF	None	11564.37	127.03	3	102
	11564.37	SmRC	11564.49	ESI-FT	None	11564.37	-10.34	5	93
RplV L22	12226.33	WT	12226.49	ESI-TOF	None	12226.33	-13.58	7	134
12226.33	12226.33	SmR	12226.63	ESI-TOF	None	12226.33	-24.86	9	84
	12226.33		12226.04	ESI-FT	None	12226.33	23.58		
	12226.33	SmRC	12227.28	ESI-TOF	None	12226.33	-77.95	6	140
RplW L23	11199.14	WT	11199.23	ESI-TOF	None	11199.14	-7.50	3	-
11199.14	11199.14	SmR	11199.29	ESI-TOF	None	11199.14	-13.13	1	-
	11199.14		11197.29	ESI-TOF	None	11199.14	165.46		
	11199.14		11199.08	ESI-FT	None	11199.14	5.76		
	11199.14	SmRC	11199.12	ESI-FT	None	11199.14	2.27	4	-
	11199.14		11198.52	ESI-TOF	None	11199.14	55.63		
RplX L24	11316.23	WT	11185.06	ESI-FT	-MET	11185.03	-2.43	8	377

Subunit & Sequence-predicted Ave. Mass	Calculated Ave. Mass	Derivative	Experimental Mass	Source	Predicted PTM	Calculated Mass (plus PTM)	Prediction Error (ppm)	*Unique Peptides	^Mascot MS & MS/MS Score
11316.23	11316.23		11185.18	ESI-TOF	-MET	11185.03	-13.41		
	11316.23	SmR	11185.05	ESI-FT	-MET	11185.03	-1.01	8	301
	11316.23		11185.28	ESI-TOF	-MET	11185.03	-21.99		
	11316.23	SmRC	11185.03	ESI-FT	-MET	11185.03	-0.06	7	478
	11316.23		11186.08	ESI-TOF	-MET	11185.03	-93.25		
RplY L25	10693.47	WT	10693.45	ESI-FT	None	10693.47	1.98	8	176
10693.47	10693.47		10693.48	ESI-TOF	None	10693.47	-0.56		
	10693.47	SmR	10693.45	ESI-FT	None	10693.47	1.68	2	-
	10693.47		10693.65	ESI-TOF	None	10693.47	-16.55		
	10693.47		10693.20	ESI-TOF	None	10693.47	25.34		
	10693.47	SmRC	10693.35	ESI-FT	None	10693.47	11.36	8	194
	10693.47		10694.29	ESI-TOF	None	10693.47	-76.96		
RpmA L27	9124.49	WT	8993.04	ESI-FT	-MET	8993.30	28.48	2	420
9124.49	9124.49	SmR	8993.45	ESI-FT	-MET	8993.30	-16.73	2	309
	9124.49		8993.60	ESI-TOF	-MET	8993.30	-33.36		
	9124.49	SmRC	8993.27	ESI-FT	-MET	8993.30	2.56	1	130
RpmB L28	9006.52	WT	8874.79	ESI-FT	-MET	8875.32	59.17	1	265
9006.52	9006.52	SmR	8875.31	ESI-TOF	-MET	8875.32	0.68	1	299
	9006.52	SmRC	8875.42	ESI-FT	-MET	8875.32	-11.80	1	282

Subunit & Sequence-predicted Ave. Mass	Calculated Ave. Mass	Derivative	Experimental Mass	Source	Predicted PTM	Calculated Mass (plus PTM)	Prediction Error (ppm)	*Unique Peptides	^Mascot MS & MS/MS Score
RpmC L29	7273.47	WT	7273.22	ESI-FT	None	7273.47	34.79	3	245
7273.47	7273.47		7273.29	ESI-TOF	None	7273.47	24.34		
	7273.47	SmR	7272.76	ESI-FT	None	7273.47	97.26	1	135
	7273.47		7273.56	ESI-TOF	None	7273.47	-12.65		
	7273.47	SmRC	7273.49	ESI-FT	None	7273.47	-2.85	3	141
	7273.47		7274.33	ESI-TOF	None	7273.47	-117.83		
RpmD L30	6541.82	WT	6410.17	ESI-FT	-MET	6410.62	69.73	2	104
6541.82	6541.82		6410.62	ESI-TOF	-MET	6410.62	-0.16		
	6541.82	SmR	6410.31	ESI-FT	-MET	6410.62	47.66	1	-
	6541.82		6410.45	ESI-TOF	-MET	6410.62	25.89		
	6541.82	SmRC	6410.67	ESI-FT	-MET	6410.62	-7.65	2	95
	6541.82		6411.74	ESI-TOF	-MET	6410.62	-175.49		
RpmE L31	7871.11	WT	7871.26	ESI-FT	None	7871.11	-19.82	0	74
7871.11	7871.11		7869.20	ESI-TOF	None	7871.11	242.41		
	7871.11	SmR	7871.05	ESI-FT	None	7871.11	7.20	0	-
	7871.11		7869.26	ESI-TOF	None	7871.11	235.16		
	7871.11	SmRC	Not Observed					1	62
RpmF L32	6446.40	WT	6314.68	ESI-FT	-MET	6315.20	82.79	1	382

Subunit & Sequence-predicted Ave. Mass	Calculated Ave. Mass	Derivative	Experimental Mass	Source	Predicted PTM	Calculated Mass (plus PTM)	Prediction Error (ppm)	*Unique Peptides	^Mascot MS & MS/MS Score
6446.40	6446.40	SmR	6314.65	ESI-FT	-MET	6315.20	87.27	3	-
	6446.40		6314.62	ESI-TOF	-MET	6315.20	92.63		
	6446.40	SmRC	6314.65	ESI-FT	-MET	6315.20	87.27	2	-
RpmG L33	6371.60	WT	6254.57	ESI-FT	-MET, CH2	6254.42	-23.59	1	387
6371.60	6371.60		6239.70	ESI-TOF	-MET	6240.40	112.81		
	6371.60		6254.21	ESI-TOF	-MET, CH2	6254.42	34.04		
	6371.60		6267.80	ESI-TOF	-MET, 2-CH2	6268.45	103.82		
	6371.60	SmR	6254.32	ESI-FT	-MET, CH2	6254.42	16.87	1	223
	6371.60		6238.60	ESI-TOF	-MET	6240.40	289.08		
	6371.60		6254.24	ESI-TOF	-MET, CH2	6254.42	29.08		
	6371.60		6268.12	ESI-TOF	-MET, 2-CH2	6268.45	52.45		
	6371.60	SmRC	6254.45	ESI-FT	-MET, CH2	6254.42	-3.62	1	211
	6371.60		6240.80	ESI-TOF	-MET	6240.40	-63.46		
	6371.60		6254.96	ESI-TOF	-MET, CH2	6254.42	-85.24		
	6371.60		6268.00	ESI-TOF	-MET, 2-CH2	6268.45	71.92		
RpmH L34	5380.41	WT	Not Observed					0	281
5380.41	5380.41	SmR	Not Observed						-
	5380.41	SmRC	Not Observed						-

Subunit & Sequence-predicted Ave. Mass	Calculated Ave. Mass	Derivative	Experimental Mass	Source	Predicted PTM	Calculated Mass (plus PTM)	Prediction Error (ppm)	*Unique Peptides	^Mascot MS & MS/MS Score
RpmI L35	7288.96	WT	7157.32	ESI-FT	-MET	7157.76	61.47	0	38
7288.96	7288.96		7157.58	ESI-TOF	-MET	7157.76	25.29		
	7288.96	SmR	7157.99	ESI-TOF	-MET	7157.76	-31.71		-
	7288.96	SmRC	Not Observed						-
RpmJ L36	4364.36	WT	4362.77	ESI-TOF	None	4364.45	385.98	0	25
4364.36	4364.36	SmR	4363.27	ESI-TOF	None	4364.45	271.42		42
	4364.36		4363.79	ESI-TOF	None	4364.45	152.28		
	4364.36		4364.72	ESI-FT	None	4364.45	-60.05		
	4364.36	SmRC	4313.78	ESI-TOF	unknown	-	-		-

Table A1 - Summary of top-down/bottom-up MS protein analyses. Summary of ribosomal protein identifications and PTM isoforms predicted in the combined top-down/bottom-up analysis of *E. coli* ribosomal proteins showing FT-ICR generated (green) and BioTOFII generated (orange) data. *Unique peptides observed in the bottom-up (shotgun) datasets. ^Mascot MS and MS/MS scores from the analysis of MALDI MS/MS datasets.