12-2009

# Targeted Proteomics for the Characterization of Enriched Microbial Protein Isolates and Protein Complexes

William Judson Hervey

*University of Tennessee-Knoxville/Oak Ridge National Laboratory Graduate School of Genome Science and Technology*, judson.hervey@gmail.com

To the Graduate Council:

I am submitting herewith a dissertation written by William Judson Hervey entitled "Targeted Proteomics for the Characterization of Enriched Microbial Protein Isolates and Protein Complexes." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Life Sciences.

Gregory B. Hurst, Major Professor

We have read this dissertation and recommend its acceptance:

Dale A. Pelletier, Robert L. Hettich, Kurt H. Lamour, Frank W. Larimer

Accepted for the Council:

Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

To the Graduate Council:

I am submitting herewith a dissertation written by William Judson Hervey, IV entitled "Targeted Proteomics for the Characterization of Enriched Microbial Protein Isolates and Protein Complexes." I have examined the final electronic copy of this dissertation for form and content and recommended that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Life Sciences.

Gregory B. Hurst
Major Professor

Dale A. Pelletier
Major Professor

We have read this dissertation
and recommended its acceptance:

Robert L. Hettich

Kurt H. Lamour

Frank W. Larimer

Accepted for the Council:

Carolyn R. Hodges
Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

# Targeted Proteomics for the Characterization of Enriched Microbial Protein Isolates and Protein Complexes

A Dissertation
Presented for the
Doctorate of Philosophy
Degree
The University of Tennessee, Knoxville

William Judson Hervey, IV
December 2009

# Dedication

This dissertation is dedicated to my father, William Judson Hervey, III.

Without his selfless sacrifices, steadfast love, and unwavering support throughout my *entire* education, the work presented would not have been possible.

## Acknowledgements

It is necessary that I first acknowledge the effort Dr. Gregory B. Hurst has put forth during my mentorship. Throughout fluxes of experimental success and failure, Greg has guided me with upmost professionalism. It is not possible for me to articulate the sense of gratitude I have for Greg's mentorship.

I also acknowledge the members of my doctoral committee for directing my research: Dr. Dale A. Pelletier, Dr. Robert L. Hettich, Dr. Kurt H. Lamour, and Dr. Frank W. Larimer. I thank Dr. Pelletier and members of his laboratory for extending their biological knowledge and facilities to me. It is also necessary to thank Dr. Hettich for his participation in the GST curriculum, which has significantly contributed to my knowledge of mass spectrometry.

I acknowledge the Organic and Biological Mass Spectrometry (OBMS) Group as a whole. I especially thank Becky R. Maggard of the OBMS group for clerical assistance.

I acknowledge Dr. M. Brad Strader (NIH), Dr. Nathan VerBerkmoes (OBMS), Dr. Chongle Pan (ORNL), Dr. David L. Tabb (Vanderbilt), and Dr. W. Hayes McDonald (Vanderbilt) for contributing their expertise in biochemistry and mass spectrometry during my studies. I also acknowledge Mrs. Patricia K. Lankford and Mr. Keiji G. Asano for administering technical assistance while conducting this research.

This research would not have been possible without sponsorship by the Genomics:GTL program of the Office of Biological and Environmental Research, U.S. Department of Energy. I acknowledge the University of Tennessee-Knoxville and Oak Ridge National Laboratory Graduate School of Genome Science and Technology for providing me with this unique opportunity.

Further, I acknowledge numerous professional and personal relationships with fellow GST graduate students throughout the years. If it were not for the GST program, it is unlikely that I would have met my best friend, Mr. Adam M. Tebbe (Transform Pharmaceuticals).

At this time, I would also like to acknowledge the memory of my grandparents for their collective contributions to my education. From a very young age, all four of my grandparents instilled the value of education through their actions and words.

On a more personal note, I acknowledge the support of Karyn L. Massi, D. V. M. over the course of the past 3 years.

I also acknowledge that I may have inadvertently omitted contributions others have made on my behalf. In this event, I apologize in advance.

## Abstract

The field of proteomics encompasses the study of identities, interactions, and dynamics of all proteins expressed by a living system. Research in this dissertation blends biochemical and quantitative proteomics techniques to increase the latitude of biological applications for the bottom-up mass spectrometry proteomics approach. Together, isolation of selected protein "targets," such as multiprotein complexes, and quantitative characterization yields information essential for more detailed understanding of microbial cell function.

Often, a challenging aspect of characterizing a variety of biochemically enriched samples is limited protein yield. This dissertation describes an enzymatic proteolysis protocol employing an organic/aqueous solvent that alleviates excessive handling steps to reduce losses during sample preparation for small quantities of protein samples.

Presence of artifactual, non-specific proteins in enriched protein complex isolates complicates biological interpretation of specific protein interactions. Heterologous expression of affinity-tagged bait proteins may also cause unintended collateral effects. A series of local and global protein isotope ratio measurements were performed to differentiate authentic interactions from artifactual interactions among affinity-isolated complexes and assess collateral effects, respectively.

Protein localization provides clues regarding protein function. To infer protein localization, quantitative proteomics techniques were used to estimate protein enrichment of cold osmotic shock periplasmic isolates. Protein isotope ratios indicating enrichment, combined with identification of amino-terminal signal peptide cleavages, increase confidence of periplasmic localization.

Collectively, this dissertation provides a framework for tailoring biochemical and quantitative techniques for targeted characterization of microbial protein isolates.

# Table of Contents

# Chapter 1
# Introduction

## *1.1 First Sight to Systems Insight*

Some of the greatest discoveries in life science have arisen from interdisciplinary technological applications. In the late 17th century, Antonie van Leeuwenhoek placed scrapings from his own mouth on his microscope stage, likely out of curiosity. As a result, van Leeuwenhoek was the first person to see bacteria and fathered the field of microbiology [1]. Though this discovery was over 300 years ago, it represents a reoccurring theme throughout the study of life. Curiosity, combined with cutting-edge technology, the microscope in this case, resulted in a significant stride forward in life science: the discovery of organisms unable to be seen with the naked eye.

More recently, much of the 20th century was spent defining life at the molecular level, probing for the fundamentals upon which life is conducted. Rather than view external features and characteristics of life solely in an observational fashion, the advent of molecular biology opened the study of dynamics among biomolecules that comprise living systems. The central dogma of molecular biology was among the most profound discoveries of this era [2]. This concept states that genetic information, encoded in nucleotides of deoxyribonucleic acid (DNA), is transcribed to messenger ribonucleic acid (RNA), and is translated into the amino acid sequences of proteins. Thus, hereditary information encoded in DNA's smaller units, or genes, is passed from one generation to the next containing an organism's "blueprint." The "building blocks" of the DNA "blueprint" are proteins, the biomolecules that maintain the flow of genetic information and perform an array of functions essential to life.

In the current era, technological advancements have dramatically changed the perspective from which life is viewed, not unlike Antonie van Leeuwenhoek's time. The focus of life science has shifted from studies of single genes, transcripts, or proteins to studies encompassing the entire suite of biomolecules produced by an organism, or living "system." Dubbed Systems Biology, this paradigm transition was initially spurred by acquisition of large amounts of DNA sequences, as reviewed by Ideker *et al.* [3, 4]. Application of automated DNA sequencing technology allowed measurement of an organism's entire set of genes, or genome. To obtain a more "holistic" view of organisms consistent with the central dogma, the fields of transcriptomics and proteomics have developed for systems study of mRNA transcripts and proteins, respectively. The proteome, or entire component of expressed proteins, could be considered a cast of protein entities performing fundamental life processes [5]. Proteome studies may provide the identities and implicate functional "roles" for each protein, or "cast member," present in a system. In contrast to the other "-omics" fields, proteomics offers a "snapshot" view of proteins performing essential life processes, such as maintaining the flow of genetic information.

Potential knowledge of a global "snapshot" of the proteome has generated considerable interest in this interdisciplinary field of study. However, proteome studies are challenging for several reasons. First, as the most versatile biomolecules of the cell, proteins are remarkably dynamic: their expression, function, and abundance in the cell are not fixed or static. Secondly, the proteome is a very complex, heterogenous mixture of proteins, differing in physical and chemical properties and cellular abundances. Finally, proteins may undergo a wide range of post-translational modifications (PTMs)

influencing their function, localization, and/or regulation. Based upon these attributes, characterization of the proteome is more complex than its systems biology counterparts of genomics and transcriptomics. Despite these formidable challenges facing proteomics studies, a number of technological advances have converged to enable protein study on a systems level by mass spectrometry-based approaches [6].

## *1.2 The "Bottom-up" Mass Spectrometry-based Approach to Proteome Study*

Mass spectrometry-based approaches [7] are at the forefront for proteome characterization due to a confluence of technological advancements in analytical instrumentation, multi-dimensional chromatographic separations, and bioinformatics. One widely-used proteomics approach is to perform qualitative analysis of proteolytic peptides by one dimensional (1D) or two dimensional (2D) chromatographic separation coupled via electrospray ionization (ESI) [8] with tandem mass spectrometry (LC-MS-MS), generating peptide fragmentation spectra that are compared to translated genomic databases [9, 10]. The combination of this suite of analytical technologies is commonly referred to as the "bottom-up" or "shotgun" proteomics approach, as reviewed by VerBerkmoes *et al.* [6].

Nominally, the bottom-up approach refers to the characterization of proteolytic peptide mixtures from protein samples, as opposed to the "top-down" approach of directly characterizing intact protein mixtures. The scope of this dissertation focuses on application of the bottom-up approach for proteome study; however, each of these approaches has distinct strengths and weaknesses. For instance, the top-down approach provides valuable information regarding the intact protein molecular mass and potential

post-translational modifications. However, the top-down approach yields fewer protein identifications and is challenging to directly interface with liquid chromatographic separations. Coupling liquid separations for a complex mixture of proteins, expressed at different abundances, physical, and chemical properties is a formidable challenge. Alternatively, the bottom-up approach provides a direct manner to identify the protein content of a sample mixture based upon its constituent tryptic peptides. Though providing limited information regarding intact proteins, the bottom-up approach is straightforwardly coupled to liquid chromatographic separations [6]. Furthermore, this approach has benefitted from numerous software tools to identify peptides and constituent proteins from complex mixtures. At first, increasing the complexity of a protein mixture to a peptide mixture via proteolysis may seem counterintuitive. However, it creates a more uniform mixture of analytes (tryptic peptides) that are more tractable for existing liquid chromatographic separations. In general, complex proteolytic peptide mixtures characterized by the bottom-up approach do not exhibit the diversity of physical and chemical properties that their constituent proteins represent.

The robust bottom-up approach has been applied, for example, to identifying proteins from complex mixtures of one or more organisms [11, 12], less abundant samples obtained from biochemically-enriched multiprotein complexes [13], and affinity isolations of single proteins along with interacting partners [14, 15].

Successful application of the bottom-up approach for proteome characterization would not be possible without technological advancements in mass spectrometry instrumentation. Three essential components of a mass spectrometer are an ionization source, a mass analyzer, and a detector. Over the course of the past 30 years, innovations

in ionization techniques and mass analyzers have benefitted the characterization of biomolecules by mass spectrometry, and, consequently, the bottom-up proteomics approach.

The use of mass spectrometry-based approaches for the characterization of biomolecules would not be possible without development of "soft" ionization techniques. Electrospray ionization (ESI) developed by Fenn *et al.* [8], and matrix assisted laser desorption ionization (MALDI), developed by Hillenkamp *et al.* [16], both generate gas phase ions from biomolecules gently, without destroying or damaging them in the process. In the above described implementation of the bottom-up approach, ESI is used to directly ionize peptides from the solution phase to the gas phase by application of high voltage (2-4 kV). This produces a number of multiply-charged peptide ions which enter the mass spectrometer for further manipulation. The impact of these two ionization techniques was so great that Fenn and Koichi Tanaka were jointly awarded the 2002 Nobel Prize in Chemistry for applying the ionization techniques of ESI and MALDI, respectively, to proteins.

In addition to ionization techniques, developments in analytical instrumentation have contributed significantly to the bottom-up proteomics approach. Specifically, ion trap mass spectrometers, such as the three dimensional (3D) quadrupole ion trap [17] and two dimensional (2D) linear ion trap [18] instruments, have proved robust "workhorses" in the bottom-up proteomics approach. The ion trap mass analyzers of these instruments allow storage, manipulation, and fragmentation of ions by tandem mass spectrometry (MS-MS). Performing MS-MS of peptide ions by collision-induced dissociation (CID) [19] with a neutral gas, such as helium or nitrogen, results in a characteristic

fragmentation pattern of the peptide's amino acid sequence [20, 21]. These properties of ion trap mass spectrometers make them extremely useful for measurement of complex mixtures of peptides, as peptide ions may be collected, sorted by their mass-to-charge ratio (m/z), stored, fragmented to yield their amino acid sequence, and ejected from the ion trap. Programming iterative cycles of these operations and data-dependent MS-MS acquisition are possible under automated control of instrument software.

Two commercially-available ion trap mass spectrometers manufactured by ThermoFinnigan were used for data acquisition in this dissertation: the LCQ DecaXP 3D ion trap mass spectrometer and the 2D LTQ ion trap mass spectrometer. In its own right, the LCQ DecaXP is a robust "workhorse" of an instrument when implemented as part of a bottom-up proteomics approach. During the course of this dissertation, the LTQ instrument was developed [18] and commercially made available. Though analogous to the LCQ DecaXP with respect to electronics and scanning functions, the LTQ instrument provides more comprehensive proteome characterization due to the 2D shape of its ion trap. The linear shape of this mass analyzer permits an increased volume of peptide ions for storage, manipulation, and fragmentation at greater trapping efficiency. The linear shape also permits radial ejection of ions to two electron multiplier detectors, one on either side of the mass analyzer. In comparison to the LCQ DecaXP, these improvements make the LTQ instrument more adept at identifying peptides of lower abundance amid peptides of higher abundance in a given sample, effectively increasing the dynamic range of proteome measurement. The faster scanning rate and two electron multiplier detectors also increase the duty cycle of the LTQ relative to the LCQ DecaXP. Qualitative [22] and quantitative [23] comparisons between these two instruments have shown that the

characteristics of the LTQ result in up to a 5-fold increase in the number of protein identifications among microbial proteomes. Thus, many advancements in analytical instrumentation have made proteome studies by the bottom-up approach possible.

A second advancement contributing to the effectiveness of the bottom-up approach is the relative ease of coupling liquid chromatographic separations directly to mass spectrometer instruments by ESI. A key development in increased sensitivity of the bottom-up approach has been application of multidimensional nanoflow high performance liquid chromatographic (nanoHPLC) separations. In general, nanoHPLC forms smaller liquid droplets more easily desolvated, leading to more efficient ESI. The proteome is expressed at high complexity in very different physical and chemical properties varying from protein to protein. Proteolytic digestion of a protein sample with an enzyme that has predictable cleavage properties, such as trypsin, does yield a peptide sample of even greater complexity; however, a mixture of tryptic peptides is less diverse in its physical and chemical properties than whole proteins. Thus, tryptic peptides are more amenable to liquid chromatographic separation than whole proteins.

Development of the multidimensional protein identification technology (MudPIT) [24-26] by the laboratory of John R. Yates combines online strong cation exchange (SCX) and reverse phase ($C_{18}$) nanoHPLC separation to identify complex mixtures of proteolytic peptides. The MudPIT technology permits a "gel-free" method to separate peptides in two dimensions: based on affinity to the negatively-charged SCX phase and hydrophobicity to the $C_{18}$ reverse phase. This nanoHPLC separation allows greater sensitivity for characterization of complex peptide mixtures. When MudPIT separations

are coupled to LTQ linear ion trap mass spectrometers for peptide measurement, it is possible to identify up to 2,500 proteins in a microbial proteome [6].

The third advancement permitting efficient proteome characterization by the bottom-up proteomics approach is in bioinformatics. Though the commercially-licensed SEQUEST algorithm [9, 10] was distributed with ThermoFinnigan instruments, several informatics bottlenecks prohibited automated peptide and protein identification in proteomics experiments. The SEQUEST algorithm uses cross-correlation comparisons between theoretical peptide fragmentation patterns from a FASTA-formatted protein sequence database with observed peptide fragmentation patterns from MS-MS spectra in an experiment. SEQUEST results were provided in the form of thousands of output (*.out) text files that listed cross-correlation identification scores, peptide sequences, descriptions of their representative protein, and other spectral data, such as scan number. A significant bottleneck in the informatics workflow was filtering accurate identifications from inaccurate identifications among the SEQUEST output files, organizing accurate peptide identifications by loci (into predicted proteins), and comparing LC-MS-MS experiments with each other.

The development of DTASelect and Contrast programs [27] provided a suite of tools to overcome the above described computational bottleneck. The DTASelect program allows a user to select cross-correlation score thresholds for peptide identification, providing a means to parse thousands of output text files in an automated fashion. DTASelect further organizes peptide identifications passing user-specified thresholds into their constituent proteins and allows further advanced classification options. These data are conveniently displayed in html-formatted output. An additional

useful feature of the program is its ability to output results in tab-delimited format for further manipulation in relational databases, such as Microsoft Access, for further data mining. The Contrast program provides output for visual comparisons between LC-MS-MS experiments in html-formatted output.

Through the use of Perl scripting and more compact output file formats [28], it is possible to automate peptide and protein identification combining SEQUEST and DTASelect. Since the development of DTASelect and Contrast, peptide and protein identification from data-dependent LC-MS-MS experiments has progressively become more automated. In this dissertation, an automated method of peptide and protein identification using these tools is employed. Though numerous open-source algorithms have been developed in recent years, the combination of SEQUEST and DTASelect are acceptable methods to report peptide and protein identifications in primary literature.

## *1.3 Biological Applications of Proteomics*

The bottom-up proteomics approach is remarkably versatile and may be adapted to study different features of the proteome depending upon the biological goal(s) of the study. The approach may be applied in at least five different ways to yield unique biological insight:

a. **Proteome inventory profiling**: the objective of proteome inventory measurement is to identify the entire complement of proteins expressed in an organism under a distinct set of biological conditions. Inventory profiling provides a starting point for subsequent proteome investigations by identifying expressed proteins from predicted genes. Thus, this area is very closely tied to methods of gene prediction [6, 29].

b. **Comparative proteome profiling**: the objective of comparative measurement is to detect changes among the proteome profile by qualitative or quantitative methods. In contrast to inventory profiling, comparative profiling may implicate specific proteins in a biological function under a set of biological conditions or between a set of biological conditions.

c. **Protein Interaction Analyses**: the objective of interaction analyses is to measure physical protein interactions which may infer membership in a protein complex. These analyses often combine small quantities of biochemically-enriched proteins with either qualitative or quantitative measurement by the bottom-up proteomics approach for identification [30, 31].

d. **Protein Localization**: the objective of localization studies is to infer the distribution of a single protein or multiple proteins throughout the cell. This may involve biochemical enrichment of a subcellular protein fraction, organelle, or structure combined with either qualitative or quantitative measurement by the bottom-up proteomics approach for identification [32, 33].

e. **Post-translational Modification (PTM) Identification**: the objective of PTM analysis is to pinpoint modification of specific amino acid residues among peptide identifications measured by either qualitative or quantitative methods. These measurements, in turn, may implicate the constitutive protein in a specific biological function or cellular location [34-36]

During the course of this dissertation, quantitative proteomics techniques have been applied to the above five areas of proteome study, reviewed in [37, 38]. Quantitative proteomics techniques provide measurement of either absolute [35, 39] or relative protein abundance. Often, experimental designs incorporating quantitative proteomics techniques yield measurement of significant protein changes in abundance from measurement of proteins that do not change. Depending on experimental goals, quantitative measurements may assist with biological understanding and identify potential target proteins for follow-up studies.

For the scope of this dissertation, quantitative proteomics techniques using stable isotope labeling with enriched $^{15}N$ were performed in microbial species. In general, these techniques allow relative measurements of abundance between $^{15}N$-labeled peptides and $^{14}N$-labeled peptides. Metabolic labeling with the stable isotope of nitrogen that is relatively low in natural abundance, $^{15}N$, in microbial species is a straightforward established technique, performed by Meselson and Stahl in the 1950s to study the replication of DNA in *Escherichia coli* [40]. Introducing the isotopic label as a salt in minimal growth medium ensures that all amino acids synthesized will contain enriched $^{15}N$. Another consideration is the reduction of potential sources of error between $^{15}N$-labeled and unlabeled samples, as losses between the two samples should be equivalent. In-house development of the ProRata program by Pan *et al* [41] was a principal factor enabling use of $^{15}N$ quantitative measurements. This software performs relative abundance measurement of peptide and protein isotope ratios between unlabeled samples, most abundant in $^{14}N$, and samples labeled with $^{15}N$.

## *1.4 Dissertation Focus*

The focus of this dissertation is to demonstrate the utility of the bottom-up proteomics approach to address several selected microbiological questions ***beyond*** full proteome inventory methods. The bottom-up proteomics approach has proven effective for proteome inventory profiling of microbial species [6]. Rather than being viewed as an end result, the protein inventories from such experiments may be considered a starting point for more detailed proteome investigation *beyond* protein identification. To this end, this dissertation tailors biochemical protein enrichment methods and quantitative proteome techniques with the bottom-up approach to characterize targeted protein isolates.

Measurements of protein samples isolated from a model microbial species, *Escherichia coli*, and an environmental microbial species of interest in the Department of Energy's Genomics:GTL program, *Rhodopseudomonas palustris*, are performed in this dissertation. Chapter 2 highlights the commercial bottom-up proteomics platform used for LC-MS-MS measurements.

The interdisciplinary field of proteomics often involves close collaboration between biochemists and analytical chemists, who perform protein isolation and LC-MS-MS measurements, respectively. An area of technical concern between laboratories engaged in this field of research is sample preparation and proteolytic digestion of the protein sample. A method for proteolytic digestion of affinity-isolated protein complexes, previously developed at ORNL (Appendix A), was adapted for digestion of small protein quantities. In Chapter 3, a comparison of this digestion protocol with other frequently used protocols is performed.

In Chapter 4, protein-protein interactions are further examined using a general system for plasmid-encoded bait proteins developed at ORNL (Appendix B). A challenge in the study of protein interactions is distinguishing authentic interactions from background interactions. In addition, any collateral effects arising from bait protein expression should be monitored. Collectively, local and global isotope ratio measurements employed by quantitative proteomics techniques address both of these concerns.

In Chapter 5, protein localization to the periplasm is explored by means of quantitative proteomics techniques. Measurement of periplasmic to whole cell lysate (PP:WCL) protein isotope ratios among protein mixtures provides an estimate of protein enrichment useful for potentially inferring periplasmic localization. In addition, experimental identification of the PTM of amino-terminal signal peptide cleavage positions among enriched proteins implicates protein transit via the general secretory pathway. Taken together, these measurements increase the confidence of periplasmic localization in microbial species. Chapter 6 concludes the dissertation and offers a perspective regarding the future of the field.

In a manner similar to that of van Leeuwenhoek's use of the microscope, the overall goal of this dissertation is to provide unique perspective of life. Both the microscope and mass spectrometry-based proteomics are tools that have changed the way life is studied. Here, the bottom-up approach is extended to proteome studies beyond surveys of all expressed proteins. Tailoring biochemical protein enrichment and quantitative proteome techniques expands the scope of biological questions that may be addressed by the bottom-up proteomics approach. The combination of biochemical and

analytical tools provide different perspectives of the proteome focused on detailed examination of a protein fraction or protein complex.

# Chapter 2
# Experimental Platform for Targeted Characterization of Enriched Microbial Protein Isolates and Protein Complexes

## *2.1 Introduction*

This chapter briefly outlines the experimental platform for targeted characterization of microbial protein isolates by the "bottom-up" proteomics approach. All experimental measurements were performed on commercial quadrupole ion trap mass spectrometers. Experimental methods of peptide separation by liquid chromatography and analysis by tandem mass spectrometry (MS-MS) are explained in greater detail in the Experimental Procedures section of each chapter.

Different characteristics of microbial protein isolates analyzed by the bottom-up approach in each chapter required that experimental methods be developed for each chapter. Here, a generalized outline of the ORNL proteome analysis platform is provided in Figure 2.1a, with steps of the workflow modified for targeted proteome characterization in this dissertation shown in Figure 2.1b. General steps in the workflow include microbial cell growth, protein isolation, sample preparation, liquid chromatography and tandem mass spectrometry (LC-MS-MS), and proteome informatics. More detailed information regarding protein isolation from different microbial isolates is provided in the Experimental Procedures section of each chapter.

## *2.2 Microbial Cell Growth*

All studies in this dissertation involve growth of bacterial cells from glycerol stocks to larger "production" cultures. In general, glycerol stocks of the desired species and strain were removed from storage at -80°C and streaked on an agar plate. Isolated

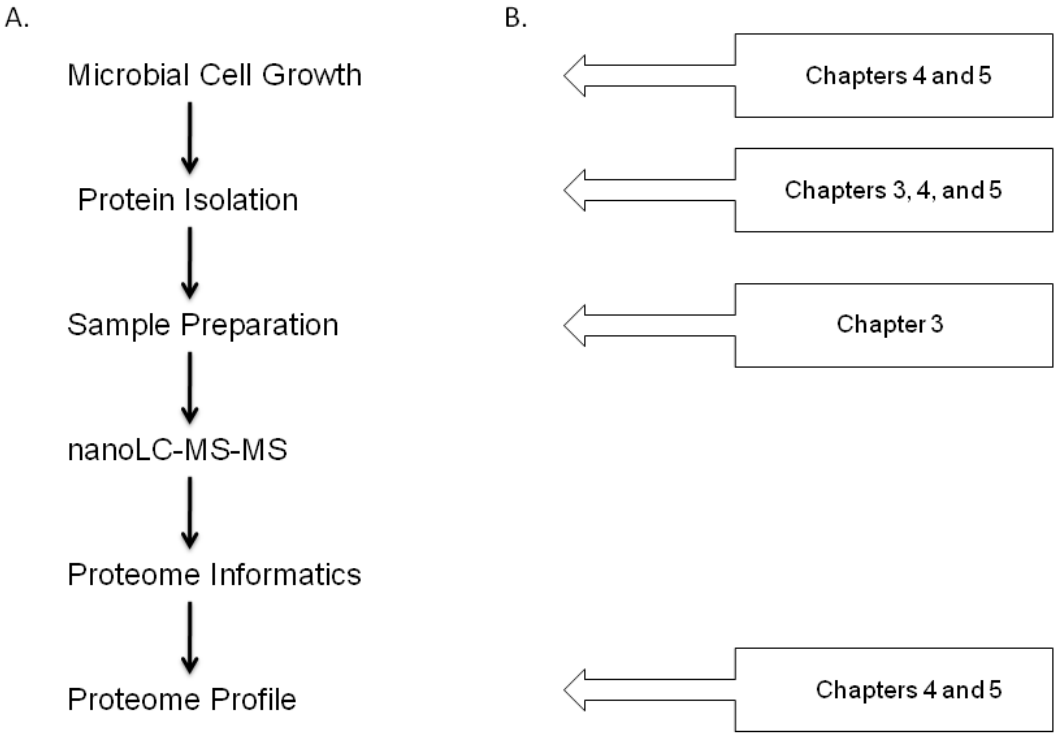



**Figure 2.1.** a. General experimental design for proteomics experiments using the bottom-up proteomics approach. b. Steps of the experimental design modified for targeted characterization of microbial protein isolates which are the focused on in the specified chapters.

colonies grown from agar plates were transferred to small liquid incolum cultures (ranging from 10 – 25 mL) for further growth. The volume of the innoculum culture was transferred to larger production cultures (ranging from 500 mL – 1 L) and growth was conducted to mid-log phase. For production cultures, all growth was conducted in minimal salt medium under optimal growth conditions for each bacterial species. *Rhodopseudomonas palustris* production cultures were grown in specialized bottles designed by Dr. Dale A. Pelletier at ORNL, shown in Figure 2.2. More details regarding the cellular growth of *R. palustris* are provided in Chapter 3 and $^{15}$N-metabolic labeling of this species is described in Chapter 4. *E. coli* production cultures were grown in a similar manner, with further details provided in Chapters 4 and 5 describing $^{15}$N-metabolic labeling.

## *2.3 Protein Isolation*

The studies in this dissertation involve several different methods of protein isolation from microbial cells, each of which is discussed in more detail for each specific chapter. In general, cell cultures were harvested by centrifugation, resuspended in a salt buffer, and transferred to 25 mL Oak Ridge centrifugation tubes. After obtaining this microbial cell pellet, protein isolation steps differ by chapter and are described in more detail in the appropriate Experimental Procedures sections.

In Chapter 3, sucrose density centrifugation is used to enrich the *R. palustris* 70*S* ribosomal protein complex, which has been previously described by Strader *et. al*. [42]. The other protein sample used in Chapter 3, microtubule-associated proteins (MAP), was a gift from the laboratory of Timothy J. Mitchison (Department of Systems Biology, Harvard Medical School).

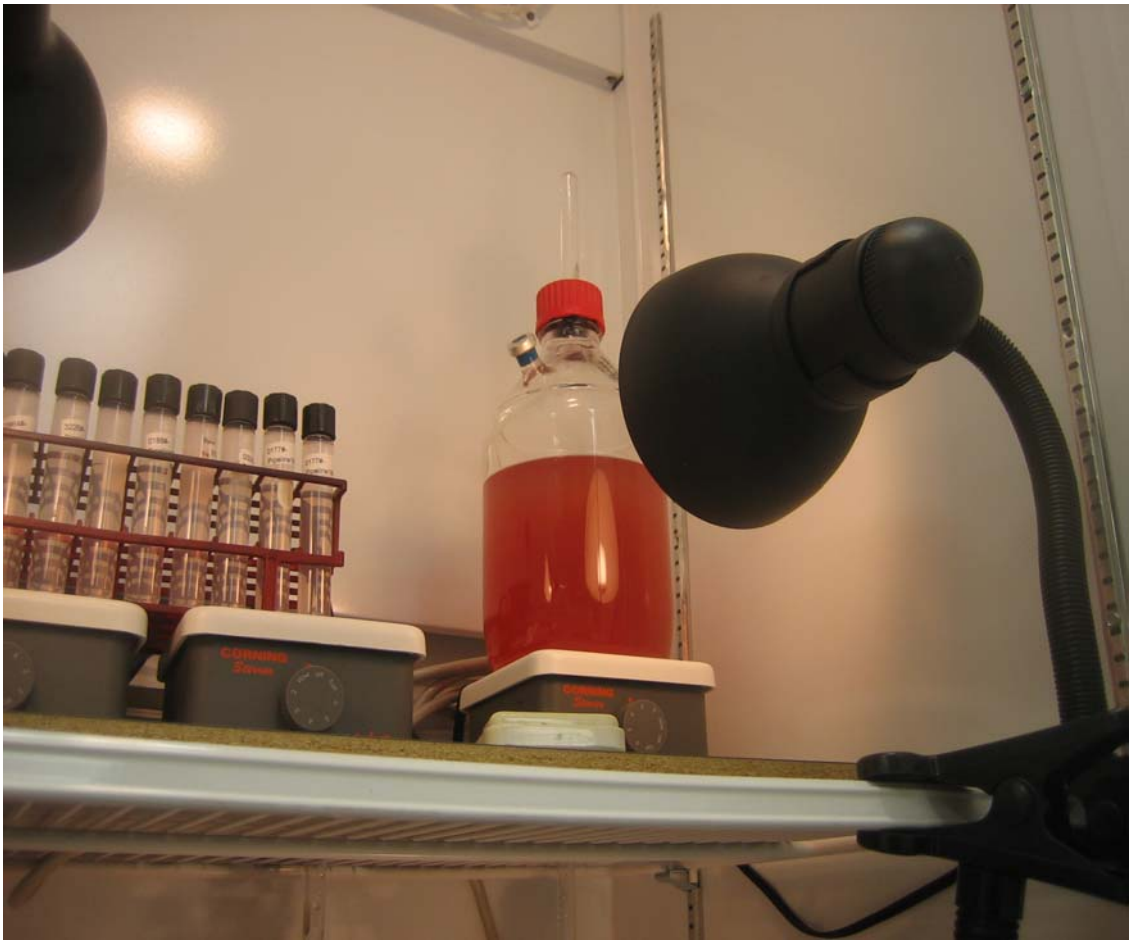


**Figure 2.2.** Growth of *R. palustris* cell cultures. To the left, 25 mL tubes containing cultures of *R. palustris* are shown. A larger 1.5 L production culture of *R. palustris* is shown to the right.

In Chapter 4, dual-affinity isolation is used to enrich the RNA polymerase protein complex from *E. coli* and *R. palustris*. The methods used for isolating this protein complex were adapted from Butland *et al*. [15] and described in more detail in Chapter 3. This chapter also features comparative proteome profiling of a whole cell lysate protein fraction. The whole cell lysate fraction was isolated using a standard protocol developed at ORNL and described in VerBerkmoes *et al.*[43].

In Chapter 5, periplasmic isolate was obtained from *E. coli* K12 using cold osmotic shock protocols [44, 45]. A whole cell lysate fraction was isolated using a standard protocol developed at ORNL and described in VerBerkmoes *et al.*[43].

## *2.4 Sample Preparation*

Methods of sample preparation include measurement of protein quantity, proteolytic digestion of protein isolate with trypsin, and solid phase extraction of peptides prior to LC-MS-MS analysis. As Chapter 3 is a comparison of proteolysis protocols itself, the sample preparation steps are described there in more detail. In Chapter 4, affinity-isolated protein complexes were digested according to the 80% $CH_3CN$ protocol described in Chapters 3 and 4. Complex protein mixtures containing whole cell lysate fractions described in Chapters 4 and 5 were prepared using a standard protocol developed at ORNL and described in VerBerkmoes *et al.* [43].

## *2.5 LC-MS-MS Analysis of Targeted Protein Isolates*

In Chapter 3, LC-MS-MS analysis of ribosomal and MAP protein isolates is performed by 1D reverse phase $C_{18}$ nanoHPLC coupled to a 3D quadrupole ion trap mass spectrometer, the ThermoFinnigan LCQ DecaXP [17]. This specific platform was selected due to the characteristics of the two protein samples studied in this chapter. Both

the ribosomal and MAP protein isolates are of low complexity, representing a simpler mixture of proteins relative to the complexity of a proteome. Also, the quantities of sample processed were also small (2.5 μg protein). Thus, proteolytic digestion of these samples produced peptide mixtures which were adequately separated by 1D reverse phase nanoHPLC. The LCQ DecaXP instrument was adequate for identification, for these samples were of low complexity and did not require measurement of a large dynamic range.

Chapters 4 and 5 feature quantitative LC-MS-MS analyses performed on a linear quadrupole ion trap mass spectrometer, theThermoFinnigan LTQ [18]. In Chapter 4, measurements of affinity-isolated protein complexes were performed using the 1D reverse phase nanoHPLC (as described above), as these protein samples were of low complexity. The LTQ was used for quantitative measurement of affinity-isolated protein complexes due to the instrument's rapid scanning speed. This allowed acquisition of more full MS scans to detect peptide chromatographic features necessary for quantification by the ProRata program [41]. The quantitative proteome profiles measured in Chapters 4 and 5 were obtained by 2D nanoHPLC (strong cation exchange and reverse phase or MudPIT) [26] coupled to the LTQ instrument. These samples were far more complex peptide mixtures, thus they required more extensive chromatographic separation. This complexity resulted in measurements of an increased dynamic range. By design, the LTQ instrument is more adequately suited for such measurements, as its linear ion trap may accumulate a greater number of ions relative to the LCQ [22, 23]. This permits measurement of peptide ions of lower abundance in the sample, likely representing proteins of low abundance. As described above, the rapid full MS scans

performed on the LTQ were important for accurate peptide quantification, permitting collection of chromatographic features (or peaks).

## *2.6 Proteome Informatics and Data Analysis*

Tandem mass spectrometry data were searched with the SEQUEST algorithm [9, 10] for peptide identification. Sources and descriptions of the FASTA-formatted protein sequences (or "databases") used are provided in each chapter. Filtering, sorting, and organization of peptide identifications were performed by the DTASelect program [27] as described in each chapter.

Within the framework of the general experimental design shown in Figure 2.1a, this process is time consuming. In Chapters 4 and 5, the use of $^{15}$N-labeling necessitates a second SEQUEST search of mass spectrometry data for identification of $^{15}$N-labeled peptides. This effectively doubles the computing resources and time required for data analysis. During the course of this dissertation, an automated method of data analysis using SEQUEST and DTASelect was developed by Dr. W. Hayes McDonald at ORNL. This method consisted of a series of Perl scripts (GitRSeq.pl and GitRSeq1.pl) for automation and featured use of a more compact series of data file formats [28]. These series of Perl scripts were modified for use under Unix/Linux (or *nix) operating systems (Appendix C) to decrease search time and reduce use of shared computer resources, such as Windows-based workstations. Peptide and protein quantification was performed by the ProRata program [41] as described in Chapters 4 and 5 in more detail.

In general, all chapters employed further data analysis of tab-delimited output from the DTASelect and ProRata programs. Microsoft Access was used for further data analysis, as described in each specific chapter.

# Chapter 3
# Comparison of Digestion Protocols for Microgram Quantities of Enriched Protein Samples

All of the data presented in this chapter have been adapted from the following published journal article:

W. Judson Hervey, IV, Michael Brad Strader, and Gregory B. Hurst. Comparison of Digestion Protocols for Microgram Quantities of Enriched Protein Samples. *Journal of Proteome Research.* 2007, 6 (8), 3054-3061.

*B. taurus* microtubule associated protein (MAP) isolates were a generous gift from the laboratory of Dr. Timothy J. Mitchison, Department of Systems Biology, Harvard Medical School.

*R. palustris* cell growth and isolation of the 70***S*** ribosomal protein complex was conducted in the laboratory of Dr. Dale A. Pelletier by W. Judson Hervey, IV. Sample preparation of all protein isolates, LC-MS-MS measurements, data analysis, and responses to peer review was performed by W. Judson Hervey, IV.

Supplementary Figures S1-S6 referred to in the Results and Discussion section are available freely on the internet at http://pubs.acs.org/doi/abs/10.1021/pr070159b .

## *Abstract*

Standard biochemical techniques that are used for protein enrichments, such as affinity isolation and density gradient centrifugation, frequently yield high nanogram to low microgram quantities at a significant expenditure of resources and time. The characterization of selected protein enrichments by the "shotgun" mass spectrometry approach is often compromised by the lack of effective and efficient in-solution proteolysis protocols specifically tailored for these small quantities of proteins. This study compares the results of five different digestion protocols that were applied to 2.5 µg portions of protein isolates from two disparate sources: *Rhodopseudomonas palustris* 70***S*** ribosomal proteins, and *Bos taurus* microtubule-associated proteins (MAPs). Proteolytic peptides produced according to each protocol in each type of protein isolate were analyzed by one-dimensional liquid chromatography-tandem mass spectrometry (LC-MS/MS). The effectiveness of each digestion protocol was assessed on the basis of three parameters: number of peptide identifications, number of protein identifications, and sequence coverage. The two protocols using a solvent containing 80% acetonitrile ($CH_3CN$) for trypsin digestions performed as well as, and in some instances better than, protocols employing other solvents and chaotropes in both types of protein isolates. A primary advantage of the 80% $CH_3CN$ protocol is that it requires fewer sample manipulation steps.

## *3.1 Introduction*

The analysis of protein mixtures by mass spectrometry-based techniques [7] has become possible due to a confluence of advancements in analytical instrumentation,

multi-dimensional chromatographic separations, and bioinformatics. One widely-used proteomics strategy is to perform qualitative analysis of proteolytic peptides by one dimensional (1D) or two dimensional (2D) chromatographic separation coupled via electrospray with tandem mass spectrometry (LC-MS-MS), generating peptide fragmentation spectra that are compared to translated genomic databases [7, 46]. This robust approach has been applied, for example, to identifying proteins from complex mixtures [24], less abundant samples obtained from biochemically-enriched multiprotein complexes [13, 42], and affinity isolations of single proteins along with interacting partners [14, 47-51].

The LC-MS-MS analysis of a protein sample from a biochemical enrichment or affinity isolation presents unique challenges to the workflow of "bottom-up" LC-MS-MS proteomics approaches. The limitation imposed by the small amounts of protein typically available from the biological enrichment requires that all aspects of downstream sample processing be as efficient and effective as possible, including enzymatic proteolysis. Improvements in enzyme activity have been reported when enzymatic reactions are conducted in organic media [52, 53], and more significantly, with trypsin directly in reverse-phase HPLC solvents [54, 55]. Recently, our laboratory has demonstrated the effectiveness of conducting trypsin digestions of protein isolates in a solvent containing 80% acetonitrile ($CH_3CN$), resulting in an increased number of peptide and protein identifications, increased sequence coverage, and more thorough proteolysis for low microgram to high nanogram quantities of standard protein mixtures, ribosomes purified by density gradient centrifugation, and an affinity-isolated protein complex [56].

In the present report, we compare the effectiveness of two digestion protocols performed in 80% $CH_3CN$ [56] with three other protocols commonly used for the digestion of proteins. These protocols include a modified version of the manufacturer's instructions from the sequencing-grade modified trypsin [57, 58], a sequential digestion with Lys-C and trypsin [59, 60], and a protocol utilizing the acid-labile surfactant (ALS) RapiGest™ SF [61, 62]. The enzymatic proteolysis of 2.5 μg of protein, a yield representative of biochemical protein enrichment or affinity isolation experiments, was performed on two distinct types of protein isolates: the 70***S*** ribosome of *Rhodopseudomonas palustris* [42], and microtubule-associated proteins (MAPs) [59, 63] of the cow, *Bos taurus*, using the digestion protocols described above. The two protocols employing the 80% $CH_3CN$ solvent for trypsin digestions performed as well as or better than the other protocols tested in this study. In as little as one hour, trypsin digestions in 80% $CH_3CN$ yielded either comparable or more numerous tryptic peptide identifications and increased sequence coverage in both the 70***S*** ribosomal isolates and the MAPs.

## *3.2 Experimental Procedures*

Reagents obtained from Sigma Chemical Company (St. Louis, MO), were: Trizma Hydrochloride, Trizma Base, iodoacetamide (IAA, Ultrapure), hydrochloric acid (molecular biology grade), and urea (Ultrapure). Other materials used in this study include: lyophilized sequencing-grade modified trypsin (Promega, Madison, WI); endoproteinase Lys-C (Roche Applied Science, Penzberg, Germany); formic acid (EM Science SupraPur), calcium chloride (JT Baker, 99% purity), HPLC-grade acetonitrile and water (Burdick and Jackson, Muskegon, MI); tris(2-carboxyethyl)phosphine hydrochloride (TCEP) and micro bicinchoninic acid (BCA) assay reagent (Pierce,

Rockford, IL). RapiGest™ SF was obtained from Waters (Milford, MA). Ultrapure 18 MΩ water used for dialysis of the protein samples was obtained from a Milli-Q system (Millipore, Bedford, MA). HPLC-grade water was used for all sample buffers. Nonstick microcentrifuge tubes were obtained from VWR International (West Chester, PA).

**Preparation of protein isolates.** The 70***S*** ribosomal protein complex was enriched from *R. palustris* CGA010 by sucrose density fractionation, as described previously [42]. Microtubule-associated proteins were a gift from the Mitchison laboratory (Department of Systems Biology, Harvard Medical School) [59]. Protein isolates were dialyzed for 16 h against Ultrapure water in a 3500 MWCO Slide-A-Lyzer dialysis cassette (Pierce, Rockford, IL), quantified using the micro BCA assay (Pierce), aliquotted into 2.5 μg portions, and concentrated to a volume of approximately 10 μL by centrifugal evaporation in a SpeedVac (Savant).

**Protein Digestion by Enzymatic Proteolysis.** Digestions of 2.5 μg portions of 70***S*** ribosomal isolates and 2.5 μg portions of MAPs were performed in duplicate according to five different protocols (Figure 3.1 and Table 3.1). For all trypsin digestions, trypsin was reconstituted at a concentration of 20 ng/μL in 50 mM Trizma-HCl 10 mM $CaCl_2$ at pH 7.6 and added to each protein isolate at a trypsin to protein ratio of 1:10 (wt/wt) in a

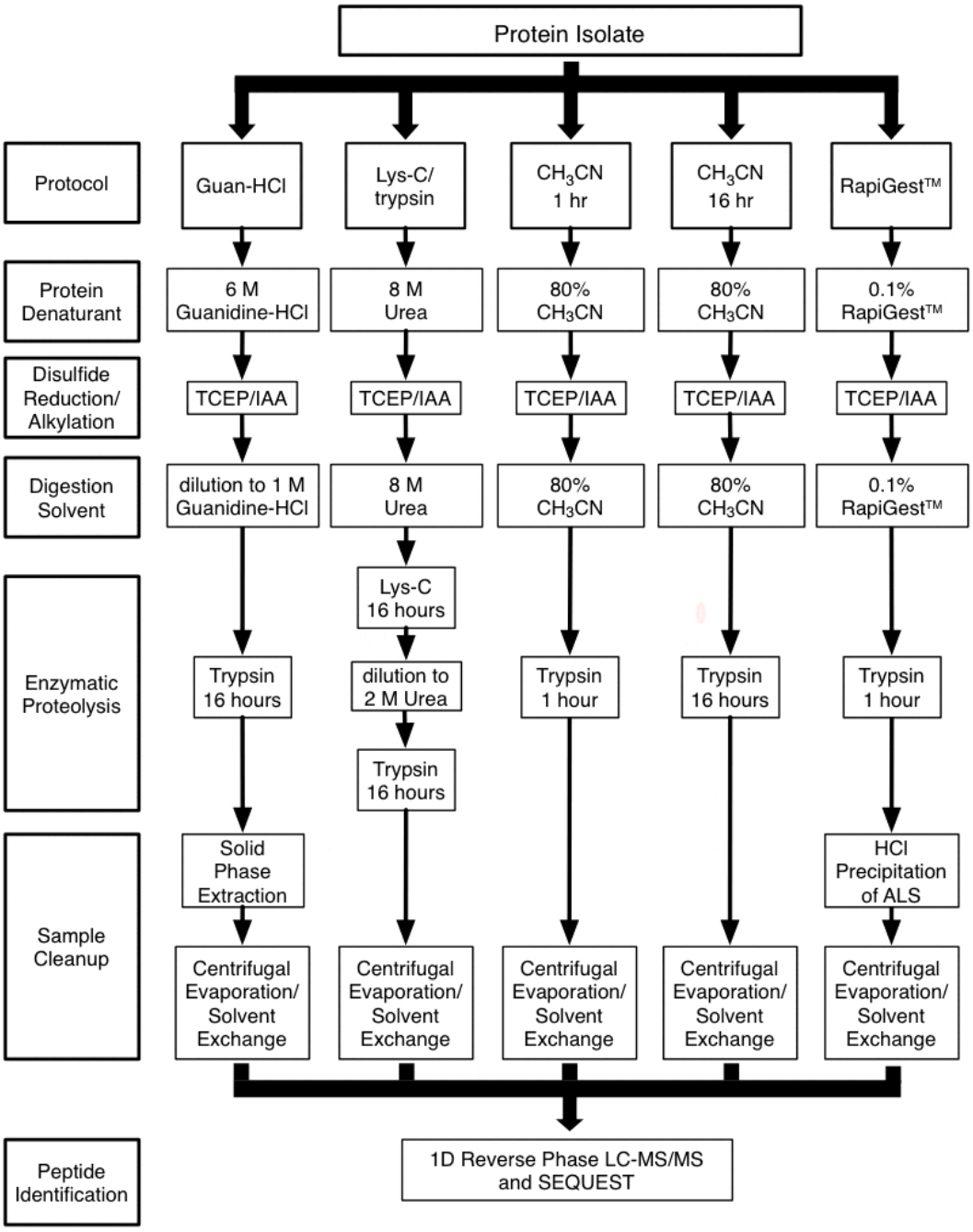



**Figure 3.1.** Flowchart representing the digestion protocols compared in this study.

**Table 3.1.** Solvents and Incubation Times for the Digestion Protocols.

| **Protocol Designation** | **Solvent/Buffer for Trypsin Digestion** | **Incubation Time at 37°C** |
|---|---|---|
| Guan-HCl | 1 M guanidine hydrochloride, 50 mM Trizma-HCl, 10 mM $CaCl_2$ pH 7.6 | 16 hours |
| Lys-C/ trypsin | 2 M urea, 100 mM Trizma-HCl pH 8.5 | 32 hours |
| $CH_3CN$ 1 hr | 80% $CH_3CN$, 20% 50 mM Trizma-HCl 10 mM $CaCl_2$ pH 7.6 (v/v) | 1 hour |
| $CH_3CN$ 16 hr | 80% $CH_3CN$, 20% 50 mM Trizma-HCl 10 mM $CaCl_2$ pH 7.6 (v/v) | 16 hours |
| RapiGest™ | 0.1% RapiGest™ SF, 50 mM $NH_4HCO_3$ (w/v) | 1 hour |

nonstick microcentrifuge tube (VWR). Following each trypsin digestion, peptides were concentrated by centrifugal evaporation (SpeedVac) to a volume of ~30 μL, diluted HPLC solvent A (95% $H_2O$ 5% $CH_3CN$ 0.1% formic acid) to a volume of 123 μL, and stored at -80ºC until LC-MS-MS analysis. Duplicate LC-MS-MS analyses of each duplicate digestion were performed, providing a total of 4 data sets for each combination of sample type and digestion protocol.

**Guanidine hydrochloride denaturation and 16 hour trypsin digestion (*Guan-HCl*) [57, 58].**

Protein isolates were suspended in 90 μL of 6 M guanidine hydrochloride dissolved in 50 mM Trizma-HCl, pH 7.6, heated at 60ºC for 1 hr, and cooled to laboratory temperature (approximately 25ºC). Disulfide bonds were reduced with the addition of TCEP to a concentration of 2 mM and incubated at ambient laboratory temperature (approximately 25ºC). Disulfide bonds were reduced with the addition of TCEP to a concentration of 2 mM and incubated at ambient laboratory temperature for 20 min. Cysteine residues were carboxyamidomethylated with the addition of IAA to a concentration of 10 mM and incubated at ambient laboratory temperature for 15 min in the dark. The concentration of guanidine hydrochloride was diluted to 1 M with the addition of 50 mM Trizma-HCl 10 mM $CaCl_2$ at pH 7.6. Proteolysis was conducted with trypsin for 16 hr at 37ºC. Following digestion, samples were concentrated by centrifugal evaporation to a total volume of 100 μL. Solid phase extraction (SPE) was performed using 100 μL C18 OMIX Tips (Varian) and extracted peptides were concentrated by centrifugal evaporation and dilution in Solvent A as detailed above.

**Urea denaturation, 16 hour Lys-C digestion, and 16 hour trypsin digestion (*Lys-C/trypsin*) [59, 60]**. Protein isolates were suspended in 35 μL of 8 M urea dissolved in 100 mM Trizma-HCl, pH 8.5. Disulfide bonds were reduced and cysteine residues alkylated as described in the *Guan-HCl* protocol. In the first proteolysis reaction, 100 ng of Lys-C was added to each protein isolate and incubated at 37ºC for 16 hr. Following proteolysis with Lys-C, the urea concentration was diluted to 2 M urea using 100 mM Trizma-HCl at pH 8.5 and proteolysis with trypsin was conducted at 37ºC for 16 hr.

**Organic solvent and 1 hour trypsin digestion (*$CH_3CN$ 1 hr*) [56]**. Protein isolates were suspended in 80% $CH_3CN$, 20% 50 mM Trizma-HCl, 10 mM $CaCl_2$ at pH 7.6 to a volume of 100 μL. Disulfide bonds were reduced and cysteine residues alkylated as described in the *Guan-HCl* protocol. Proteolytic digestion was performed with trypsin at 37ºC for 1 hr.

**Organic solvent and 16 hour trypsin digestion (*$CH_3CN$ 16 hr*) [56].** All steps were performed identically to the *$CH_3CN$ 1 hr* protocol except proteolysis, which was extended to 16 hr at 37ºC.

**Acid-labile surfactant and 1 hour trypsin digestion (*RapiGest*™) [61, 62]**. Protein isolates were suspended in 90 μL of 0.1% (w/v) RapiGest™ SF (Waters) 50 mM $NH_4HCO_3$ and equilibrated at 37ºC for 2 min. Disulfide bonds were reduced and cysteine residues alkylated as described in the *Guan-HCl* protocol. Proteolytic digestion was performed with trypsin at 37ºC for 1 hr. Protein isolates were acidified by addition of 20 μL of 500 mM HCl, incubated at 37ºC for 45 min, and centrifuged at 13,000 *g* for 10 min to precipitate the RapiGest™ SF. The supernatant peptides were removed and concentrated as described above.

**Liquid Chromatography and Tandem Mass Spectrometry (LC-MS-MS).** Mass spectrometry data were collected using an LCQ Deca XP Plus quadrupole ion trap mass spectrometer (ThermoFinnigan, San Jose, CA) equipped with a nanospray source, coupled to a one dimensional reverse phase HPLC system (LCPackings/Dionex), as described previously [56]. Briefly, 50 μL of each digestion, representing approximately 1.2 μg of digested protein isolate, was loaded onto the reverse phase HPLC system by a Famos autosampler (LCPackings/Dionex), concentrated on a C18 reverse phase trapping cartridge (300 μm x 5 mm, 300 Å PepMap, LCPackings/Dionex), and desalted for 10 min with solvent A at a flow rate of 30 μL/min, pumped by a Switchos II unit (LCPackings/Dionex). By means of a switching valve on the Switchos, a lower solvent flow (see below) was then directed in the opposite direction across the trapping cartridge, backflushing the trapped peptides onto a C18 reverse phase analytical column. The analytical column consisted of an 18 cm length of Jupiter C18 reverse phase resin with 5 micron diameter particles and 300 Å pore size (Phenomenex) packed into a 100 μm ID PicoTip Emitter (New Objective) using a pressure cell (New Objective). Peptides were eluted from the reverse phase column by an Ultimate HPLC pump (LCPackings/Dionex) at a total flow rate of 150 μL/min, with flow splitting provided by a NAN-75 calibrator, for a column flow rate of ~250 nL/min. The gradient elution consisted of a linear ramp over 130 min from 100% solvent A to 50% solvent A/50% solvent B (30% $H_2O$ 70% $CH_3CN$ 0.1% formic acid), followed by a 1 min ramp to, and 19 min hold at, 100% solvent C (5% $H_2O$ 95% $CH_3CN$ 0.1% formic acid). The mass spectrometer was operated in data dependent mode, in which the four most intense ions in the precursor ion scan within the mass-to-charge range of 400 – 2000 Da were subjected to tandem mass

spectrometry. Dynamic exclusion [64] was enabled on the instrument with the repeat count set at two, as described previously [42, 56].

**Peptide Identification and Data Analysis.** Identification of peptides from tandem mass spectra was performed using the SEQUEST algorithm (version 27) [65] without enzyme specificity. Databases used in the SEQUEST analysis were the predicted proteomes of *R. palustris* [43, 66] and *B. taurus* [67] for ribosomal and MAP protein isolates, respectively, FASTA protein sequences for *R. palustris* were obtained from the ORNL microbial genome annotation pipeline [68] and the protein sequences for *B. taurus* (IPI Cow 3.09) [67] were obtained from the European Bioinformatics Institute (EBI). FASTA protein sequences of common contaminants (e.g. keratins), protein standard mixture components (e.g. myoglobin, hemoglobin), and proteolytic enzymes (trypsin, lys-c) were concatenated to the predicted proteomes of each organism, resulting in a total of 45,843 proteins for *B. taurus* and 4,890 proteins for *R. palustris*. A dynamic modification of 57.0 Da was specified for carboxyamidomethylation of cysteine residues.

DTASelect (version 2.06) [69] filtered peptide identifications from each search results file [28] according to the default cross-correlation ($X_{Corr}$) scores ($\geq$1.8, $\geq$2.5, and $\geq$3.5 for the 1+, 2+, and 3+ charge state, respectively) and $\Delta Cn \geq 0.08$. Peptide identifications above these score thresholds were assembled by locus into protein identifications, requiring a minimum of two peptide identifications per protein for a positive protein identification. Redundant locus information was discarded by DTASelect. Peptide and protein identification data of replicate chromatographic runs were exported from DTASelect with the organizer.pl Perl script [70], modified to use the thresholds quoted above, and organized for further analysis in a relational database (Microsoft Access).

## *3.3 Results and Discussion*

The characterization of protein enrichments by mass spectrometric identification of peptides depends critically on an effective and efficient proteolysis protocol that is specifically tailored for low microgram to high nanogram quantities of protein. This report compares the effectiveness of several protocols for the digestion of protein isolates of this nature. To this end, 2.5 μg portions of two isolates, the *R. palustris* 70***S*** ribosomal protein complex and *B. taurus* microtubule-associated proteins (MAPs), were digested according to five different digestion protocols, shown in Figure 1. The five digestion protocols that we studied encompass a variety of protein denaturation methods, digestion solvents (see Table 3.1), proteolytic enzymes, and incubation periods. These parameters were not comprehensively varied, as described in previous studies [71]; rather, we sought to implement the digestion protocols as described in technical or manufacturer's literature directly. The comparison among the various protocols is based on numbers of identified peptides and proteins, and sequence coverage for proteins.

The two sets of proteins chosen for this study are distinct in their biological origin, sample complexity, and physical characteristics. First, the 70***S*** bacterial ribosome is isolated from a prokaryote, *R. palustris*, whereas MAP isolates are from the cow, a eukaryotic organism. Figure 3.2 further illustrates the differences between the two sets of proteins with respect to sample complexity and the physical characteristics of molecular

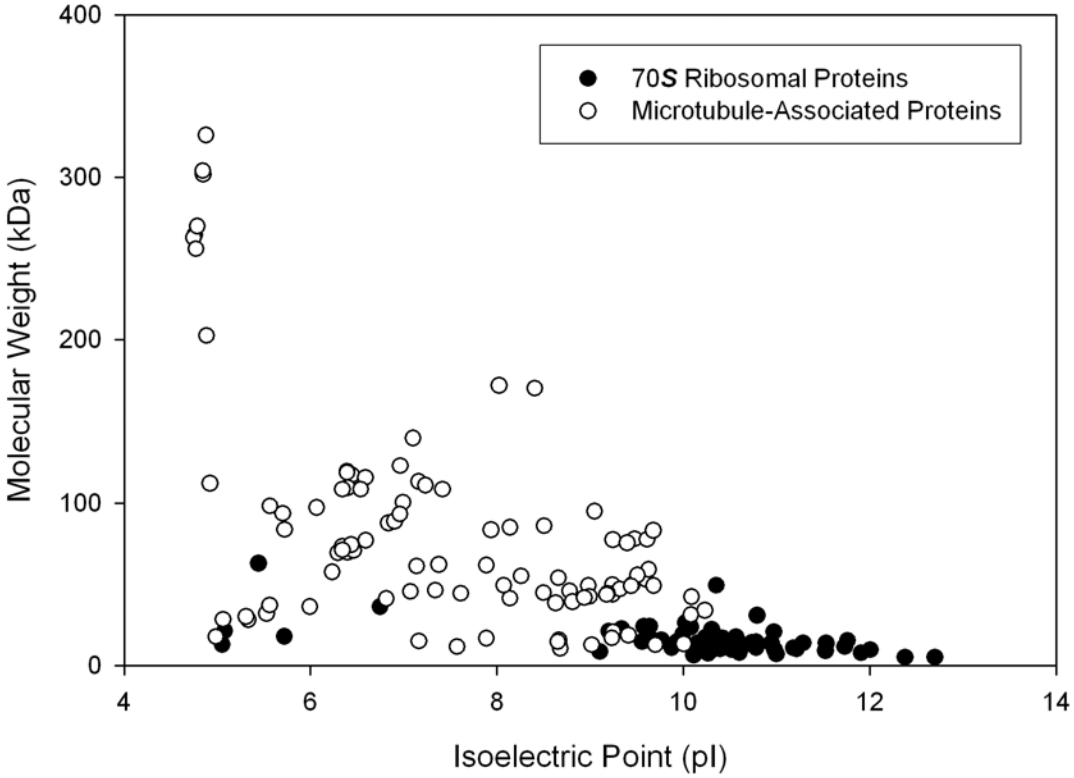



**Figure 3.2.** Physical characteristics of proteins annotated as *R. palustris* ribosomal proteins or *B. taurus* MAPs.

weight and isoelectric point (p[10]). The more complex MAP isolates cover a wider molecular weight range than ribosomal proteins; isoforms of the lighter tau protein are between 55-62 kDa, while the larger MAPs range from 180 to 350 kDa. Additionally, MAP pIs range from <5 to >10. Most ribosomal proteins are of low molecular weight, and, with the exception of a few acidic proteins, have high pIs, as Figure 2 shows. As the 70***S*** ribosome functions within the cytoplasm of the prokaryotic cell, the protein component of the complex is soluble at physiological pH. Thus, the different characteristics of these two protein samples provided a diverse test set for comparing the various digestion protocols.

***R. palustris* 70*S* ribosomal isolates.** The ribosome is a ribonucleic acid-protein complex that functions in the synthesis of proteins. This large macromolecular protein complex is readily sedimented away from other cellular components by sucrose density fractionation. With an estimated 20,000 copies per cell [72], the 70***S*** ribosome of eubacteria has been chosen to be characterized by LC-MS-MS-based approaches due to its intracellular abundance, relative ease of isolation by existing biochemical techniques, and biological significance [42].

Table 3.2 shows the average number of peptide identifications from digestions of ribosomal protein isolates. The *$CH_3CN$ 1 hr* protocol produced both the most total peptide identifications and the most tryptic peptide identifications of any protocol tested, followed by the *$CH_3CN$ 16 hr*, *Guan-HCl*, *RapiGest*, and *Lys-C/trypsin* protocols, in that order. Extension of the digestion in 80% $CH_3CN$ to 16 hours resulted in a number of peptide identifications similar to the 1 hour digestion. In accord with other studies [60],

**Table 3.2.** Average Peptide Identifications from *R. palustris* 70***S*** Ribosomal Protein Isolates by Digestion Protocol[a]

| Digestion Protocol | Tryptic[a] Peptides | Semi-tryptic Peptides[b] | Non-Specific Peptides[c] | Total Peptide Identifications[d] | Percent Tryptic[e] | 3-4 Missed Cleavages[f] |
|---|---|---|---|---|---|---|
| **Guan-HCl** | 293 ± 19 | 37 ± 2 | 0.3 ± 0.5 | 331 ± 21 | 88.3 – 89.0 | 3 ± 1 |
| **Lys-C/ trypsin** | 203 ± 16 | 36 ± 16 | 6.5 ± 1.3 | 245 ± 29 | 78.8 – 89.3 | 14 ± 3 |
| **$CH_3CN$ 1 hour** | 391 ± 33 | 66 ± 9 | 1.3 ± 1.3 | 458 ± 26 | 81.5 – 87.4 | 40 ± 8 |
| **$CH_3CN$ 16 hours** | 380 ± 49 | 71 ± 23 | 2.5 ± 1.7 | 453 ± 47 | 75.7 – 88.1 | 0 |
| **RapiGest** | 261 ± 27 | 46 ± 5 | 0.3 ± 0.5 | 307 ± 32 | 84.9 – 85.4 | 6 ± 1 |

Average values with standard deviation are shown from replicate experiments (see text for details). [a]Peptides with two termini resulting from cleavage C-terminal to lysine or arginine residues. [b]Peptides with one terminus resulting from cleavage C-terminal to a lysine or arginine residue. [c]Peptides with neither terminus resulting from cleavage C-terminal to a lysine or arginine residue. [d]Total peptide identifications regardless of enzyme specificity. [e]Percentage of total peptide identifications with two tryptic termini. [f]Tryptic peptides containing either three or four missed trypsin cleavage sites.

the extension of digestion times beyond 1 hour provided no increase in the number of peptide identifications. In fact, the *Lys-C/trypsin* protocol, in which proteolysis was conducted for the longest total time (32 hours), produced the fewest peptide identifications in this sample. Although the *Guan-HCl* and *$CH_3CN$ 16 hr* protocols each had incubation periods of 16 hours, approximately 100 more tryptic peptides were detected in the latter (Table 2). It is possible that the *Guan-HCl* and *Lys-C/trypsin* protocols did not yield as many peptides as the *$CH_3CN$ 16 hr* protocol due to the combined effects of protein or peptide losses to sample tube surfaces from the chaotrope dilution. Furthermore, it is also possible that SPE was an additional source of peptide loss for the *Guan-HCl* protocol. Base peak chromatograms for one run from each digestion protocol are shown in Figure S1 (Supporting Information).

An examination of the physical characteristics of tryptic peptide identifications from each digestion protocol revealed that the majority of peptide identifications were distributed over a range of 800-1800 Da with p*I* values ranging from 3 to 12, regardless of digestion protocol (Figure S2, Supporting Information).

Table 3.3 lists the number of ribosomal and non-ribosomal protein identifications from each digestion protocol. ANOVA indicates that a null hypothesis that the five digestion protocols yield similar numbers of protein identifications can be rejected at a >99% confidence level; the *$CH_3CN$ 1 hr* protocol led on average to identification of the largest number of ribosomal proteins, and the *Lys-C/trypsin* protocol led to the smallest. The 40 to 49 ribosomal protein identifications, along with a low number of co-sedimenting non-ribosomal proteins, are typical of LC-MS-MS analysis of tryptic digests of ribosomal

**Table 3.3.** Protein Identifications from *R. palustris* 70***S*** Ribosomal Protein Isolates by Digestion Protocol

| **Digestion Protocol** | **Ribosomal / Non-Ribosomal[a]** |
|---|---|
| | 46 / 2 |
| **Guan-HCl** | 49 / 1 |
| | 46 / 2 |
| | 46 / 2 |
| | 41 / 1 |
| **Lys-C/trypsin** | 44 / 0 |
| | 40 / 1 |
| | 40 / 0 |
| | 47 / 3 |
| **$CH_3CN$ 1 hour** | 48 / 3 |
| | 49 / 4 |
| | 49 / 3 |
| | 45 / 1 |
| **$CH_3CN$ 16 hours** | 45 / 6 |
| | 47 / 4 |
| | 47 / 5 |
| | 44 / 2 |
| **RapiGest™** | 40 / 2 |
| | 45 / 3 |
| | 44 / 4 |

[a]Based on functional annotation of the protein.

proteins isolated by sucrose density fractionation. In our previous report, a single 110-minute capillary one dimensional reverse phase LC-MS-MS separation of 10 μg trypsin digestions of ribosomal isolates yielded the identification of 186 peptides matching 41 *R. palustris* proteins [42].

Figure 3.3 compares the digestion protocols in terms of the proportion of identifications from 70***S*** ribosomal protein isolates with average sequence coverage greater than 30%. The *$CH_3CN$ 1 hr* protocol produced the largest proportion of protein identifications with greater than 30% coverage, followed by the *Guan-HCl*, *$CH_3CN$ 16 hr*, *RapiGest*, and *Lys-C/trypsin* protocols, in that order. The *RapiGest* and *Lys-C/trypsin* protocols resulted in ≥30% sequence coverage for less than half of the identified proteins. Figure S4 in Supporting Information summarizes sequence coverage and physical characteristics for a total of 53 70*S* ribosomal proteins identified via any digestion protocol. The *Guan-HCl* protocol yielded the most protein identifications with the highest sequence coverage followed by the *$CH_3CN$ 1 hr* protocol; together, these two protocols provided the highest sequence coverage for 46 of the 53 proteins. Figures S5 and S6 (Supporting Information) show the distribution of sequence coverage values of proteins over 4 p*I* ranges and 3 molecular weight ranges.

For the 70***S*** ribosomal isolates, the *$CH_3CN$ 1 hr* protocol produced the highest number of peptide identifications and ribosomal protein identifications, and resulted in the highest

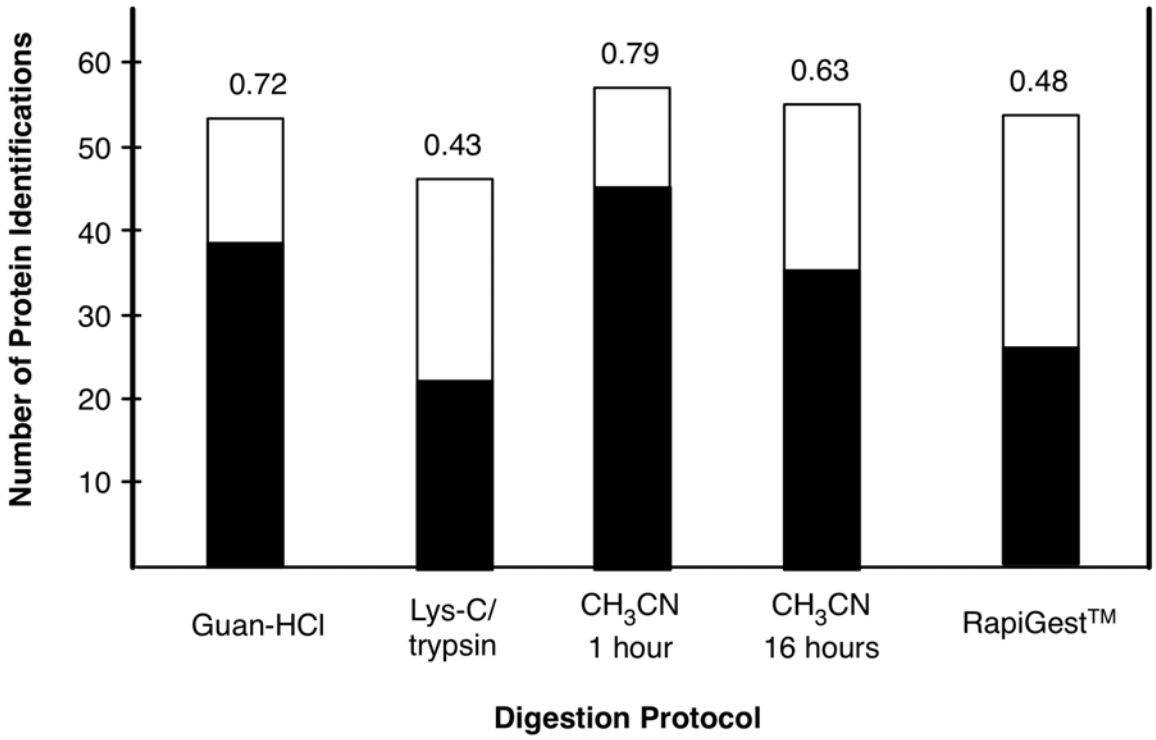



**Figure 3.3.** Proportion of Identifications from *R. palustris* 70***S*** Ribosomal Protein Isolates with average sequence coverage greater than 30% by digestion protocol. The height of each bar indicates the total number of protein identifications from tryptic peptides by digestion protocol. The solid portion of each bar indicates the number of proteins with an average sequence greater than 30%; the proportional value is indicated at the top of each bar. Numbers of total protein identifications accumulate all proteins identified in any replicate measurement, and therefore appear larger than suggested by Table 3.

proportion of protein identifications with sequence coverage values greater than 30%. The effectiveness of this protocol on this sample may be due to one or more of the following factors. First, the 80% acetonitrile solvent need not be diluted prior to trypsin digestion, while dilution of the chaotrope in the *Guan-HCl* and *Lys-C/trypsin* protocols is required to avoid inactivating trypsin. Such an increase in reaction volume for the digestion results in more solvent assessable tube surface area, which increases the potential for adsorptive loss of proteins or peptides along the inside of a centrifuge tube. Second, it is plausible that trypsin activity is affected in different ways by the various solvents used in these protocols, which span a range of chaotropes, pH values, and water concentrations. The 80% acetonitrile solvent may promote protease activity, as demonstrated in other mixed aqueous-organic solvents [53, 73]. Finally, different sample treatments for downstream compatibility with liquid chromatography are required. While the 80% acetonitrile solvent may easily be removed by centrifugal evaporation, the *Guan-HCl* and *RapiGest* protocols that we followed [42, 57, 58],[61, 62] both specify additional sample processing steps (SPE and acid precipitation, respectively) to remove the chaotrope, presenting an additional potential mechanism for loss of peptides.

***B. taurus* microtubule-associated protein (MAP) isolates.** As a subset of microtubule-binding proteins, MAPs govern the structural dynamics of the cytoskeleton by assembling and stabilizing tubulin heterodimers into microtubules, facilitating essential cellular processes such as mitosis [72]. The subset of characterized MAPs includes the lower molecular weight tau protein and the larger MAP1, MAP2, and MAP3 proteins, with each subspecies present in distinct isoforms as a result of alternative mRNA splicing. The tau and MAP2 proteins are principally expressed in neuronal tissue,

while other MAPs have been identified and are distributed throughout other tissues and cell types in the body [74]. The subset of neuronal MAPs are of considerable biological interest, as these proteins have been linked to the formation of cytoskeletal neurofibrillary tangles, a defining characteristic of Alzheimer's disease [63]. The methodology to isolate MAPs from mammalian brain tissue involves at least two cycles of temperature-dependent depolymerization of tubulin proteins followed by phosphocellulose chromatography of the resulting tubulin-enriched fraction, yielding enriched amounts of the higher molecular weight MAPs and some residual amounts of tubulins [74, 75].

Table 3.4 shows the average number of peptide identifications from digestions of 2.5 µg of MAPs according to the five digestion protocols. The *$CH_3CN$ 16 hour* protocol yielded the highest number of both tryptic peptide and total peptide identifications, followed by the *Lys-C/trypsin*, the *$CH_3CN$ 1 hour*, *RapiGest*, and *Guan-HCl* protocols. (It should be noted that we observed a 20 minute delay in detection of major chromatographic peaks during one technical replicate of the *$CH_3CN$ 1 hour* protocol,

**Table 3.4.** Average Peptide Identifications from *B. taurus* MAP Isolates by Digestion Protocol[a]

| Digestion Protocol | Tryptic[a] Peptides | Semi-tryptic Peptides[b] | Non-Specific Peptides[c] | Total Peptide Identifications[d] | Percent Tryptic[e] | 3-4 Missed Cleavages[f] |
|---|---|---|---|---|---|---|
| **Guan-HCl** | 60 ± 23 | 5 ± 2 | 3.0 ± 1.4 | 67 ± 24 | 83.7 – 91.4 | 8 ± 4 |
| **Lys-C/ trypsin** | 296 ± 44 | 71 ± 18 | 10.0 ± 4.1 | 377 ± 65 | 76.0 – 81.0 | 1 ± 1 |
| **$CH_3CN$ 1 hour** | 283 ± 137 | 51 ± 19 | 5.5 ± 3.0 | 339 ± 156 | 72.6 – 81.0 | 16 ± 3 |
| **$CH_3CN$ 16 hours** | 427 ± 48 | 75 ± 16 | 3.3 ± 1.5 | 505 ± 64 | 83.1 – 87.0 | 4 ± 4 |
| **RapiGest** | 281 ± 73 | 65 ± 25 | 0.0 ± 0.0 | 346 ± 97 | 79.4 – 86.3 | 13 ± 5 |

Average values with standard deviation are shown from replicate experiments (see text for details). [a]Peptides with two termini resulting from cleavage C-terminal to lysine or arginine residues. [b]Peptides with one terminus resulting from cleavage C-terminal to a lysine or arginine residue. [c]Peptides with neither terminus resulting from cleavage C-terminal to a lysine or arginine residue. [d]Total peptide identifications regardless of enzyme specificity. [e]Percentage of total peptide identifications with two tryptic termini. [f]Tryptic peptides containing either three or four missed trypsin cleavage sites.

Leading to the large standard deviation in the number of peptide identifications.) The numbers of peptide identifications covered a larger range for the 5 digestion protocols applied to the MAPs than for the ribosomal proteins, perhaps reflecting the greater dynamic range and the total number of proteins of the MAPs. There was no clear trend relating digestion time and number of peptide identifications; the two 16-hour digestions had the highest and lowest number of peptide identifications, while the *RapiGest*, *$CH_3CN$ 1 hr*, and *Lys-C/trypsin* protocols all yielded similar numbers of peptide identifications. Incubation times of 16 hours or longer resulted in fewer missed cleavage sites within peptide identifications relative to those protocols where incubation was conducted for only 1 hour. This is significant, as the *Lys-C/trypsin* protocol requires more digestion time and total number of reagents. Digestion of the MAPs according to the *Guan-HCl* protocol resulted in by far the fewest peptide identifications, perhaps due to protein or peptide losses caused by dilution of the guanidine hydrochloride, and SPE, as described above.

Table 3.5 lists the numbers of MAP and non-MAP protein identifications from each digestion protocol. ANOVA indicates that a null hypothesis that the five digestion protocols yield similar numbers of protein identifications can be rejected at a >99% confidence level. The various digestion protocols resulted in the identification of 8 to 19 neuronal MAPs. The *Lys-C/trypsin*, *$CH_3CN$ 1 hour*, *$CH_3CN$ 16 hours*, and *RapiGest* protocols produced comparable numbers of MAP identifications, ranging from 14 to 19, while the number of identifications for the *Guan-HCl* protocol ranged from 8 to 14. The difference in number of identifications of non-MAP proteins differed more drastically by

**Table 3.5.** Protein Identifications from *B. taurus* MAP Isolates by Digestion Protocol

| Digestion Protocol | MAP / Non-MAP[a] |
|---|---|
| | 9 / 11 |
| **Guan-HCl** | 8 / 9 |
| | 14 / 13 |
| | 13 / 16 |
| | 18 / 78 |
| **Lys-C/trypsin** | 18 / 80 |
| | 18 / 61 |
| | 17 / 77 |
| | 18 / 59 |
| **$CH_3CN$ 1 hour** | 18 / 56 |
| | 19 / 3 |
| | 14 / 55 |
| | 19 / 87 |
| **$CH_3CN$ 16 hours** | 19 / 88 |
| | 18 / 80 |
| | 18 / 74 |
| | 16 / 43 |
| **RapiGest™** | 19 / 65 |
| | 17 / 42 |
| | 17 / 52 |

[a]Based on functional annotation of the protein.

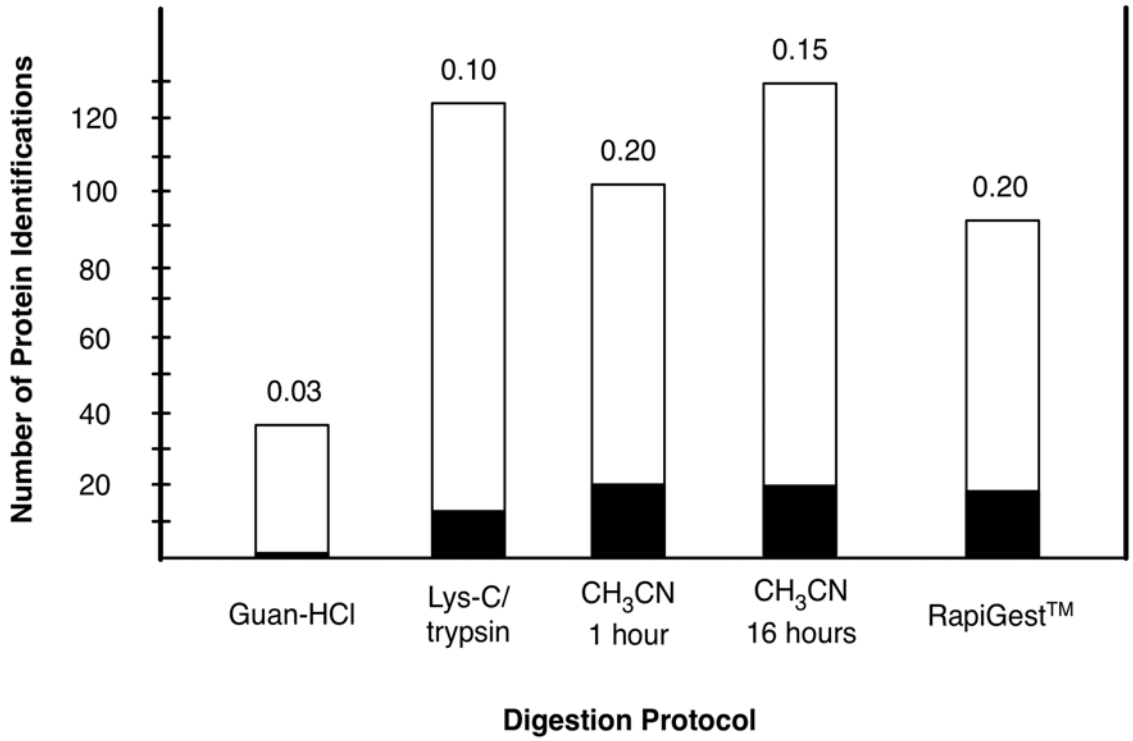



**Figure 3.4.** Proportion of Identifications from *B. taurus* MAP isolates with average sequence coverage values greater than 30% by digestion protocol. The height of each bar indicates the total number of protein identifications from tryptic peptides by digestion protocol. The solid portion of each bar indicates the number of proteins with an average sequence greater than 30%; the proportional value is indicated at the top of each bar. Numbers of total protein identifications accumulate all proteins identified in any replicate measurement, and therefore appear larger than suggested by Table 3.5.

producing a relatively large number of total protein identifications, yielded only 12 protein identifications at ≥30% sequence coverage. In sharp contrast to the other four protocols, the *Guan-HCl* protocol produced only one protein with a sequence coverage ≥30%.

A previous report of nanospray LC-MS-MS analyses of *B. taurus* MAPs ranged from 147 peptide identifications (64% were fully-tryptic peptides) matching 26 protein identifications for a one dimensional reverse-phase chromatographic separation to 431 peptide identifications (57% were fully-tryptic peptides) matching 62 protein identifications for a 3-phase MudPIT separation [59]. The present study is thus comparable to previous results on a comparable analytical platform. For the MAP isolates, the *$CH_3CN$ 16 hr* protocol was most effective in terms of both total and tryptic peptide identifications, and yielded numbers of protein identifications and sequence coverage values greater than 30% that were among the highest. The *Guan-HCl* protocol produced the fewest peptide identifications, fewest total protein identifications, and lowest number of proteins with ≥30% sequence coverage values for MAPs. As discussed above for ribosomal proteins, this protocol has the most sample processing steps, namely dilution of the guanidine hydrochloride and SPE, which offer the potential for the peptide loss. Digestion protocols conducted for as little as 1 hour produced a similar number of peptide identifications as the *Lys-C/trypsin* protocol, suggesting that the combination of two proteases and digestion times beyond 16 hours was not advantageous for peptide identifications in this sample. It is likely that in addition to the dilution step, more peptide losses resulted from the SPE in the *Guan-HCl* protocol, where only 1 protein was identified with greater than 30% sequence coverage.

## *3.4 Conclusions*

Protein enrichment protocols often provide yields in the low microgram to high nanogram range. The characterization of such protein isolates by LC-MS-MS of peptides requires the digestion of proteins into peptides prior to the measurements of the intact mass and fragmentation pattern of the peptides. This digestion step is frequently taken for granted, especially for the digestion of limited or small quantities of protein, e. g. those isolated by affinity purification. Here we compared the performance of existing protocols for the digestion of small quantities of "real world" protein isolates and further demonstrated the effectiveness of the 80% acetonitrile solvent for trypsin digestions.

In both types of protein isolates examined, protocols where the trypsin digestion was conducted in the 80% acetonitrile solvent performed as well as, and in many instances better than, the other protocols tested. For isolates of the *R. palustris* 70***S*** ribosomal protein complex, a 1 hour digestion in 80% acetonitrile resulted in the highest number of peptide identifications. A 16 hour digestion in 80% acetonitrile yielded in the highest number of peptide identifications for the *B. taurus* MAPs. Digestion in 80% acetonitrile solvent can be advantageous because it eliminates steps in the workflow that can lead to losses of protein or peptides, which becomes a more significant problem for samples that contain small quantities of protein. Chaotrope dilution steps, such as those in the *Guan-HCl* and *Lys-C/trypsin* protocols, increase the surface area to which the sample is exposed inside microcentrifuge tubes, increasing the possibility of adsorptive losses of proteins or peptides. To prevent interference in chromatography, chaotropes are frequently removed from peptide mixtures in a sample cleanup step prior to LC-MS-MS by SPE or precipitation, as in the *Guan-HCl* and *RapiGest* protocols. These steps

increase the amount of sample handling, surface area to which the sample is exposed, and exacerbate the potential for peptide loss. The 80% acetonitrile solvent is volatile and may quickly be removed from peptide mixtures by centrifugal evaporation, apparently leading to fewer losses than SPE or precipitation of acid-labile surfactants. Regardless of the type of protein isolates examined, we consistently found the minimization of sample handling steps in the digestion protocols tested to be beneficial. Finally, although not investigated in this study, it is possible that the activity of trypsin differs among the solvents used in the various protocols, which span a considerable range of composition and pH.

# Chapter 4
# Evaluation of Affinity-Tagged Protein Expression Strategies using Local and Global Isotope Ratio Measurements

All of the data presented in this chapter have been adapted from the following published journal article:

W. Judson Hervey, IV, Gurusahai Khalsa-Moyers, Patricia K. Lankford, Elizabeth T. Owens, Catherine K. McKeown, Tse-Yuan Lu, Linda J. Foote, Keiji G. Asano, Jennifer L. Morrell-Falvey, W. Hayes McDonald, Dale A. Pelletier, and Gregory B. Hurst. *Journal of Proteome Research.* 2009, 8 (7), 3675-3688.

Chromosome-integrated *E. coli* sequential peptide affinity (SPA) clones were a generous gift from the laboratory of Andrew Emili, Donnelly Centre for Cellular and Biomolecular Research (CCBR), University of Toronto, Toronto, ON.

Development of the SPA2 plasmid was performed by Gurusahai Khalsa-Moyers, Jennifer L. Morrell-Falvey, W. Hayes McDonald, Linda J. Foote, and Dale A. Pelletier. Molecular cloning of DNA, primer design, and Invitrogen Gateway recombination reactions were performed by Tse-Yuan Lu, Elizabeth T. Owens, Patricia K. Lankford, and Catherine K. McKeown.

Microbial cell growth, $^{15}N$ stable isotope metabolic labeling, isolation of affinity-tagged bait proteins, isolation of soluble proteome fractions, sample preparation, LC-MS-MS measurements, and data analysis was performed by W. Judson Hervey, IV.

Supplementary Information referred to in the Results and Discussion section are available freely on the internet at http://pubs.acs.org/doi/suppl/10.1021/pr801088f

*Abstract*

Elucidation of protein-protein interactions can provide new knowledge on protein function. Enrichments of affinity-tagged (or "bait") proteins with interaction partners generally include background, non-specific protein artifacts. Furthermore, *in vivo* bait expression may introduce additional artifacts arising from altered physiology or metabolism. In this study, we compared these effects for chromosome and plasmid encoding strategies for bait proteins in two microbes: *Escherichia coli* and *Rhodopseudomonas palustris.* Differential metabolic labeling of strains expressing bait protein relative to the wild-type strain in each species allowed comparison by liquid chromatography tandem mass spectrometry (LC-MS-MS). At the local level of the protein complex, authentic interacting proteins of RNA polymerase (RNAP) were successfully discerned from artifactual proteins by the ***i***sotopic ***d***ifferentiation of ***i***nteractions as ***r***andom or ***t***argeted (I-DIRT; A. J. Tackett et al. *J. Proteome Res.* **2005**, *4* (5), 1752-1756). To investigate global effects of bait protein production, we compared proteomes from strains harboring a plasmid encoding an affinity-tagged subunit (RpoA) of RNAP with the corresponding wild-type strains. The RpoA abundance ratios of 0.8 for *R. palustris* and 1.7 for *E. coli* in plasmid strains versus wild type indicated only slightly altered expression. While most other proteins also showed no appreciable difference in abundance, several that did show altered levels were involved in amino acid metabolism. Measurements at both local and global levels proved useful for evaluating *in vitro* and *in vivo* artifacts of plasmid-encoding strategies for bait protein expression.

## *4.1 Introduction*

Many biological processes are conducted by macromolecular assemblies of individual polypeptide chains, or protein "complexes." Elucidation of physical interactions among protein complexes is essential for improved understanding of cellular function. Interaction analyses of entire biological systems are now possible through integration of genomic, analytical, and computational tools [3]. Typically, such large-scale studies encompass systematic surveys of genome-wide fusion protein expression libraries with techniques to measure physical protein interactions, such as two-hybrid, phage display, or affinity isolation [2]. The combination of affinity purification of a tagged "bait" protein and its interaction partners with mass spectrometry (AP-MS) [30, 76] has emerged as an attractive, robust, high-throughput approach to performing such measurements. Mass spectrometric identification of the proteins that are co-purified along with the bait enables the inference of physical interactions among enriched proteins, implying membership in a protein complex and providing insights into function. Amid authentic*, bona fide* protein-protein interactions, however, it is common to observe non-specific protein artifacts. In the absence of follow-up experimental verification, identifying the authentic, biologically significant protein interactions from AP-MS datasets relies upon statistical techniques [77], corroborating predictions from bioinformatics, and/or computational algorithmic development [78].

Most AP-MS studies have been conducted in model organisms for which genetic systems have been developed that enable expression of chromosome-encoded bait fusion proteins [47, 79]. Such chromosome encoding strategies have the advantage that bait proteins are expressed under control of the native promoter gene sequence and, in

prokaryotic organisms, are transcribed as a unit with other genes in the same operon. In order to extend the utility of AP-MS to the growing number of microbes for which genome sequencing is complete, but for which tools for chromosomal integration have not yet been developed, our laboratory recently described a plasmid-based strategy for expression of fusion proteins for AP-MS studies across a range of Gram-negative species [80]. In plasmid-based strategies, bait proteins are encoded in extra-chromosomal DNA that has been introduced into the cell. Expression of the bait from plasmid is often under control of a promoter sequence that cannot be regulated. Although effects such as altered levels of bait protein expression or other collateral physiological consequences are likely [81], plasmid-based systems offer advantages such as implementation with general molecular biological tools and commercial reagents, no requirement of a species-specific genetic system, and applicability for organisms beyond "model" biological species.

In order to differentiate authentic from artifactual interactions, isotope ratio mass spectrometry measurements have been applied to the study of protein-protein interactions by AP-MS [30, 31, 76]. In particular, the *i*sotopic *d*ifferentiation of *i*nteractions as *r*andom or *t*argeted (I-DIRT) method has been demonstrated using chromosomal fusion expression strategies in *S. cerevisiae* [82] and *E. coli* [83]. In this method, a strain that has been genetically modified to express a bait protein is grown in medium prepared from components with naturally occurring stable isotope distributions (mostly $^{12}C$, $^{14}N$, etc.), so that newly synthesized proteins that interact with the bait *in vivo* during culture growth will contain predominantly the naturally occurring distribution of isotopes. A second culture of an unmodified strain, grown in medium enriched in a distinctive stable isotope (e.g. $^{13}C$ or $^{15}N$), is mixed with the bait-expressing strain. Proteins from the latter strain

that interact with the bait *in vitro* after mixing are affinity isolated along with the bait protein. These proteins can be distinguished via their isotope abundance ratios in mass spectra; authentic interactors will contain mostly naturally abundant isotopes, while proteins interacting with the bait after mixing will contain more of the distinctive stable isotope.

In this study, we compared different strategies of bait fusion protein expression for their effects on detection of microbial protein interactions by AP-MS. We used the DNA-directed RNA polymerase (RNAP) as a model complex to investigate strategies for encoding bait proteins in two phylogenetically distinct microbial species: *Escherichia coli*, a model species, and *Rhodopseudomonas palustris* [66], a metabolically diverse and environmentally uibiquitous species. For the chromosome-encoded bait protein strategy, we used sequential peptide affinity (SPA) [50, 79] strains expressing the *E. coli* RNA polymerase proteins RpoA/SPA, RpoB/SPA, and RpoC/SPA. For the plasmid-encoded bait protein strategy, we encoded SPA-tagged RpoA in *E. coli* and *R. palustris* using a medium-copy vector backbone [80]. To differentiate authentic protein-protein interactions from co-isolating artifactual interactions, we looked locally, i.e. at the level of the protein complex, by applying the I-DIRT method to affinity-isolated RNAP using different bait proteins. To examine more global consequences of the plasmid encoding strategy on the cell, we performed whole-proteome quantitative MS measurements of the differentially labeled plasmid-expressing strains relative to their wild-type counterparts. For both *E. coli* and *R. palustris*, expression levels of plasmid encoded RpoA were similar to wild type, and most other proteins were within 2-fold of their abundance in wild type strains. Nevertheless, some 2-fold or greater differences in abundances of

proteins, especially those involved in amino acid biosynthesis, were observed. At both local and global levels, chromosome- and plasmid-encoding strategies for the genes studied yielded comparable results for detection of authentic protein-protein interactions.

## *4.2 Experimental Procedures*

Reagents obtained from Sigma Chemical Company (St. Louis, MO) included Trizma Hydrochloride (UltraPure), Trizma Base (UltraPure), $NH_4HCO_3$ (UltraPure), guanidine hydrochloride (UltraPure), 2-mercaptoethanol (≥98% molecular biology grade), $KH_2PO_4$ (≥ 99.0% purity), ethylene glycol-bis(aminoethylether)-tetraacetic acid (EGTA) (SigmaUltra), glycerol (SigmaUltra), Triton® X-100 (SigmaUltra), ANTI-FLAG® M2 affinity gel, $CH_3CO_2NH_4$ (99.999% purity), and disodium ethylenediamine-tetraacetic acid (EDTA) (99+% molecular biology grade). Other reagents included dithiothreitol (DTT) (OmniPur), glucose (USP purity), Novagen Benzonase® nuclease, and formic acid (SupraPur) (EMD Chemicals, Darmstadt, Germany); $^{15}NH_4Cl$ (98+ atom % excess $^{15}N$) and $(^{15}NH_4)_2SO_4$ (98+ atom % excess $^{15}N$) (Isotec, Miamisburg, OH); $NH_4Cl$ (99.5% purity), $(NH_4)_2SO_4$ (99.5% purity), $CaCl_2$ (99% purity) and NaCl (molecular biology grade) (Mallinckrodt Baker Chemicals, Phillipsburg, NJ); AcTEV™ protease (Invitrogen, Carlsbad, CA); Calmodulin Sepharose™ 4B (GE Healthcare, Piscataway, NJ); bicinchoninic acid (BCA) Protein Assay Kit (ThermoFisher Scientific, Rockford, IL), lyophilized sequencing-grade modified trypsin (Promega, Madison, WI), HPLC-grade acetonitrile and water (Burdick and Jackson, Muskegon, MI); $MgSO_4 \cdot 7H_2O$ (99.9% purity) and $Na_2HPO_4$ (USP) (Fischer Scientific, Fair Lawn, NJ). Ultrapure 18 MΩ water used for protein isolation buffers was obtained from a Milli-Q system (Millipore, Bedford, MA). HPLC-grade water was used in trypsin buffer (50 mM

Trizma-HCl, 10 mM $CaCl_2$, pH 7.6). HPLC-grade water, HPLC-grade acetonitrile, formic acid (SupraPur), and $CH_3CO_2NH_4$ (99.999% purity) were used in preparation of HPLC mobile phases.

Table 4.1 lists microbial strains, metabolic labeling, and growth media used in three bait encoding strategies (outlined in Figure 4.1). The SPA [50, 79] strains used in the chromosome encoding strategy were a generous gift from the laboratory of Andrew Emili, Donnelly Centre for Cellular and Biomolecular Research (CCBR), University of Toronto, Toronto, ON. For the plasmid encoding strategy, the SPA2 plasmid was designed to encode a bait protein with a C-terminal calmodulin binding peptide epitope, TEV cleavage sites, and FLAG epitopes, encoding an affinity tag comparable to chromosomally-integrated SPA bait proteins [84]. DNA sub-cloning, recombination into a broad-host-range vector backbone that allows encoding of fusion proteins bearing a variety of affinity tags, and strain sub-culturing was performed as described elsewhere in more detail [80].

*Escherichia coli* and *Rhodopseudomonas palustris* strains were grown to exponential phase in $^{14}N$- or $^{15}N$-enriched minimal medium. *E. coli* cultures were grown on M9 minimal medium [85] to a density of $OD_{600}$=0.6, with $NH_4Cl$ as the sole source of nitrogen [40]. Growth was conducted in baffled Erlenmeyer flasks under aerobic conditions in a shaking incubator (New Brunswick Innova) at 37°C, with agitation at approximately 1,440 rpm. *R. palustris* cultures were grown on photosynthetic medium (PM-10) under anaerobic, photoheterotrophic conditions with succinate as a carbon source, to a density of $OD_{660}$=0.6 [41, 86]. In PM-10 medium, ammonium sulfate (0.014 M) is the sole source of nitrogen available for incorporation into biosynthesized proteins;

**Table 4.1.** Microbial Species, Affinity-Tagged Bait Protein Expression Strategy, and Metabolic Labeling used in this study.

| ***Species*, strain, Bait Protein, or (Gene Number)** | **RNAP Subunit** | **Molecular Weight[a]** | **Encoding Strategy[b]** | **Growth Medium; Nitrogen Source** | **Source** | **Reference** |
|---|---|---|---|---|---|---|
| *Escherichia coli* | | | | | | |
| RpoA/SPA (ECK3282) | α | 36.5 kDa | chromosome | M9; $NH_4Cl$ | A. Emili | 8, 15 |
| RpoB/SPA (ECK3978) | β | 150.6 kDa | chromosome | M9; $NH_4Cl$ | A. Emili | 8, 15 |
| RpoC/SPA (ECK3979) | β' | 155.2 kDa | chromosome | M9; $NH_4Cl$ | A. Emili | 8, 15 |
| RpoA/SPA2 (ECK3282) | α | 36.5 kDa | plasmid | M9; $NH_4Cl$ | our lab | 9, 16 |
| RpoB/SPA2 (ECK3978) | β | 150.6 kDa | plasmid (-) | M9; $NH_4Cl$ | our lab | 9, 16 |
| RpoC/SPA2 (ECK3979) | β' | 155.2 kDa | plasmid (-) | M9; $NH_4Cl$ | our lab | 9, 16 |
| K12 MG1655 (wild-type strain) | | | - | M9: $^{15}NH_4Cl$ | ATCC (700926) | 25 |
| *Rhodopseudomonas palustris* | | | | | | |
| RpoA/SPA2 (RPA3226) | α | 37.5 kDa | plasmid | PM-10; $(NH_4)_2SO_4$ | our lab | 9, 16 |
| RpoC/SPA2 (RPA3267) | β' | 156.1 kDa | plasmid (-) | PM-10; $(NH_4)_2SO_4$ | our lab | 9, 16 |
| RpoB/SPA2 (RPA3268) | β | 154.6 kDa | plasmid (-) | PM-10; $(NH_4)_2SO_4$ | our lab | 9, 16 |
| CGA010 (wild-type strain) | | | - | PM-10; $(^{15}NH_4)_2SO_4$ | C. Harwood | 14, 18 |

[a] Molecular weight of the native, untagged protein determined by the amino acid sequence of the predicted proteome. Molecular weight of the SPA affinity tag is approximately 8.1 kDa[8]. [b] Entries labeled "plasmid (-)" indicate fusion proteins that did not express.

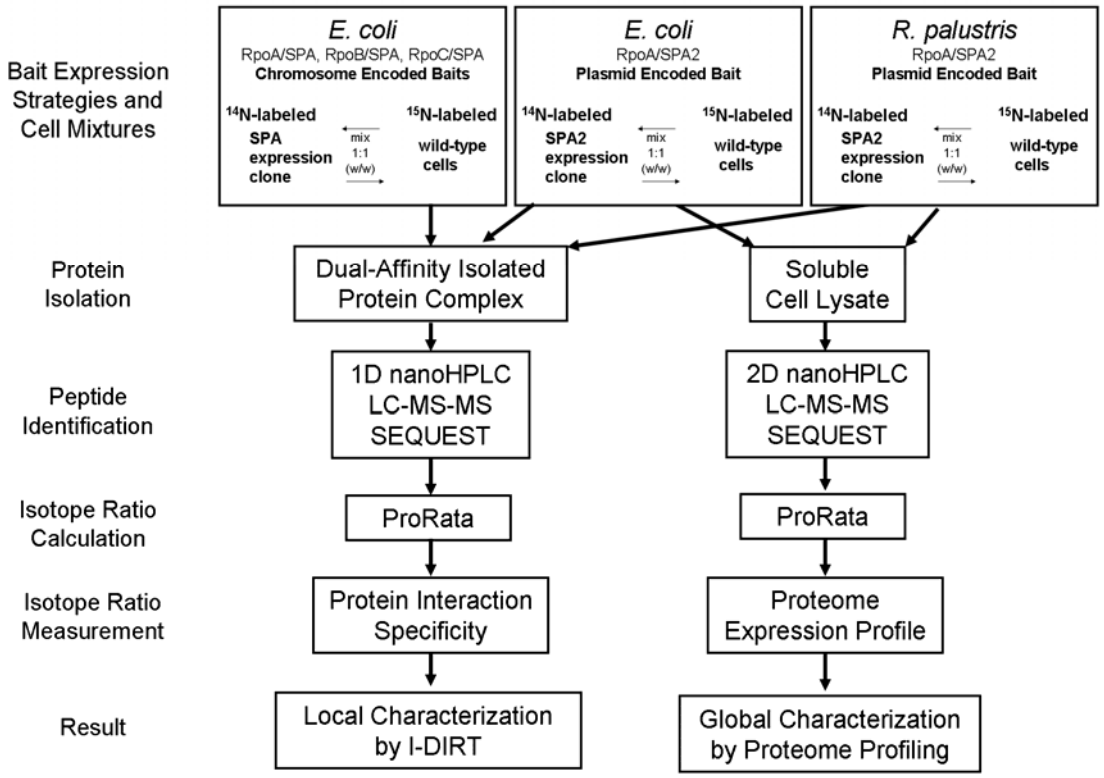



**Figure 4.1.** Experimental outline for evaluation of encoded affinity-tagged (or "bait") protein strategies presented in this study.

A small concentration of *p*-aminobenzoic acid (0.015 mM) provides a negligible amount of $^{14}N$ contamination. Duplicate cell cultures were grown as outlined in Figure 4.2. Each cell culture was harvested by centrifugation [80]. Mixtures of bait-expressing strains and isotopologous wild-type strains were created at 1:1 (w/w) wet cell paste mass ratios on ice [82]. Wet cell paste mass of each mixture was measured on an analytical balance (Mettler, Toledo, OH); total cell paste masses were approximately 1.5 g.

**Protein Complex Analysis.** Following the left-hand workflow shown in Figure 1, isolation of the RNA polymerase (RNAP) protein complex was performed using protocols appropriate for the SPA tag [50, 79], with slight modifications for the I-DIRT method [82, 83]. Each cell mixture was resuspended at 0.2 g/mL in M2 buffer (10 mM Trizma-HCl, 100 mM NaCl, 10% glycerol, 0.1% Triton® X-100, 1 μg/mL Benzonase® Nuclease), and incubated on ice for 30 min. Cell lysis was performed on ice using a Branson Sonifier (Danbury, CT) [41, 86] and cellular debris were removed by centrifugation [80]. RNAP was isolated from the resulting cleared, soluble protein fraction of cell lysate for all bait proteins [50, 79], with the following modifications: cell lysate was incubated with ANTI-FLAG® M2 affinity gel at 4°C for 4 h; AcTEV™ protease reaction was conducted at ambient laboratory temperature (approximately 25°C) for 1 h; and, AcTEV™ cleavage product was incubated with Calmodulin Sepharose™ 4B at 4°C for 1 h. The final elution from Calmodulin Sepharose™, containing the affinity-isolated protein enrichment, was digested with trypsin in a solvent of 80% $CH_3CN$ [87].

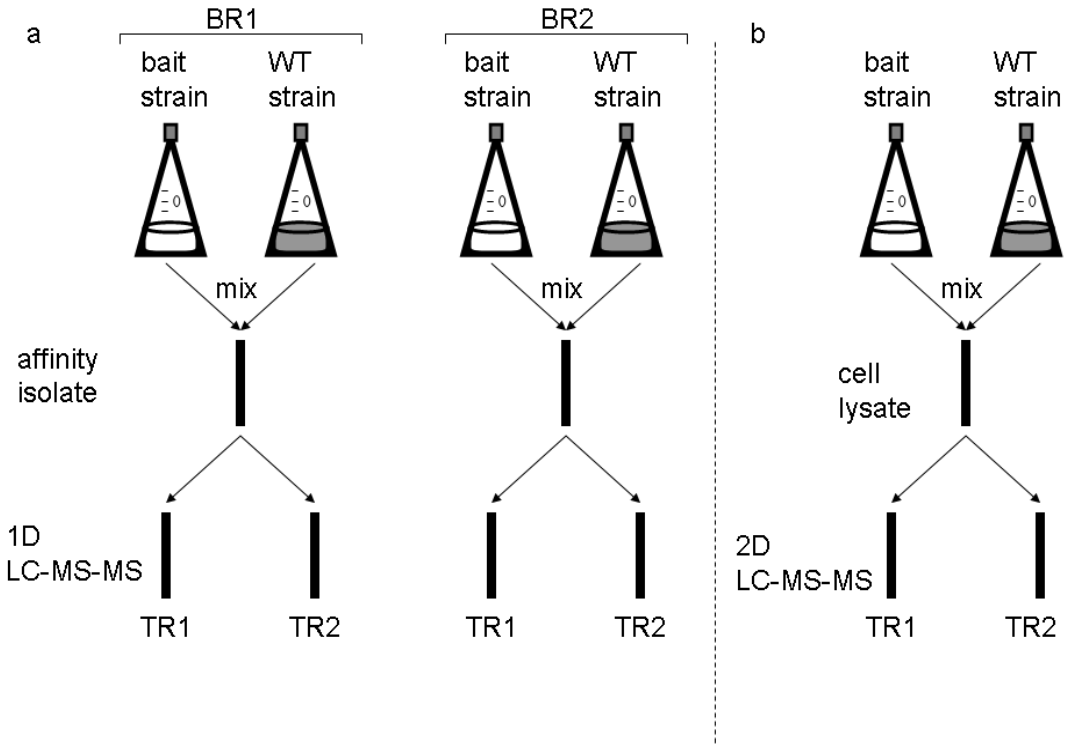



**Figure 4.2.** Replication scheme for experiments outlined in Figure 1. (a) Local (complex-specific) measurements; (b) global (whole proteome) measurements. BR= biological replicate (i.e., culture), TR = technical replicate (i.e. LC-MS-MS run).

The resulting mixtures of proteolytic peptides were separated by an automated one dimensional reverse phase $C_{18}$ nanoHPLC system, composed of a FAMOS autosampler, Switchos, and UltiMate HPLC pump (LC-Packings/Dionex, Sunnyvale, CA), described in more detail previously [87]. The nanoHPLC system was coupled directly to the nanoelectrospray source (Proxeon Biosystems, Odense, Denmark) of a linear ion trap mass spectrometer (ThermoFinnigan LTQ, San Jose, CA), under control of the Xcalibur software (version 2.0.9, ThermoFinnigan). The LTQ instrument scanned the mass-to-charge range of 400-1700 in data dependent mode and subjected the three most intense ions to MS/MS analysis at 35% normalized collisional energy. Dynamic exclusion [64] was enabled and conducted under the following settings: repeat count 1, repeat duration 60 sec., exclusion size list 300, exclusion duration 180 sec., exclusion width by mass (low and high) 1.5 Da, with expiration disabled. The default charge state was set to 3. Full-scan mass spectra were averaged from four microscans; tandem mass spectra were averaged from two microscans.

**Proteome Analysis.** Soluble proteome fractions were isolated from isotopologous cell mixtures of plasmid-bearing strains and wild-type strains, shown in the right-hand workflow in Figure 4.1. For *E. coli,* a single 1:1 (w/w) cell mixture was created between a strain encoding the RpoA protein in the SPA2 plasmid and the K12 MG1655 wild-type strain. Similarly, for *R. palustris,* a single mixture was created between the RpoA/SPA2 plasmid-bearing strain and the CGA010 wild-type. Proteome samples were processed as previously described in more detail [41, 86]. Briefly, the protein content of cleared, soluble proteome fractions from each cell mixture (obtained as described above) was measured by the BCA Assay (ThermoFisher). Approximately 2 mg of each proteome

fraction was digested with trypsin and resulting peptides desalted by solid phase extraction [41, 86]. The resulting mixtures of proteolytic peptides were separated by a two dimensional (strong cation exchange and reverse phase) nanoHPLC system [41, 86], configured for a split-phase 12-step MudPIT separation [26], coupled to a nanoelectrospray source (Proxeon Biosystems, Odense, Denmark) of a linear ion trap mass spectrometer (ThermoFinnigan LTQ, San Jose, CA). The LTQ mass spectrometer was operated in data dependent mode with dynamic exclusion [64] enabled, as described above. Duplicate LC-MS-MS analyses were conducted for the *E. coli* mixture and for the *R. palustris* mixture as outlined in Figure 4.2.

**Peptide Identification, Protein Quantification, and Data Analysis**. Identification of peptides from tandem mass spectra was performed using the SEQUEST algorithm (version 27) [10] without enzyme specificity. *E. coli* FASTA-formatted protein sequences were obtained from SwissProt [88], while *R. palustris* FASTA-formatted protein sequences (version 3.0) were obtained from the ORNL genome annotation pipeline [89]. To estimate the false positive peptide identification rate, the predicted proteome of each microbial species was reversed and concatenated to the FASTA-formatted protein sequence file [90] by Perl scripting. Additional proteins encoded by the vector backbone [80] (antibiotic resistance proteins, and affinity tag sequences), and common contaminant proteins (in-house standard mixture proteins, proteolytic enzymes, and keratins, etc. specified as "contaminant" in the FASTA protein description) were added to each FASTA sequence file, resulting in 8,712 protein sequences in the *E. coli* file and 9,700 protein sequences in the *R. palustris* file. Two separate SEQUEST searches were performed, with amino acid molecular masses calculated based on $^{14}N$ for

the first, and $^{15}N$ for the second search [41]. Results from these $^{14}N$ and $^{15}N$ SEQUEST searches were merged using the "Merge Directories" tool of the ProRata program GUI (version 1.1) [41].

Peptide identifications were filtered, organized, and assembled by locus using the DTASelect program (version 2.06) [27]. DTASelect retained identifications based on singly-, doubly-, and triply–charged peptide ions with $X_{Corr}$ scores $\geq$1.8, $\geq$2.5, and $\geq$3.5, respectively, and $\Delta$CN scores $\geq$0.08. Identification at the protein level required identification of two or more non-redundant tryptic peptides. Multiple MS-MS scans for identified peptide sequences were retained by specifying the DTASelect locus filter *–t 0*. DTASelect discarded redundant locus information by default; contaminant proteins (proteolytic enzymes, keratins, etc.) were removed using the filter *–l contaminant*. The false positive identification rates were estimated by DTASelect as $\leq$0.3% at the peptide level and $\leq$0.7% at the protein level.

The ProRata program extracted selected ion chromatograms, estimated peptide abundance ratios, and assembled peptide quantifications into protein quantifications according to default specifications [41], with the additional requirement of $\geq$3 peptide quantifications per protein. Quantification data from biological replicates of protein complex analyses was further filtered by the ProRata program [86] to include only those proteins quantified in all 4 total LC-MS-MS analyses per bait protein, yielding a single isotopic protein abundance ratio and 90% confidence interval. Interaction specificity with RNAP was based upon the combination of protein abundance ratio, known interactors determined by standard biochemical techniques [91, 92], and those reported in previous AP-MS studies [50, 80, 83]. For proteome analyses, the ProRata program

estimated peptide and protein abundance ratios as described above; however, isotope ratio results were reported only for those proteins for which isotope ratios could be determined in both replicate LC-MS-MS analyses. Proteins were considered more abundant in the plasmid encoded bait expression strains if the $\log_2$ abundance ratio of the protein was $\geq$1, and less abundant if the $\log_2$ abundance ratio was $\leq$-1 [86]. *E. coli* proteins differing in abundance were classified into clusters of orthologous groups of proteins (COGs) functional category assignments using the COGnitor tool [93]. COG functional categories for *R. palustris* proteins differing in abundance were retrieved from the ORNL genome annotation pipeline [89]. KEGG pathways for proteins were downloaded from the Kyoto Encyclopedia of Genes and Genomes [94]. Predicted operons for *E. coli* and *R. palustris* genes were obtained from the Virtual Institute for Microbial Stress and Survival Operon Prediction website [95]. Tab-delimited output of protein identifications, abundance ratios, and COGs were further analyzed in a relational database (Microsoft Access).

## *4.3 Results and Discussion*

In this study, we have applied isotope ratio proteomics techniques to compare affinity-tagged bait fusion protein expression from chromosome and plasmid encoding strategies for detection of microbial protein interactions. We devised the current experimental design, shown in Figure 4.1, using RNAP as a model protein complex. This study used isotope ratio mass spectrometry in both a local approach to characterize the proteins that were isolated along with the bait in the affinity purification, and in a global approach to examine effects of expression of a plasmid-encoded bait on protein abundances across the proteome. The local approach, I-DIRT, used $^{14}N$:$^{15}N$ isotopic abundance ratios to

distinguish authentic protein-protein interactions with the bait protein from artifactual interactions. The global approach, shotgun proteome profiling, used metabolic stable isotope labeling to enable abundance ratio measurements for proteins in plasmid-bearing strains relative to wild-type strains. In combination, these two measurements allowed us to assess particular protein-protein interactions, the extent of bait protein over-expression, and broader effects on the proteome profile arising from the "metabolic load" [81] imposed by the presence of the plasmid.

Numerous classical biochemical experiments encompassing decades of research have deciphered the role of DNA-dependent RNA polymerase in the flow of genetic information [91, 92]. In bacterial species, the transcription of genomic DNA into mRNA is performed by a 5-subunit core polymerase, with specificity conferred by the presence of one of several gene-specific σ factors. Core RNAP consists of the α-subunit, RpoA (36.5 kDa), a β-subunit, RpoB (150.6 kDa), a β'-subunit, RpoC (155.2 kDa), and an ω – subunit, RpoZ (10.2 kDa), configured in a 2:1:1:1 stoichiometry, respectively ($\alpha_2\beta\beta'\omega$) [91, 92]. This breadth of previous biochemical knowledge regarding this protein complex has contributed to its use as a model system for demonstrating both chromosome-[50, 79] and plasmid-encoded [80] fusion bait protein strategies, and as a model to demonstrate the I-DIRT method in the hemorrhagic *E. coli* strain O157:H7 [83]. To leverage this wealth of background information, we chose specific components of RNAP as baits, and applied the I-DIRT method to *E. coli* strains expressing chromosome-encoded SPA-tagged [50, 79] bait proteins (designated RpoA/SPA, RpoB/SPA, and RpoC/SPA), and to *E. coli* and *R. palustris* strains expressing plasmid-encoded SPA2-tagged RpoA (shown in Table 4.1).

**Chromosome-Encoded Baits.** Affinity isolation by the I-DIRT method [82, 83] was performed from a 1:1 (w/w) cell mixture of an isotopically "light" strain expressing the affinity-tagged bait protein with an isotopically "heavy" wild-type strain (Figure 1). Authentic components of RNAP that interact *in vivo* with the bait protein will be isotopically light, while artifactual associations occurring *in vitro* after mixing of the cell pellets will be a mixture of light and heavy isotopologs. Thus proteins involved in authentic interactions yield greater $^{14}N$:$^{15}N$ isotopic ratios than artifactual proteins.

I-DIRT Analysis of *E. coli* RpoA/SPA Co-Isolates**.** Isotope ratios for a total of 15 proteins that co-isolated with the RpoA/SPA bait protein are shown in Figure 4.3a and Table S1 (Supporting Information). Relatively large $^{14}N$:$^{15}N$ isotopic abundance ratios of 14.9, 6.5, and 4.3 were measured for core RNAP components RpoB, RpoA, and RpoC, respectively. Among other proteins with $^{14}N$:$^{15}N$ isotopic abundance ratios indicative of specific interactions were DnaK (4.3) and RapA, formerly HepA[15] (2.6). The latter two proteins were previously identified qualitatively with affinity isolated RNA polymerase [50], and in I-DIRT analysis of the *E. coli* O157:H7 RNAP as authentic-interacting proteins [83]. In addition to DnaK, another heat-shock family protein, CH60 (GroEL), was quantified with an isotopic ratio of 2.8. This interaction is plausibly specific, as the presence of chaperones in this case could be attributable to the C-terminal affinity tag of the bait protein or other anomalies related to fusion protein folding, although literature sources, consulted as described in Experimental Procedures, offered no evidence suggest-

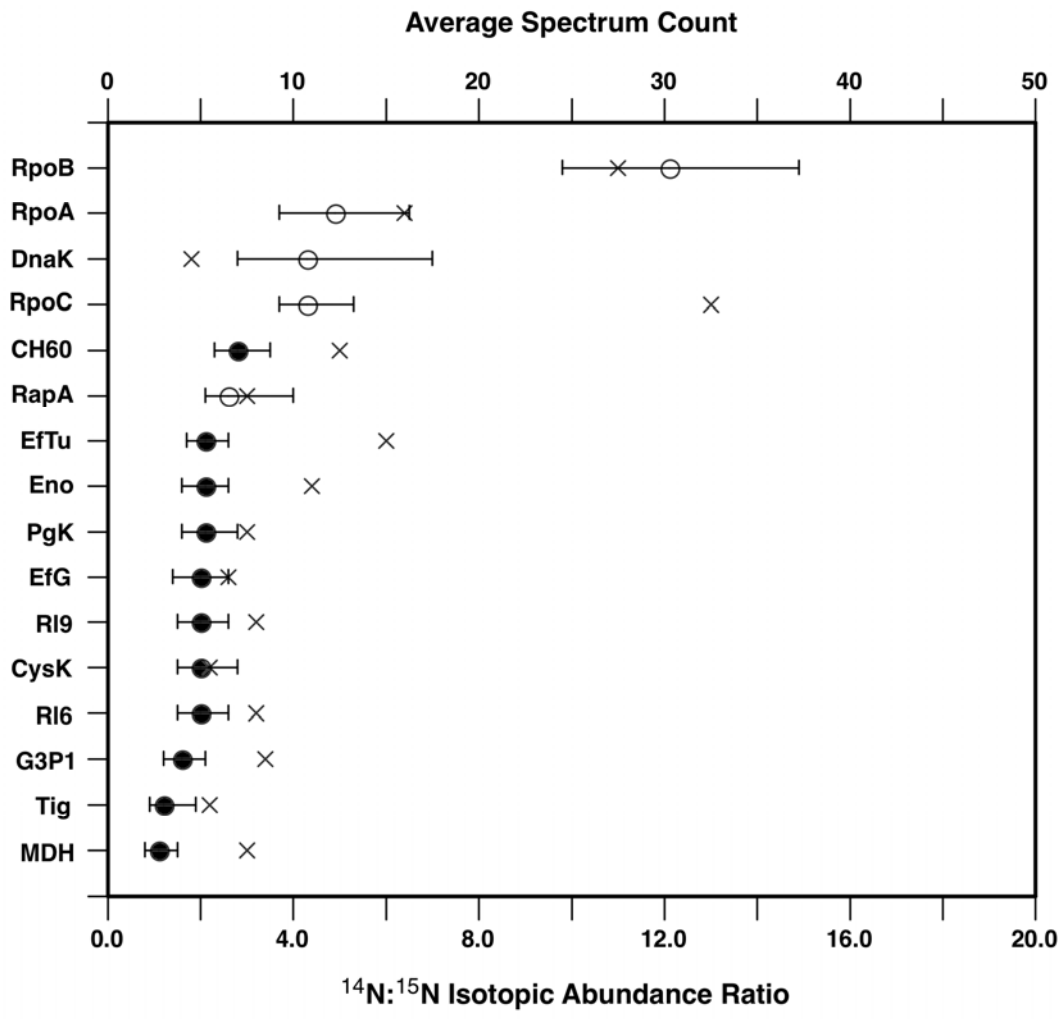



**Figure 4.3a.** I-DIRT measurements of affinity isolated RNA polymerase protein complex from the *E. coli* chromosome-encoded RpoA/SPA bait protein. Isotopic abundance ratios for each protein are indicated with a circle; open circles indicate specific protein interactions, while filled circles represent non-specific protein interactions. Scale for the abundance ratio is shown along the lower x-axis. Error bars represent the 90% confidence interval calculated by ProRata. Average spectrum count is shown for each protein and indicated with an X. The scale is provided along the upper x-axis.

(Rl9), ribosomal protein L6 (Rl6), and ribosomal-associated trigger factor (Tig), may drive the equilibrium toward association of these proteins with the bait protein or other components of the RNAP complex, leading to detection of these abundant proteins in the AP-MS experiment. As a crude estimate of absolute protein amount, Figure 3a also shows the average spectrum count, a semi-quantitative measure of protein abundance [96], associated with each detected protein. The spectrum count observed for known interactor DnaK was measurably lower than for abundant cellular proteins such as EfTu, Rl6, and Rl9. However, low $^{14}N$:$^{15}N$ ratios for these abundant proteins clearly indicated that they associated with the bait after mixing of the "light" and "heavy" cell pellets and cellular lysis, and thus are likely to be artifactual. On the other hand, spectrum count values for RpoA, RpoB and RpoC were fairly high, indicating that the affinity isolation is effective in enriching RNAP components.

I-DIRT Analysis of *E. coli* RpoB/SPA Co-Isolates**.** Isotope ratios for a total of 11 proteins were measured from isolations of the RpoB/SPA bait protein (shown in Figure 4.3b, and Table S2 in Supporting Information). Similar to results for RpoA/SPA described above, co-isolates of the RpoB/SPA bait protein yielded relatively large isotope ratios for core polymerase components RpoA (9.8), RpoC (4.6), and the bait protein itself, RpoB (13.9). The RapA protein was also classified as a specific interactor with the complex, with an isotope ratio of 4.0.

The distribution of the isotopic abundance ratios for the other 7 proteins co-isolating with the RpoB/SPA bait protein was lower and nearly uniform, with a mean of 1.9. As in the case of the RpoA/SPA bait, some of the identified proteins with lower isotopic ratios

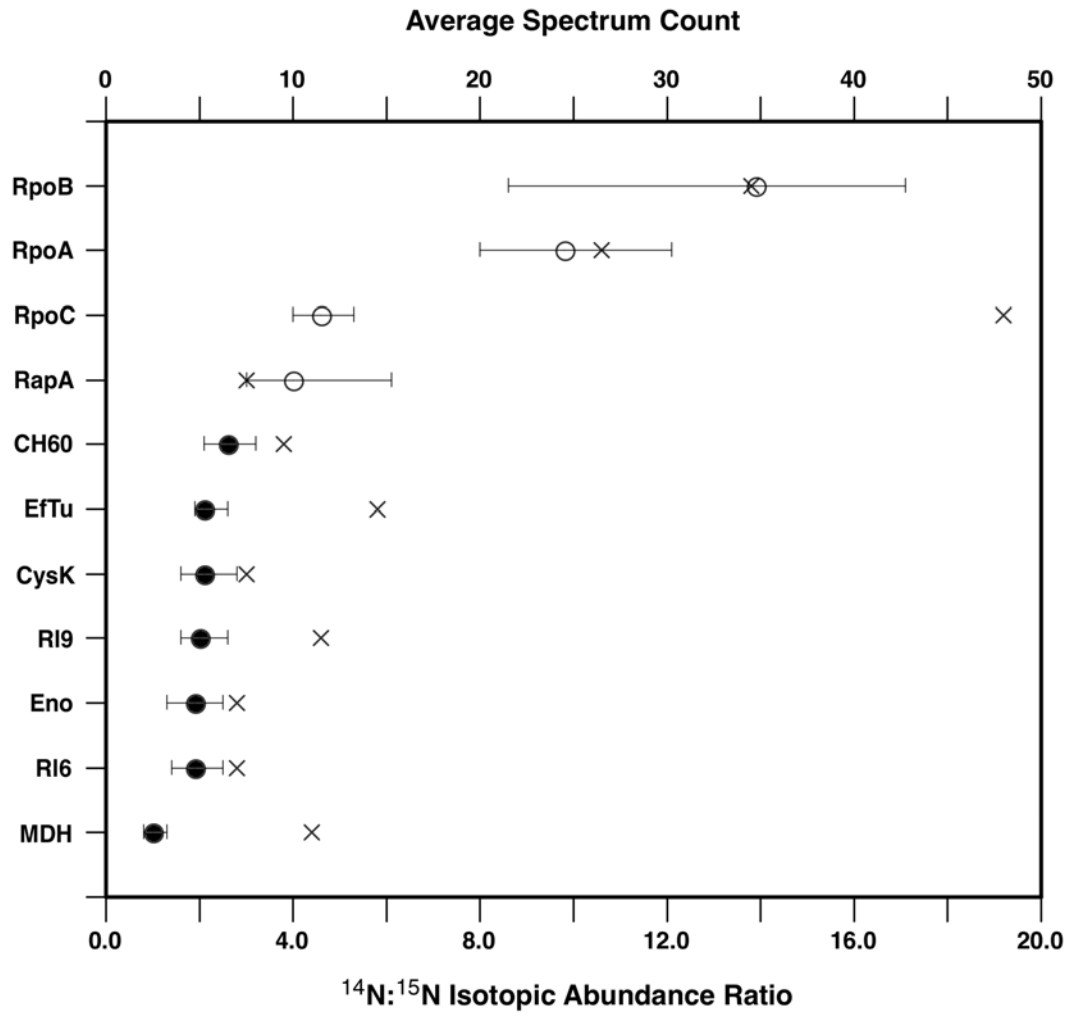



**Figure 4.3b.** I-DIRT measurements of affinity isolated RNA polymerase protein complex from the *E. coli* chromosome-encoded RpoB/SPA bait protein. Isotopic abundance ratios for each protein are indicated with a circle; open circles indicate specific protein interactions, while filled circles represent non-specific protein interactions. Scale for the abundance ratio is shown along the lower x-axis. Error bars represent the 90% confidence interval calculated by ProRata. Average spectrum count is shown for each protein and indicated with an X. The scale is provided along the upper x-axis.

Are known to be abundant cellular proteins, such as ribosomal proteins and EfTu.

I-DIRT Analysis of *E. coli* RpoC/SPA Co-Isolates**.** Isotope ratios for the 8 proteins co-isolated with the RpoC/SPA bait protein are shown in Figure 4.3c and Table S3 (Supporting Information). Isolation of the RNAP complex from the RpoC/SPA protein was the most stringent isolation among chromosome-encoded bait proteins, yielding isotope ratios for only 2 proteins likely to be background (EfTu, 2.1; and Rs16, 1.4). Just as with co-isolates of the two other chromosome-encoded fusion proteins, specific interactions with the bait were characterized by large $^{14}N$:$^{15}N$ isotopic abundance ratios for RpoA (4.0), RpoB (7.5), and RapA (4.9). Two additional specific RNAP interactions were also detected with this bait: the $\sigma_{70}$ factor, RpoD, and the ω subunit, RpoZ, yielded isotope ratios of 4.3 and 8.6, respectively. Detection of these two proteins as authentic-affiliated components of RNAP is consistent with a previous report using Protein A-tagged RpoC as bait in the I-DIRT method [83]. Interestingly, we observed the interaction between the known core subunit RpoZ and RNAP exclusively from isolations of the RpoC/SPA bait. Our result is also consistent with a biophysical study suggesting that RpoZ exhibits a chaperone-like propensity towards RpoC [97].

**Plasmid-Encoded Baits.** In recent years, the increasing rate of microbial genome sequencing has outpaced the development of homologous recombination systems for expression of chromosome-encoded bait proteins in these species. For these sequenced microbes, recombinant DNA technology and commercial systems for plasmid-based expression of fusion proteins is a useful alternative to homologous recombination. To determine whether plasmid-based strategies provide protein-protein interaction results

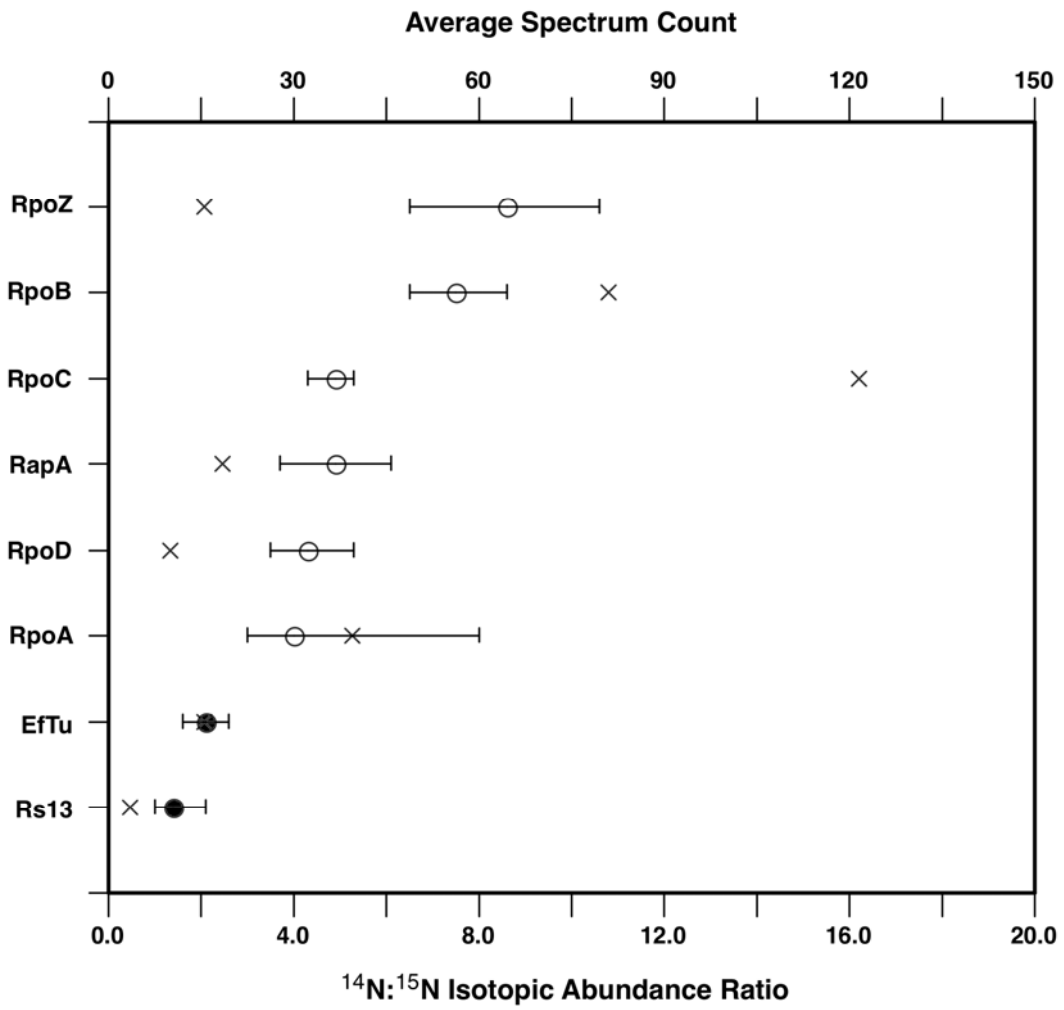



**Figure 4.3c.** I-DIRT measurements of affinity isolated RNA polymerase protein complex from the *E. coli* chromosome-encoded RpoC/SPA bait protein. Isotopic abundance ratios for each protein are indicated with a circle; open circles indicate specific protein interactions, while filled circles represent non-specific protein interactions. Scale for the abundance ratio is shown along the lower x-axis. Error bars represent the 90% confidence interval calculated by ProRata. Average spectrum count is shown for each protein and indicated with an X. The scale is provided along the upper x-axis.

Comparable to chromosomal integration, it is important to compare the two approaches. To compare the stringency of isolation between chromosome-encoded baits (described above) and plasmid-encoded baits, we implemented the left-hand workflow in Figure 1, and encoded the same genes as described above for chromosomal studies (*rpoA*, *rpoB*, *rpoC*) in a plasmid and introduced the plasmid into *E. coli* cells. To demonstrate the plasmid strategy in a less thoroughly studied species, plasmids encoding the analogous proteins were also transformed into *R. palustris*. Attempts in both species to express SPA2 fusion proteins with RpoB and RpoC from plasmid were unsuccessful, even though RpoB and RpoC could be expressed as fusions with a dual hexahistidine and V5 epitope affinity tag when cloned into the pBBR5DEST42 destination vector, derived from the same backbone [80]. Expression of fusion protein from the *rpoA* gene was successful in both species.

I-DIRT Analysis of *E. coli* RpoA/SPA2 Co-Isolates**.** $^{14}N$:$^{15}N$ isotope ratios for eight proteins were obtained in isolates of RNAP from the plasmid-encoded *E. coli* RpoA/SPA2 bait protein and are shown in Figure 4.3d and Table S4 (Supporting Information). As with chromosome-encoded fusion proteins, the isotopic abundance ratios of core components RpoB (16.0), RpoC (13.9), and RpoA (9.8) were greater than other proteins, whose mean isotope abundance ratio was 0.9. The majority of the artifactual proteins shown in Figure 3d represent abundant ribosome-associated proteins, such as EfTu and Rs7, as well as outer membrane protein A (OmpA), and pyruvate dehydrogenase E1 component (Odp1, formerly AceE).

As discussed above, we observed that average spectrum count values did not correlate

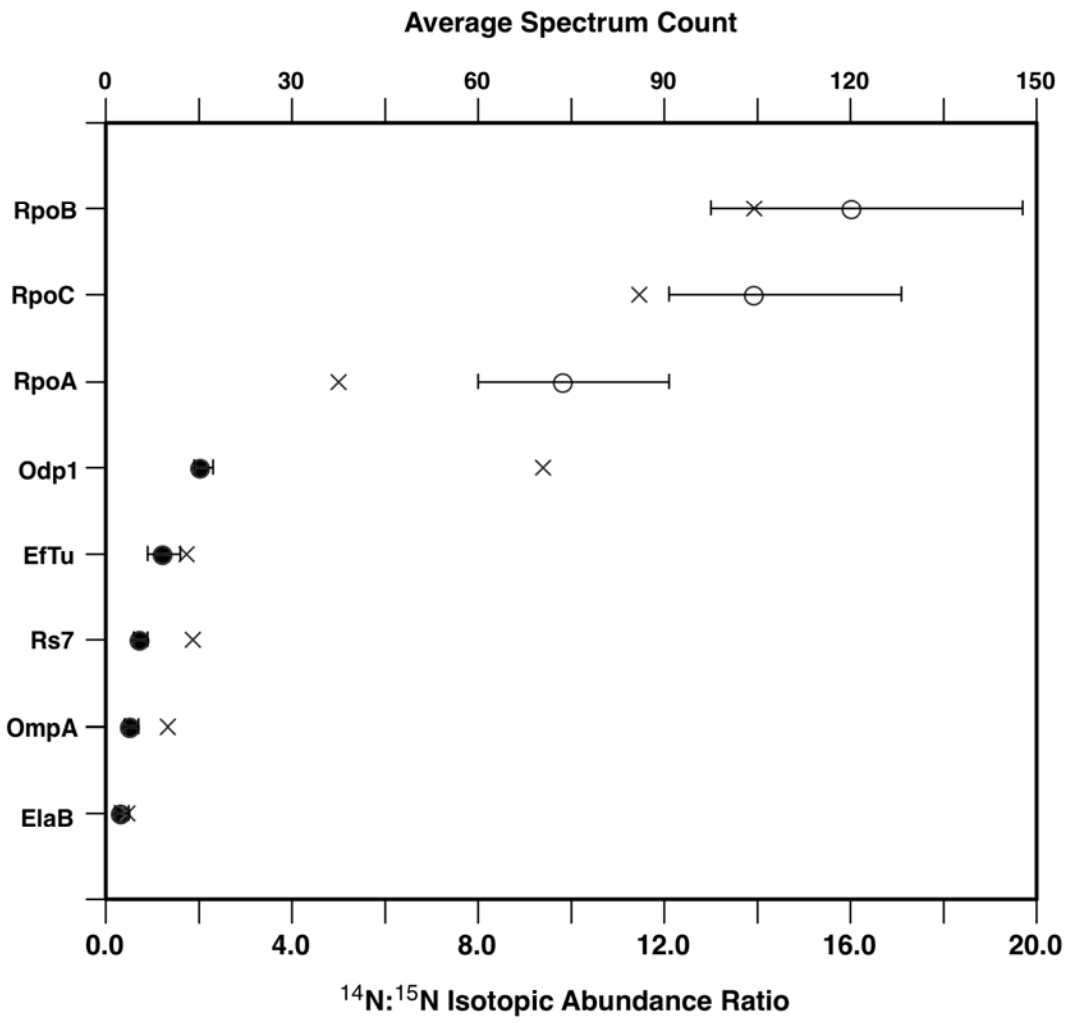



**Figure 4.3d.** I-DIRT measurements of affinity isolated RNA polymerase protein complex from the *E. coli* plasmid-encoded RpoA/SPA2 bait protein. Isotopic abundance ratios for each protein are indicated with a circle; open circles indicate specific protein interactions, while filled circles represent non-specific protein interactions. Scale for the abundance ratio is shown along the lower x-axis. Error bars represent the 90% confidence interval calculated by ProRata. Average spectrum count is shown for each protein and indicated with an X. The scale is provided along the upper x-axis. .

to interaction specificity, as measured by isotopic abundance ratios. This is further illustrated by comparing results for Odp1, likely an artifactual protein, with an RNAP component, RpoB. The average spectrum count of Odp1 was 70.5 and the isotopic abundance ratio of this protein was 2, while the corresponding values for RpoB were 104.5 and 16, respectively. The combination of average spectrum count and isotopic abundance ratio provide evidence that the Odp1 protein is abundant in the affinity-isolated enrichment, yet is likely to be an artifactual interactor with RNAP. On the other hand, these results demonstrate that RpoB is both abundant (due to enrichment by the affinity isolation) and an authentic interactor with the bait.

I-DIRT Analysis of *R. palustris* RpoA/SPA2 Co-Isolates**.** Isotope ratios were obtained for 4 proteins that co-isolated with the *R. palustris* RpoA/SPA2 bait. The isotopic abundance ratios of these proteins are shown in Figure 4.3e and Table S5 (Supporting Information). As with all other isolations of RNAP, core components RpoB (8.0), RpoC (6.5), and RpoA (3.5), were among the largest isotope ratios. The other protein, RpoD (or RPA1288), is the *R. palustris* $\sigma_{70}$ factor, a known component of RNA polymerase.

No other isotopic abundance ratios were measured, despite qualitative identification of both putatively authentic (NusA, RnpO, and RpoH) and artifactual interactions (SecA, AtpA, and AtpD) as shown in Table S5 (Supporting Information). The lack of isotopic ratios for these proteins is attributable to our conservative approach for combining ProRata results [86], which required, for each protein, the presence of an isotopic ratio in each LC-MS-MS replicate in order to calculate a combined isotope ratio for that protein (see Experimental Procedures).

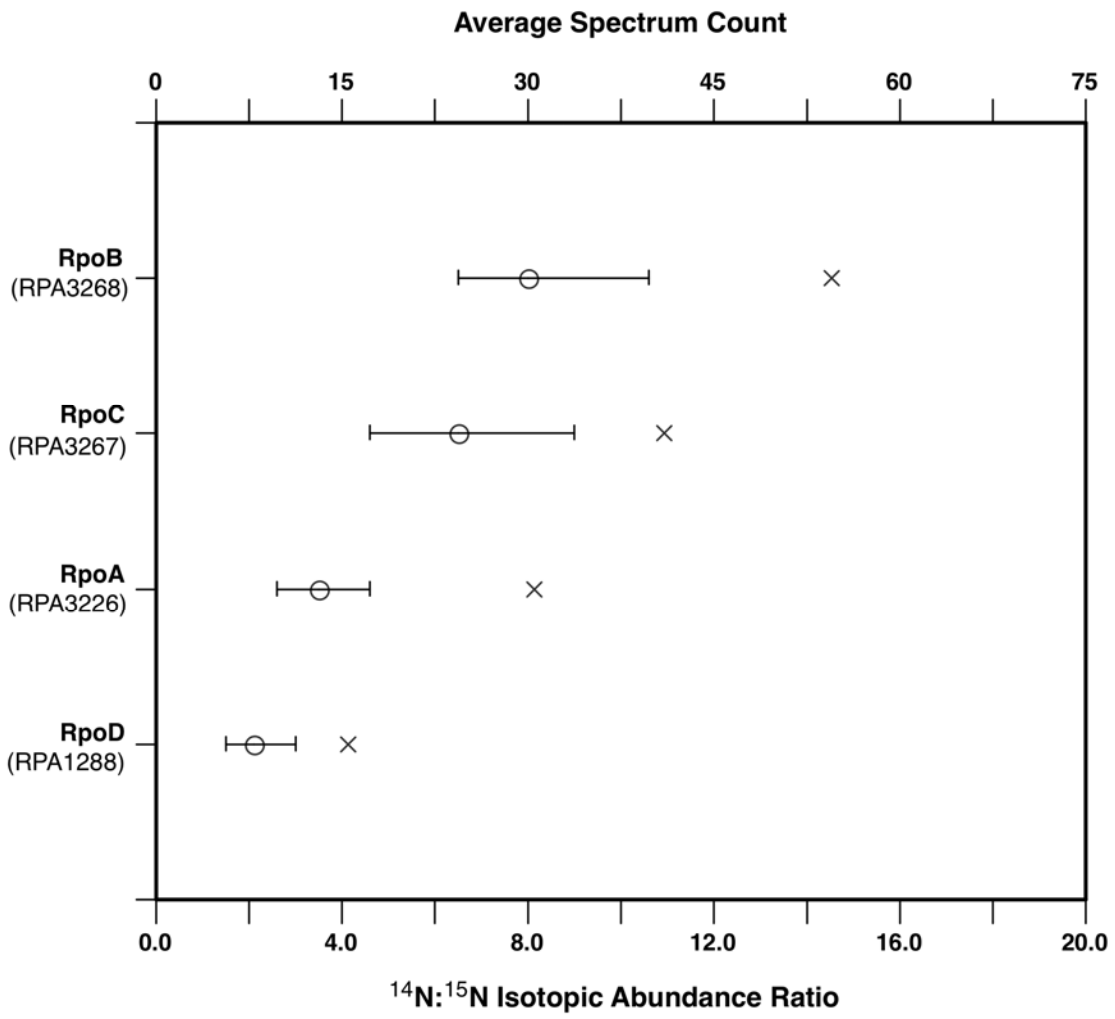



**Figure 4.3e.** I-DIRT measurements of affinity isolated RNA polymerase protein complex from the *R. palustris* plasmid-encoded RpoA/SPA2 bait protein. Isotopic abundance ratios for each protein are indicated with a circle; open circles indicate specific protein interactions, while filled circles represent non-specific protein interactions. Scale for the abundance ratio is shown along the lower x-axis. Error bars represent the 90% confidence interval calculated by ProRata. Average spectrum count is shown for each protein and indicated with an X. The scale is provided along the upper x-axis. .

**Quantitative Global Proteome Analysis of Plasmid-Bearing Strains.** A principal concern when adopting plasmid-encoded bait protein strategies for AP-MS is that an altered level of expression of the bait protein could result in collateral physiological responses in the microbe of study [81] resulting in the detection of protein interactions that are not biologically relevant. To investigate this possibility, we applied isotope labeling to perform comparative surveys of the microbial proteomes to measure the amount of bait protein produced by plasmid-bearing strains relative to its native counterpart in wild-type strains (Figure 1, right-hand workflow), as well as other proteins with higher or lower abundance.

*E. coli* RpoA/SPA2 Strain**.** From a total of 703 identified proteins (362 identified in both LC-MS-MS replicates), the ProRata program estimated $^{14}N$:$^{15}N$ isotopic ratios for 320 proteins from duplicate LC-MS-MS analyses of a single mixed pellet containing $^{14}N$ plasmid-bearing cells and $^{15}N$ wild type cells. Figure 4.4a shows the distribution of the $\log_2$ values of these ratios. (For the whole proteome measurements, it is convenient to express $^{14}N$:$^{15}N$ isotope ratios as base 2 logarithms, so that a protein with equal abundance in both strains, with $^{14}N$:$^{15}N$ = 1, has a $\log_2$ isotope ratio = 0.) Proteins of higher abundance in the $^{14}N$ strain will have positive $\log_2$ isotope ratios, while proteins of lower abundance in the $^{14}N$ strain will have negative $\log_2$ isotope ratios. A total of 51 proteins were measured with $\log_2$ $^{14}N$:$^{15}N$ isotopic ratios $\geq$1.0 (Table S6, Supporting Information), suggesting 2-fold or greater abundance in the *E. coli* RpoA/SPA2 plasmid encoded strain relative to the *E. coli* wild-type strain. A total of 24 proteins were measured at lower abundance in the plasmid-bearing strain, as indicated by $\log_2$ $^{14}N$:$^{15}N$

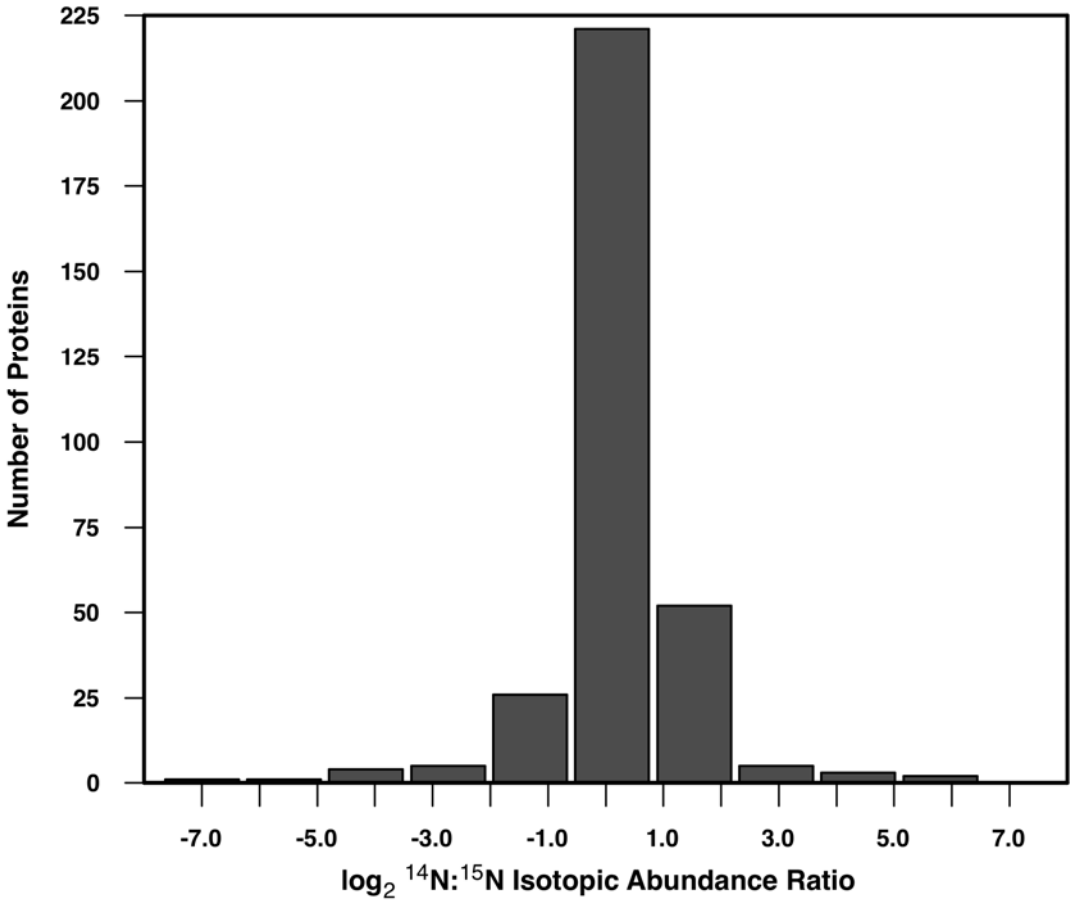



**Figure 4.4a.** Distributions of $^{14}N$:$^{15}N$ isotope ratios from proteome measurements of plasmid-encoded RpoA/SPA2 expressed in *E. coli* relative to wild-type strains.

Isotopic ratios $\leq$ -1.0. The remaining 245 proteins, representing 77% of all measured isotopic ratios, ranged from $\log_2$ -1.0 to 1.0. These data suggest that the majority of proteins in *E. coli* do not exhibit large changes in abundance attributable to the *in vivo* expression of the plasmid-encoded RpoA/SPA2 bait protein.

The $^{14}N$:$^{15}N$ ratios from measurement of 79 chromatographic features identified by ProRata, representing 15 peptides, indicated that total amount of the RpoA protein, including both native and affinity-tagged forms, was approximately 1.7 times greater ($\log_2$ value = +0.8) in the *E. coli* RpoA/SPA2 strain than for native RpoA in the *E. coli* wild type strain. This is an upper limit for the ratio of affinity-tagged RpoA from the plasmid-bearing strain to native RpoA in the wild type strain, because the plasmid-bearing strain may also contain RpoA originating from the chromosomal copy of the gene. For this protein and expression plasmid, therefore, it does not appear that the bait protein abundance is markedly different in the plasmid-bearing strain.

Among other proteins, the largest isotopic ratio measured was for the antibiotic resistance protein encoded by the SPA2 plasmid, gentamycin 3'-acetyl transferase (GenR). The $\log_2$ isotopic ratio of 5.0 indicating greater abundance in the plasmid encoded strain was expected, as the wild-type proteome does not express this protein.

Several other proteins exhibited either higher or lower abundance in the plasmid-bearing strain. While it is beyond the scope of this paper to analyze these changes in detail, we sought to identify any obvious trends by examining the classifications of these proteins into clusters of orthologous groups (COG, Table 4.2), their membership in KEGG metabolic pathways [98-100], and their membership in operons [101] (Tables 4.3-4.4).

**Table 4.2.** COG Functional Category Assignment of Proteins Measured at Different Abundances in Plasmid-bearing Microbial Strains relative to Wild-Type Strains

| Cellular Process | Functional Category | COG Code | Number of Proteins Measured at Higher Abundance[a] | | Number of Proteins Measured at Lower Abundance[a] | |
|---|---|---|---|---|---|---|
| | | | ***E. coli*** | ***R. palustris*** | ***E. coli*** | ***R. palustris*** |
| Information Storage and Processing | translation, ribosomal structure, and biogenesis | J | 1 | | 1 | |
| | Transcription | K | 3 | | | |
| | DNA replication, recombination, and repair | L | | | 1 | |
| Cellular Processes | Cell division and chromosome partitioning | D | 1 | | | |
| | Cell envelope biogenesis, outer membrane | M | 1 | 1 | 1 | |
| | post-translational modification, protein turnover | O | 5 | | 2 | 2 |
| | inorganic ion transport and metabolism | P | 3 | | 3 | |
| | signal transduction | T | | | | 2 |
| Metabolism | energy production and conversion | C | 7 | 1 | 2 | 19 |
| | amino acid transport and metabolism | E | 13 | 5 | 3 | 16 |
| | nucleotide transport and metabolism | F | 4 | | 1 | 2 |
| | carbohydrate transport and metabolism | G | 2 | | 2 | 4 |
| | coenzyme metabolism | H | 6 | | 1 | 3 |
| | lipid metabolism | I | 1 | 1 | | 8 |
| | secondary metabolites | Q | 2 | 1 | | 4 |
| Poorly Characterized Proteins | general function prediction only | R | 4 | | 2 | 11 |
| | function unknown | S | | | 2 | 1 |
| | no related COG | | 2 | 1 | 3 | 3 |
| - | GenR protein[b] | | 1 | 1 | | |

[a] Some proteins were assigned to multiple COGs, thus total number of proteins may not equal the total number of functional categories. [b] GenR protein is the antibiotic resistance protein encoded in the plasmid.

**Table 4.3.** Selected[a] proteins of increased abundance in plasmid-bearing *E. coli* relative to wild type. Multiple proteins coded by a single operon are grouped and shown in bold type.

| Accession No. | Locus | COG[b] | $\log_2$ isotope ratio (confidence interval) | Pathway | | | | | | | | | | | | | | | Protein description |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Valine, leucine and isoleucine biosynthesis | Pantothenate and CoA biosynthesis | ÿ t Pyruvate | Pyruvate metabolism | Taurine and hypotaurine metabolism | Glutamate metabolism | Alanine and aspartate metabolism | Valine, leucine and isoleucine degradation | Citrate cycle (TCA cycle) | Biosynthesis of siderophore group nonribosomal peptides | ABC transporters – General | ABC transporters – Organism-specific | Urea cycle and metabolism of amino groups | Purine metabolism | Sulfur metabolism | |
| P0A6F1 | B0032 | EF | 1.3 (1.1, 1.5) | | | | | | x | | | | | | | | | | CARA Carbamoyl-phosphate synthase small chain |
| **P30126** | **B0071** | **E** | **1.3 (0.9, 1.7)** | **x** | | | | | | | | | | | | | | | **LEUD 3-isopropylmalate dehydratase small subunit** |
| **P0A6A6** | **B0072** | **E** | **3.1 (2.7, 3.4)** | **x** | | | | | | | | | | | | | | | **LEU2 3-isopropylmalate dehydratase large subunit** |
| P0A9P0 | B0116 | C | 1.1 (0.8, 1.3) | | | | x | | | x | x | x | | | | | | | DLDH Dihydrolipoyl dehydrogenase |
| P0A790 | B0131 | H | 1 (0.5, 1.5) | | | | | | | x | | | | | | | | | PAND Aspartate 1-decarboxylase |
| P28635 | B0197 | R | 1.2 (0.8, 1.6) | | | | | | | | | | | x | x | | | | METQ D-methionine-binding lipoprotein metQ |
| **P0AEJ2** | **B0593** | **HQ** | **1.3 (1, 1.5)** | | | | | | | | | | **x** | | | | | | **ENTC Isochorismate synthase entC** |

**Table 4.3.** (Continued)

| Accession No. | Locus | COG[b] | $log_2$ isotope ratio (confidence interval) | Pathway | | | | | | | | | | | | | | | Protein description |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Valine, leucine and isoleucine biosynthesis | Pantothenate and CoA biosynthesis | Propanoate metabolism | Pyruvate metabolism | Taurine and hypotaurine metabolism | Glutamate metabolism | Alanine and aspartate metabolism | Valine, leucine and isoleucine degradation | Citrate cycle (TCA cycle) | Biosynthesis of siderophore group nonribosomal peptides | ABC transporters – General | ABC transporters – Organism-specific | Urea cycle and metabolism of amino groups | Purine metabolism | Sulfur metabolism | |
| **P10378** | **B0594** | **Q** | **1.8 (1.4, 2.2)** | | | | | | | | | | **x** | | | | | | **ENTE Enterobactin synthetase component E** |
| P0AGE9 | B0729 | C | 1.1 (0.8, 1.3) | | | x | | | | | | x | | | | | | | SUCD Succinyl-CoA ligase (ADP-forming) subunit alpha |
| **P0A6A3** | **B2296** | **C** | **1 (0.7, 1.3)** | | | **x** | **x** | **x** | | | | | | | | | | | **ACKA Acetate kinase** |
| **P0A9M8** | **B2297** | **C** | **1 (0.7, 1.3)** | | | **x** | **x** | **x** | | | | | | | | | | | **PTA Phosphate acetyltransferase** |
| P16703 | B2421 | E | 1.1 (0.3, 1.7) | | | | | | | | | | | | | | | x | CYSM Cysteine synthase B |
| P15254 | B2557 | F | 1 (0.8, 1.2) | | | | | | | | | | | | | | x | | PUR4 Phosphoribosylformylglycinamidine synthase |
| P0AFM2 | B2679 | E | 1 (0.6, 1.4) | | | | | | | | | | | x | x | | | | PROX Glycine betaine-binding periplasmic protein |
| P17846 | B2763 | P | 1.2 (0.6, 1.8) | | | | | | | | | | | | | | | x | CYSI Sulfite reductase (NADPH) hemoprotein beta-component |
| P21170 | B2938 | E | 1 (0.5, 1.4) | | | | | | x | | | | | | | x | | | SPEA Biosynthetic arginine decarboxylase |

**Table 4.3.** (Continued)

| Accession No. | Locus | COG[b] | $log_2$ isotope ratio (confidence interval) | Pathway | | | | | | | | | | | | | | | Protein description |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Valine, leucine and isoleucine biosynthesis | Pantothenate and CoA biosynthesis | Propanoate metabolism | Pyruvate metabolism | Taurine and hypotaurine metabolism | Glutamate metabolism | Alanine and aspartate metabolism | Valine, leucine and isoleucine degradation | Citrate cycle (TCA cycle) | Biosynthesis of siderophore group nonribosomal peptides | ABC transporters – General | ABC transporters – Organism-specific | Urea cycle and metabolism of amino groups | Purine metabolism | Sulfur metabolism | |
| **P0AB80** | **B3770** | **EH** | **1.8 (1.4, 2.3)** | **x** | **x** | | | | | | **x** | | | | | | | | **ILVE Branched-chain-amino-acid aminotransferase** |
| **P05791** | **B3771** | **EG** | **5.3 (4.7, 5.9)** | **x** | **x** | | | | | | | | | | | | | | **ILVD Dihydroxy-acid dehydratase** |
| P23908 | B3957 | E | 1.7 (1.4, 2) | | | | | | | | | | | | | x | | | ARGE Acetylornithine deacetylase |
| P11447 | B3960 | E | 1.4 (0.7, 1.9) | | | | | | | x | | | | | | x | | | ARLY Argininosuccinate lyase |
| **P15640** | **B4005** | **F** | **1.6 (1.2, 1.9)** | | | | | | | | | | | | | | **x** | | **PUR2 Phosphoribosylamine–glycine ligase** |
| **P15639** | **B4006** | **F** | **1.3 (1, 1.6)** | | | | | | | | | | | | | | **x** | | **PUR9 Bifunctional purine biosynthesis protein purH** |

[a]Not included are pathways containing <2 proteins, and proteins that were the sole identified member of a pathway. [b]Descriptions of COG categories are shown in Table 4.2.

**Table 4.4.** Selected[a] proteins of decreased abundance in plasmid-bearing *E. coli* relative to wild type. Multiple proteins coded by a single operon are grouped and shown in bold type.

| AccNo | Locus | COG[b] | Log2 isotope ratio (confidence interval) | Pathway: Phenylalanine, tyrosine and tryptophan biosynthesis | Pathway: Two-component system – General | Pathway: Pentose and glucuronate interconversions | Pathway: ABC transporters – General | Pathway: ABC transporters – Organism-specific | Protein description |
|---|---|---|---|---|---|---|---|---|---|
| **P00895** | **B1264** | **EH** | **-1.7 (-2.1, -1.5)** | **x** | **x** | | | | **TRPE Anthranilate synthase component 1** |
| **P00904** | **B1263** | **EH** | **-1 (-1.3, -0.8)** | **x** | **x** | | | | **TRPG Anthranilate synthase component II** |
| P0A8G3 | B3092 | G | -1.7 (-2.1, -1.2) | | | x | | | UXAC Uronate isomerase |
| P0AAB6 | B2042 | M | -1.3 (-1.7, -1) | | | x | | | GALF UTP–glucose-1-phosphate uridylyltransferase |
| P0AEM9 | B1920 | ET | -4.2 (-5.2, -3.5) | | | | x | x | FLIY Cystine-binding periplasmic protein |
| P23843 | B1243 | E | -3 (-3.2, -2.8) | | | | x | x | OPPA Periplasmic oligopeptide-binding protein |
| P37329 | B0763 | P | -5.2 (-6.3, -4.1) | | | | x | x | MODA Molybdate-binding periplasmic protein |

[a]Not included are pathways containing <2 proteins, and proteins that were the sole identified member of a pathway. [b]Descriptions of COG categories are shown in Table 4.2.

The COG functional category containing most of the proteins that were more abundant in the plasmid-bearing strain was amino acid transport and metabolism. Five of the 12 more abundant proteins in this category are involved in branched-chain amino acid (valine, leucine, and isoleucine) biosynthesis [102]. $Log_2$ $^{14}N$:$^{15}N$ isotopic abundance ratios of 5.3 for dihydroxy-acid dehydratase (IlvD), 3.1 for 3-isopropylmalate dehydratase large subunit (Leu2), 1.8 for the branched-chain-amino-acid transferase (IlvE), 1.3 for 3-isopropylmalate dehydratase subunit (LeuD), and 1.2 for acetolactate synthase isoenzyme 1 large subunit (IlvB) are indicative of higher abundance of these proteins in the plasmid encoded bait-expressing strain. Genes encoding Leu2 and LeuD occur in the same operon, as do genes encoding IlvD and IlvE, suggesting that mRNA transcription for these enzymes increases under the plasmid load. The increased abundances of these specific proteins, which collectively account for 8 of 12 total steps for branched-chain amino acid biosynthesis, could potentially be attributed to the increased protein expression due to the presence of the plasmid.

Among other proteins that were more abundant in the plasmid-bearing strain, acetate kinase and phosphate acetyltransferase, transcribed as a unit, are involved in interconversions among acetate (or propanoate), acetyl phosphate (propanoyl phosphate), and acetyl-CoA (propanoyl-CoA). Isochorismate synthase (EntC) and enterobactin synthetase component E (EntE) also share an operon, and are involved in synthesis of iron-scavenging siderophore compounds. Several heat shock and chaperone proteins in the post-translational modification and protein turnover functional category also were more abundant in the plasmid-encoded strain, consistent with a previous report [103].

Proteins with decreased abundance in the RpoA/SPA2 strain also were dominated by the amino acid transport and metabolism COG category. However, the anthranilate synthase proteins, TrpE and TrpG (encoded within the same operon), occur not in the isoleucine, leucine and valine pathway, but instead are in the phenylalanine, tyrosine and tryptophan biosynthesis pathway. Although not sharing operon membership, three periplasmic proteins annotated as ABC transporters (FliY, OppA, ModA) were significantly less abundant in the plasmid-bearing strain.

*R. palustris* RpoA/SPA2 Strain**.** From a total of 1052 identified proteins (673 identified in both LC-MS-MS replicates), $\log_2$ $^{14}N$:$^{15}N$ isotopic ratios for 598 proteins were estimated from duplicate LC-MS-MS analyses of a single mixed pellet containing $^{14}N$ plasmid-bearing cells and $^{15}N$ wild type cells. The distribution of these ratios is shown in Figure 4.4b. A total of 11 proteins were estimated with $\log_2$ $^{14}N$:$^{15}N$ isotopic ratios $\geq$1.0 (Table S7, Supporting Information), suggesting 2-fold or greater expression in the *R. palustris* RpoA/SPA2 bait-expressing strain relative to the wild-type *R. palustris* strain. A total of 71 proteins were measured at lower abundance in the plasmid-bearing strain, as indicated by $\log_2$ $^{14}N$:$^{15}N$ isotopic ratios $\leq$ -1.0. The remaining 516 proteins, representing 86% of all measured isotopic ratios, ranged from $\log_2$ -1.0 to 1.0, shown in Figure 4b. These data suggest that *in vivo* expression of the plasmid-encoded RpoA/SPA2 bait protein does not affect abundances of the majority of the *R. palustris* proteins.

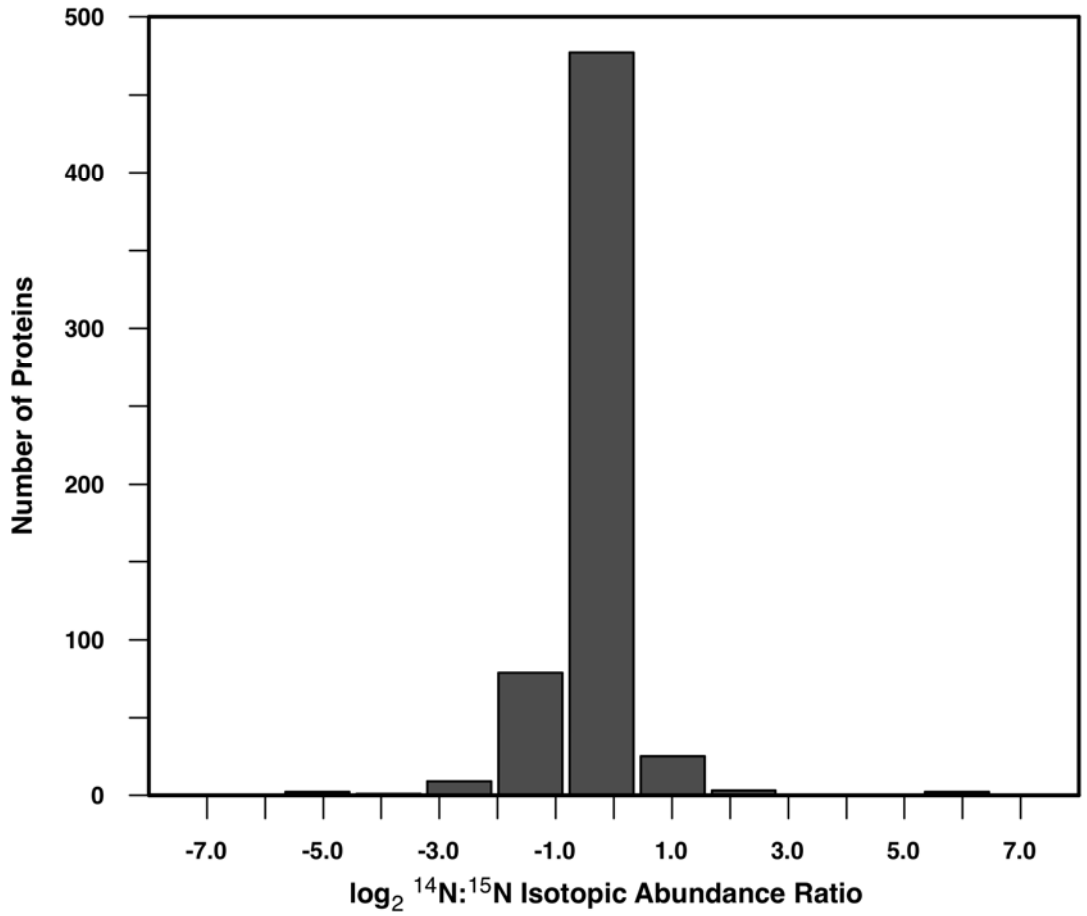



**Figure 4.4b.** Distributions of $^{14}N$:$^{15}N$ isotope ratios from proteome measurements of plasmid-encoded RpoA/SPA2 expressed in *R. palustris* relative to wild-type strains.

The $^{14}N$:$^{15}N$ ratios from measurement of 60 chromatographic features by ProRata, representing 15 peptides, indicated that total expression of the RpoA protein, including both native and affinity-tagged forms, in the *R. palustris* RpoA/SPA2 strain was approximately 0.8 times the level for native RpoA in the *R. palustris* wild type strain ($\log_2$ value = -0.3). As for the *E. coli* experiment described above, the abundance of RpoA in the plasmid-bearing strain is not significantly different from the wild type, and is actually slightly decreased. The other protein encoded only in the plasmid, GenR, yielded the second largest isotope ratio in this experiment, with $\log_2$ isotopic ratio of 5.4.

Surprisingly, the largest observed $\log_2$ isotopic ratio of 5.9 indicated that an unknown protein (RPA1422) was considerably more abundant in the plasmid-bearing strain than in the wild type. The 14 peptide quantifications measured for this protein were all of the $^{14}N$ isotopolog of this protein originating from the RpoA/SPA2 strain; no $^{15}N$ peptides expressed by the wild-type strain were identified. A recent proteome study of wild-type *R. palustris* identified the RPA1422 unknown protein as more highly abundant under two growth conditions: benzoate and nitrogen fixation [43]. In contrast, the cell cultures in the present study were grown under anaerobic photoheterotrophic growth with succinate as the carbon source. These data suggest that the presence of the SPA2 plasmid potentially affects the abundance of the RPA1422 protein.

The remainder of proteins with $\log_2$ ratios $\geq 1$ in the *R. palustris* RpoA/SPA2 strain were annotated [89] in COG functional categories shown in Table 2. In particular, two proteins annotated in the amino acid transport and metabolism functional category, MetC (RPA2350) and OhaS (RPA4251), are located in biosynthetic pathways for the sulfur-containing amino acids cysteine and methionine. Specifically, O-acetylhomoserine

sulfhydrylases MetC [104] and OhaS [105] are enzymes involved in the conversion of O-acetyl-L-homoserine and methanethiol to L-methionine and acetate. Proteins with altered abundances that are also encoded within shared operons included acetolactate synthase (higher abundance in plasmid-bearing strain, Table 4.5) and components of pyruvate dehydrogenase, alcohol/aldehyde dehydrogenase, and carbon monoxide dehydrogenase (lower abundance in plasmid-bearing strain, Table 4.6). Acetolactate synthase uses pyruvate as a substrate for synthesis of the amino acids isoleucine, leucine, and valine. However, pyruvate dehydrogenase, which also uses pyruvate but produces acetyl CoA, is over 2-fold less abundant in plasmid vs. wild type. These two results suggest that the plasmid-bearing strain preferentially converts pyruvate to products other than acetyl-CoA. Consistent with this observation is the fact that acetyl Co-A synthetase (RPA0211) was also less abundant in the plasmid-bearing strain vs. wild type. It is difficult to formulate reasonable explanations for the markedly lower abundance levels of carbon monoxide dehydrogenase subunits, as well as the aldehyde and alcohol dehydrogenases RPA1205 and 1206, in the plasmid-bearing strain.

Despite quite different metabolic characteristics for the two species, both *E. coli* and *R. palustris* strains that harbor the plasmid encoding the RpoA bait showed increased abundance in proteins involved in biosynthesis of isoleucine, leucine, and valine. Although not highly over-expressed in either species, RpoA contains a relatively large proportion of isoleucine, leucine, and valine: *E. coli* RpoA contains more of these amino acids than ~85% of predicted *E. coli* proteins, and RpoA in *R. palustris* contains more of these amino acids than ~80% of predicted *R. palustris* proteins. In the minimal media

**Table 4.5.** Selected[a] proteins of increased abundance in plasmid-bearing *R. palustris* relative to wild type. Multiple proteins coded by a single operon are grouped and shown in bold type.

| Locus | COG[b] | $log_2$ isotope ratio (confidence interval) | Pathway | | | | | | Protein description |
|---|---|---|---|---|---|---|---|---|---|
| | | | Butanoate metabolism | C5-Branched dibasic acid metabolism | Pantothenate and CoA biosynthesis | Valine, leucine and isoleucine biosynthesis | Cysteine metabolism | Methionine metabolism | |
| **RPA2031** | **EH** | **1.1 (0.9, 1.3)** | **x** | **x** | **x** | **x** | | | **ilvI acetolactate synthase (large subunit)** |
| **RPA2032** | **E** | **1.2 (0.9, 1.6)** | **x** | **x** | **x** | **x** | | | **ilvH acetolactate synthase (small subunit)** |
| RPA2317 | I | 1.2 (0.8, 1.6) | x | | | | | | gctB putative CoA transferase, small subunit B |
| RPA2350 | E | 1.7 (1.2, 2.2) | | | | | x | x | metC putative O-acetylhomoserine sulfhydrylase |
| RPA4251 | E | 1.2 (1, 1.4) | | | | | x | x | oahs O-acetylhomoserine sulfhydrylase |

[a]Not included are pathways containing <2 proteins, and proteins that were the sole identified member of a pathway. [b]Descriptions of COG categories are shown in Table 4.2.

**Table 4.6.** Selected[a] proteins of decreased abundance in plasmid-bearing *R. palustris* relative to wild type. Multiple proteins coded by a single operon are grouped and shown in bold type.

| Locus | COG[b] | $log_2$ isotope ratio (confidence interval) | Pathway | | | | | | | | | | | | | | | | | | | | | | | Protein description |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Glycolysis / Gluconeogenesis | Propanoate metabolism | Pyruvate metabolism | Reductive carboxylate cycle (CO2 fixation) | Butanoate metabolism | ABC transporters – General | Fatty acid biosynthesis | Fatty acid metabolism | Bile acid biosynthesis | 3-Chloroacrylic acid degradation | Valine, leucine and isoleucine degradation | beta-Alanine metabolism | Citrate cycle (TCA cycle) | Carbon fixation | Glyoxylate and dicarboxylate metabolism | Pantothenate and CoA biosynthesis | Nitrogen metabolism | Alanine and aspartate metabolism | Glycine, serine and threonine metabolism | Valine, leucine and isoleucine biosynthesis | Purine metabolism | Methane metabolism | Trinitrotoluene degradation | |
| RPA0211 | I | -1 (-1.2, -0.9) | x | x | x | x | | | | | | | | | | | | | | | | | | | | acetyl-CoA synthetase |
| RPA0532 | QR | -1.1 (-1.5, -0.8) | | | | | x | | | | | | | | | | | | | | | | | | | beta-ketothiolase, acetoacetyl-CoA reductase |
| RPA0668 | E | -1.4 (-1.6, -1.1) | | | | | | x | | | | | | | | | | | | | | | | | | hbaE putative ABC transporter subunit, substrate-binding component |
| RPA1110 | QR | -1 (-1.7, -0.2) | | | | | | | x | | | | | | | | | | | | | | | | | putative 3-oxoacyl-acyl carrier protein reductase |
| **RPA1205** | **C** | **-4.8 (-5.1, -4.6)** | **x** | | | | | | | **x** | **x** | **x** | | | | | | | | | | | | | | **putative alcohol dehydrogenase** |
| **RPA1206** | **C** | **-5.1 (-5.3, -4.8)** | **x** | **x** | **x** | | **x** | | | **x** | **x** | **x** | **x** | **x** | | | | | | | | | | | | **aldehyde dehydrogenase** |
| RPA1226 | C | -1.4 (-1.7, -1.1) | | | | x | | | | | | | | | x | | | | | | | | | | | putative 2-oxoglutarate ferredoxin oxidoreductase, alpha subunit |
| RPA1236 | I | -1.6 (-1.9, -1.2) | | x | | | | | | x | | | x | x | | | | | | | | | | | | putative acyl-CoA dehydrogenase |
| RPA1242 | IQ | -1 (-1.3, -0.6) | | | | | | | | x | | | | | | | | | | | | | | | | fadD1 putative long-chain-fatty-acid CoA ligase |
| RPA1298 | R | -1 (-1.2, -0.7) | | | | | | | x | | | | | | | | | | | | | | | | | putative 3-oxoacyl-acyl carrier protein synthase III |
| RPA1560 | C | -1 (-1.3, -0.6) | | | | | | | | | | | | | | x | x | | | | | | | | | cbbS bisphosphate carboxylase small chain |

**Table 4.6.** (Continued)

| Locus | COG[b] | $log_2$ isotope ratio (confidence interval) | Pathway | | | | | | | | | | | | | | | | | | | | | | | Protein description |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Glycolysis / Gluconeogenesis | Propanoate metabolism | Pyruvate metabolism | Reductive carboxylate cycle (CO2 fixation) | Butanoate metabolism | ABC transporters – General | Fatty acid biosynthesis | Fatty acid metabolism | Bile acid biosynthesis | 3-Chloroacrylic acid degradation | Valine, leucine and isoleucine degradation | beta-Alanine metabolism | Citrate cycle (TCA cycle) | Carbon fixation | Glyoxylate and dicarboxylate metabolism | Pantothenate and CoA biosynthesis | Nitrogen metabolism | Alanine and aspartate metabolism | Glycine, serine and threonine metabolism | Valine, leucine and isoleucine biosynthesis | Purine metabolism | Methane metabolism | Trinitrotoluene degradation | |
| RPA1741 | E | -2.2 (-2.4, -1.9) | | | | | | x | | | | | | | | | | | | | | | | | | possible branched-chain amino acid transport system substrate-binding protein |
| RPA1763 | IQ | -1 (-1.6, -0.4) | | | | | | | x | | | | | | | | | | | | | | | | | putative long-chain-fatty-acid CoA ligase |
| RPA1943 | H | -1.9 (-2.2, -1.5) | | | | | | | | | | | | | | | | x | | | | | | | | 91uts putative ketopantoate reductase |
| RPA2061 | C | -3.5 (-3.7, -3.3) | | | | | | | | | | | | | | | | | x | | | | | | | nosZ nitrous-oxide reductase precursor NosZ |
| **RPA2863** | **C** | **-1.4 (-1.8, -1.1)** | **x** | | **x** | | | | | | | | **x** | | **x** | | | | | **x** | **x** | | | | | **dihydrolipoamide dehydrogenase, E3 Component of Pyruvate dehydrogenase multienzyme complex** |
| **RPA2864** | **C** | **-1.3 (-1.6, -1.1)** | **x** | | **x** | | | | | | | | | | | | | | | **x** | | | | | | **Dihydrolipoamide acetyltransferase** |
| **RPA2866** | **C** | **-1.6 (-1.8, -1.4)** | **x** | | **x** | | **x** | | | | | | | | | | | | | **x** | | **x** | | | | **pyruvate dehydrogenase E1 beta subunit** |
| **RPA2867** | **C** | **-1.5 (-1.9, -1)** | **x** | | **x** | | **x** | | | | | | | | | | | | | **x** | | **x** | | | | **pyruvate dehydrogenase E1 alpha subunit** |
| RPA2975 | E | -1.2 (-1.6, -0.8) | | | | | | | | | | | | | | | | | | | x | | | | | serA1 putative phosphoglycerate dehydrogenase |
| RPA2977 | F | -1.3 (-1.7, -0.9) | | | | | | | | | | | | | | | | | | | | | x | | | nrd ribonucleotide reductase |
| RPA3289 | I | -1 (-1.3, -0.8) | | x | | | | | | x | | | x | x | | | | | | | | | | | | acyl-CoA dehydrogenase |
| RPA3472 | EG | -1 (-1.3, -0.6) | | | | | | | | | | | | | | | | x | | | | x | | | | ilvD1 dihydroxy-acid dehydratase |
| RPA3756 | C | -1.6 (-2.1, -1.2) | | | x | x | | | | | | | | | x | x | x | | | | | | | | | malate dehydrogenase-like protein |

**Table 4.6.** (Continued)

| Locus | COG[b] | $log_2$ isotope ratio (confidence interval) | Pathway | | | | | | | | | | | | | | | | | | | | | | | Protein description |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Glycolysis / Gluconeogenesis | Propanoate metabolism | Pyruvate metabolism | Reductive carboxylate cycle (CO2 fixation) | Butanoate metabolism | ABC transporters – General | Fatty acid biosynthesis | Fatty acid metabolism | Bile acid biosynthesis | 3-Chloroacrylic acid degradation | Valine, leucine and isoleucine degradation | beta-Alanine metabolism | Citrate cycle (TCA cycle) | Carbon fixation | Glyoxylate and dicarboxylate metabolism | Pantothenate and CoA biosynthesis | Nitrogen metabolism | Alanine and aspartate metabolism | Glycine, serine and threonine metabolism | Valine, leucine and isoleucine biosynthesis | Purine metabolism | Methane metabolism | Trinitrotoluene degradation | |
| RPA3848 | E | -1.3 (-1.5, -1) | | | | | | | | | | | | | | | | | x | | x | | | | | gcvT2 glycine cleavage system protein T2 |
| RPA3974 | C | -1.2 (-1.5, -1) | | | | | | | | | | | | | | | | | | | | | | x | x | putative dehydrogenase |
| RPA4019 | E | -1.7 (-1.8, -1.5) | | | | | | x | | | | | | | | | | | | | | | | | | putative branched-chain amino acid ABC transporter system substrate-binding protein |
| RPA4045 | E | -1.7 (-1.8, -1.6) | | | | | | x | | | | | | | | | | | | | | | | | | possible branched-chain amino acid ABC transport system substrate-binding protein |
| RPA4644 | C | -1.1 (-1.3, -0.9) | | | | | | | | | | | | | | x | | | | | | | | | | cbbP phosphoribulokinase (phosphopentokinase) (PRK) |
| **RPA4666** | **C** | **-2.2 (-2.5, -1.8)** | | | | | | | | | | | | | | | | | | | | | | **x** | **x** | **coxS, 92uts carbon-monoxide dehydrogenase small subunit** |
| **RPA4667** | **C** | **-2.3 (-2.5, -2.1)** | | | | | | | | | | | | | | | | | | | | | **x** | | | **coxL, CO dehydrogenase subunit** |
| **RPA4668** | **C** | **-1.7 (-2, -1.4)** | | | | | | | | | | | | | | | | | | | | | | **x** | **x** | **coxM, cutM carbon monoxide chain C** |
| RPA4813 | E | -1.1 (-1.3, -0.8) | | | | | | x | | | | | | | | | | | | | | | | | | periplasmic binding protein of ABC transporter |

[a]Not included are pathways containing <2 proteins, and proteins that were the sole identified member of a pathway. [b]Descriptions of COG categories are shown in Table 4.2.

Used in this study, the bacteria are forced to synthesize all amino acids that will be used to subsequently synthesize proteins. The somewhat increased proportion of isoleucine, leucine and valine in RpoA may result in a detectably increased activity within this pathway.

## *4.4 Conclusions*

The study of protein complexes by AP-MS involves the choice of a bait expression strategy. Several large-scale reports of protein-protein interactions for model organisms in recent years have used genetic systems for expression of chromosome-encoded bait proteins. As an alternative to such chromosomal strategies, protein-protein interaction studies could potentially be extended to a larger range of organisms by a practical and flexible plasmid-based system for expression of bait proteins in multiple gram-negative microbial species [80]. In this study, we have compared chromosome- and plasmid-encoding strategies in both a model biological species, *E. coli*, and *R. palustris*, a species that is less-studied, but environmentally ubiquitous and metabolically diverse.

For the three bait proteins studied, the major difference between chromosome and plasmid encoding was success or failure in expression of the affinity-tagged bait protein. In *E. coli*, all three bait proteins were expressed from chromosome-encoded versions, while two of the three were not expressed from the plasmid. Similarly, only one of the three baits showed expression from plasmid in *R. palustris*, despite previous success in expressing all three proteins from a plasmid encoding a different affinity tag [80]. This limited study, along with our previous results, suggests that chromosomal encoding may have greater generality across a range of genes, while plasmid encoding will require the

testing of several affinity tags and orientations (i.e., N-terminal or C-terminal fusion) to find a combination that results in successful expression of the tagged bait.

If the affinity-tagged bait can be expressed, however, it appears that both chromosome and plasmid encoding strategies perform comparably. Using RNAP as a model protein complex, the I-DIRT method showed that for both strategies, core components of the complex exhibited isotopic abundance ratios that indicated specific interaction with the bait protein. Abundant background proteins were also identified in these AP-MS experiments, but their isotopic ratios clearly indicated artifactual (post-mixing) association with the bait protein, in affinity isolations from both chromosome- and plasmid-encoded strains.

From a more global perspective of the proteome, isotope ratio measurements revealed no significant over-expression of bait protein from plasmid-bearing strains of *E. coli* or *R. palustris*, although this result is confined to a single protein (RpoA) and affinity tag (SPA). Beyond the bait, most proteins for which isotope ratios could be measured in both *E. coli* and *R. palustris* showed no significant change in abundance attributable to the presence of the plasmid, although some were either increased or decreased in abundance. The majority of proteins with altered abundances in plasmid-bearing strains were involved in amino acid biosynthesis, which may reflect the metabolic load [81] imposed by the biosynthesis of plasmid-encoded proteins in a minimal medium that requires the bacteria to produce all needed amino acids. This effect suggests that perhaps we did not successfully express RpoB and RpoC as baits in plasmid-bearing strains because these proteins are significantly larger (each > 150 kDa) than RpoA (< 40 kDa),

thus requiring production of larger amounts of amino acids.

In sum, we have demonstrated the utility of local (complex-specific) and global (proteome-level) isotope ratio measurements as diagnostic tools for objectively comparing different bait protein expression strategies for AP-MS. It is our hope that these tools will facilitate expanded studies of protein-protein interactions in bacteria, leading eventually to a better understanding of how proteins interact to form complexes and function within the cell.

# Chapter 5
# Evaluation of Periplasmic Protein Enrichment Protocols using Stable Isotope Labeling and Mass Spectrometry

The data presented here are planned for submission to the *Journal of Proteome Research*.

W. Judson Hervey, IV, Adam M. Tebbe, Dale A. Pelletier, and Gregory B. Hurst.

Microbial cell growth, $^{15}$N-metabolic labeling, protein isolation, sample preparation, primary data acquisition, and data analysis were performed by W. Judson Hervey, IV.

Dr. Gregory B. Hurst assisted with experimental design and data analysis.

*Abstract*

Determination of protein localization provides clues regarding function. Isolation of protein fractions from selected cellular structures is a prerequisite to detailed proteomic characterization of microbial cells. Information regarding protein subcellular localization is often difficult to obtain by existing qualitative proteomic techniques due to the presence of non-specific, contaminant proteins among fractions of enriched protein mixtures.

To assist with subcellular localization, we apply a quantitative proteomics approach to characterize protein localization to the periplasm in *Escherichia coli* K12. Enrichment of periplasmic protein fractions was performed by two separate cold osmotic shock protocols: $H_2O$ or $GdCl_3$. Measurements of $^{15}N$:$^{14}N$ protein isotope ratios between $^{15}N$-labeled periplasmic and $^{14}N$-labeled whole cell lystate proteome fractions, expressed as PP:WCL, provided an indication of abundance in the periplasmic fraction. PP:WCL ratios highlighted enrichment of 31 annotated periplasmic proteins by the $H_2O$ protocol and 29 annotated periplasmic proteins by the $GdCl_3$ protocol. When combined with an informatics strategy to identify amino-terminal signal peptide cleavages, a post-translational modification present among the majority of periplasmic proteins, 26 proteins were enriched in either the $H_2O$ or $GdCl_3$ protocol and included this modification. As expected, substrate binding proteins constituted the majority of these 26 proteins. However, uncharacterized proteins YnfD, YraP, YnhG, and YgiW were also detected with signal peptide cleavages, though at lower PP:WCL ratios not indicative of periplasmic abundance.

Collectively, quantitative measurement of PP:WCL isotope ratios among enriched periplasmic proteins consistent with identification of signal peptide cleavages provide increased experimental evidence of potential periplasmic localization in comparison to qualitative proteomics platforms.

## *5.1 Introduction*

The cellular location of a protein imparts significant information regarding its biological function(s). Frequently, protein localization studies involve combinations of biochemical methods and/or molecular techniques to determine the cellular distribution of a single protein [2]. Standard biochemical methods may involve fractionation of a specific cellular structure and subsequent activity measurement to determine presence of an enzyme in the fraction. Molecular techniques may encompass fusion of an exogenous tag to a single protein, such as a fluorophore or affinity epitope, with measurement by confocal microscopy to determine cellular location(s). Implementations of biochemical and molecular methods are quite effective for examining localization of a single protein throughout the cell. Often, combinations of biochemical and molecular methods require each protein to be studied on an individual basis, reducing study throughput to a "one-by-one" approach. Considering the dynamic nature of the cell, undergoing simultaneous synthesis, degradation, and trafficking thousands of proteins at any given time, it would be beneficial to monitor distribution of multiple proteins for more detailed understanding of protein localization.

A viable alternative is integration of proteomics approaches in conjunction with biochemical and/or molecular techniques to significantly expand the scope of protein

localization studies. Proteome studies of a "targeted" subset of proteins, such as a protein fraction, organelle, or other cellular structure combine biochemical isolation of a selected group of proteins with characterization by qualitative liquid chromatography tandem mass spectrometry (LC-MS-MS) [32, 106, 107]. Integration of proteome techniques in protein localization studies is particularly advantageous by providing identification of the entire protein content of targeted biochemical isolations in a single data acquisition, providing localization insight for multiple proteins. Furthermore, proteome techniques permit study of a wild-type organism, thus capturing an organism's native biological state free from potential perturbations introduced by genetic modification. Though more robust, a disadvantage of using proteome techniques is that numerous qualitative protein identifications obtained from LC-MS-MS analysis are representative of the desired or "targeted" proteins of interest amid abundant, co-isolating proteins detected as an artifact of protein isolation. Implementations of qualitative subtractive proteomic methods yielded a number of novel candidates localizing to the eukaryotic nuclear envelope [108], obtained through the subtraction of qualitative protein identifications common to two distinct protein isolations. Subtraction of common protein identifications between the protein isolations accounted for contaminant, co-isolating proteins occurring as an artifact of the biochemical isolation.

More recently, protein localization studies have benefitted from application of quantitative techniques to targeted proteome analyses by imparting advantages of both molecular tagging and biochemical separations. Incorporation of isotopic signatures via quantitative techniques effectively "tags" an entire proteome at the molecular level,

alleviating a constraint of singular, "one-by-one" molecular tagging approaches. Biochemical separation techniques provide methods to enrich, or "target," the subset of proteins bearing an isotopic signature. Measurement of isotopic signatures by LC-MS-MS provides a means to distinguish "targeted" proteins of interest from contaminant, co-isolates amid biochemical enrichments.

Applications of quantitative techniques to targeted proteome analyses have focused on organelles and structures of higher eukaryotic organisms [32, 33, 107, 109]. Despite the lack of membrane-bound organelles, prokaryotic organisms also use complex, dynamic mechanisms of localization to traffic the proteome. Targeted analyses of the periplasmic proteome have been performed using the combination of protein isolation by cold osmotic shock-based protocols [44, 45], followed by their identification by qualitative LC-MS-MS in *Escherichia coli* [110], and other species [111, 112]. The periplasm, or "periplasmic space," is the region of the cell located between the inner membrane and outer membrane, as shown in Figure 5.1. Periplasmic proteins function in this region, performing important roles in electron transport, binding of nutrients, cell wall biosynthesis, and modification of molecules that will eventually enter the cytoplasm [113]. Thus, periplasmic proteins are of considerable biological interest in elucidating interactions between microbial cells and the external environment. As an intermediate transport zone between the inner cell and surrounding external environments, the periplasm may best be conceptualized as an exceptionally dynamic region of the cell rather than a strictly rigid compartmentalized cellular structure [113]. A number of proteins that function in the periplasm may be performing a function

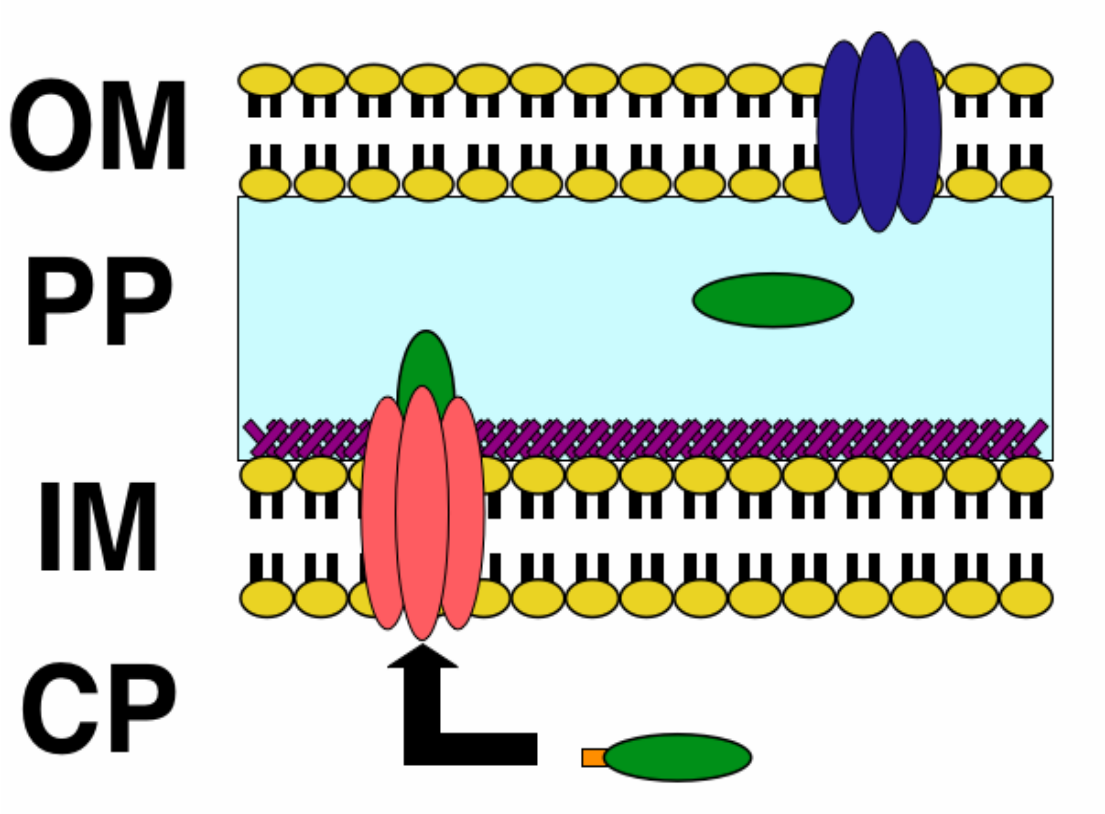



**Figure 5.1.** Protein export to the periplasm via the general secretory (Sec) pathway. Four general structures of the gram-negative cell membrane are shown above: the outer membrane (OM), the periplasm (PP), the inner membrane (IM), and the cytoplasm (CP). The majority of periplasmic proteins enter the periplasm via the general secretory (Sec) pathway. Here, an example periplasmic protein (green) bearing Sec-dependent amino-terminal signal peptide (orange) is directed through an inner membrane protein to the periplasm.

elsewhere within the cell, or be *en route* to the periplasm via the general secretory pathway after protein synthesis at the ribosome. Detailed characterization of localization among this group of proteins by existing qualitative LC-MS-MS techniques has proven difficult, as periplasmic protein enrichments frequently contain "contaminant" proteins released as an artifact of enrichment [110]. This previous study examined protein database entries for putative general secretory (Sec) pathway amino-terminal signal peptides among identifications to indicate protein localization to the periplasm, rather than direct experimental measurement by mass spectrometry. The elucidation of the twin-arginine translocation (TAT) pathway [114] has demonstrated that periplasmic proteins may also contain an amino-terminal signaling motif of twin arginine residues. Also, some TAT substrate proteins translocate the inner membrane bound to other proteins in complex that do not contain a recognizable amino-terminal signaling motif. Thus, sole presence of a Sec pathway amino-terminal signal peptide sequence motif may not be sufficient to implicate periplasmic function among proteins enriched from the periplasm.

In this study, we describe a quantitative proteomics strategy to differentiate $^{15}$N-labeled periplasmic isolate from $^{14}$N-labeled whole cell lysate among protein mixtures of the *Escherichia coli* K12 proteome. Figure 5.2 outlines our strategy of incorporating isotopic signatures via $^{15}$N metabolic labeling, followed by cold osmotic shock isolation to obtain a "heavy" enrichment of periplasmic proteins (PP). The "targeted,"

$^{15}$N-labeled fraction of enriched periplasmic proteins is then mixed with a whole cell lysate (WCL) protein fraction containing predominantly the most abundant natural

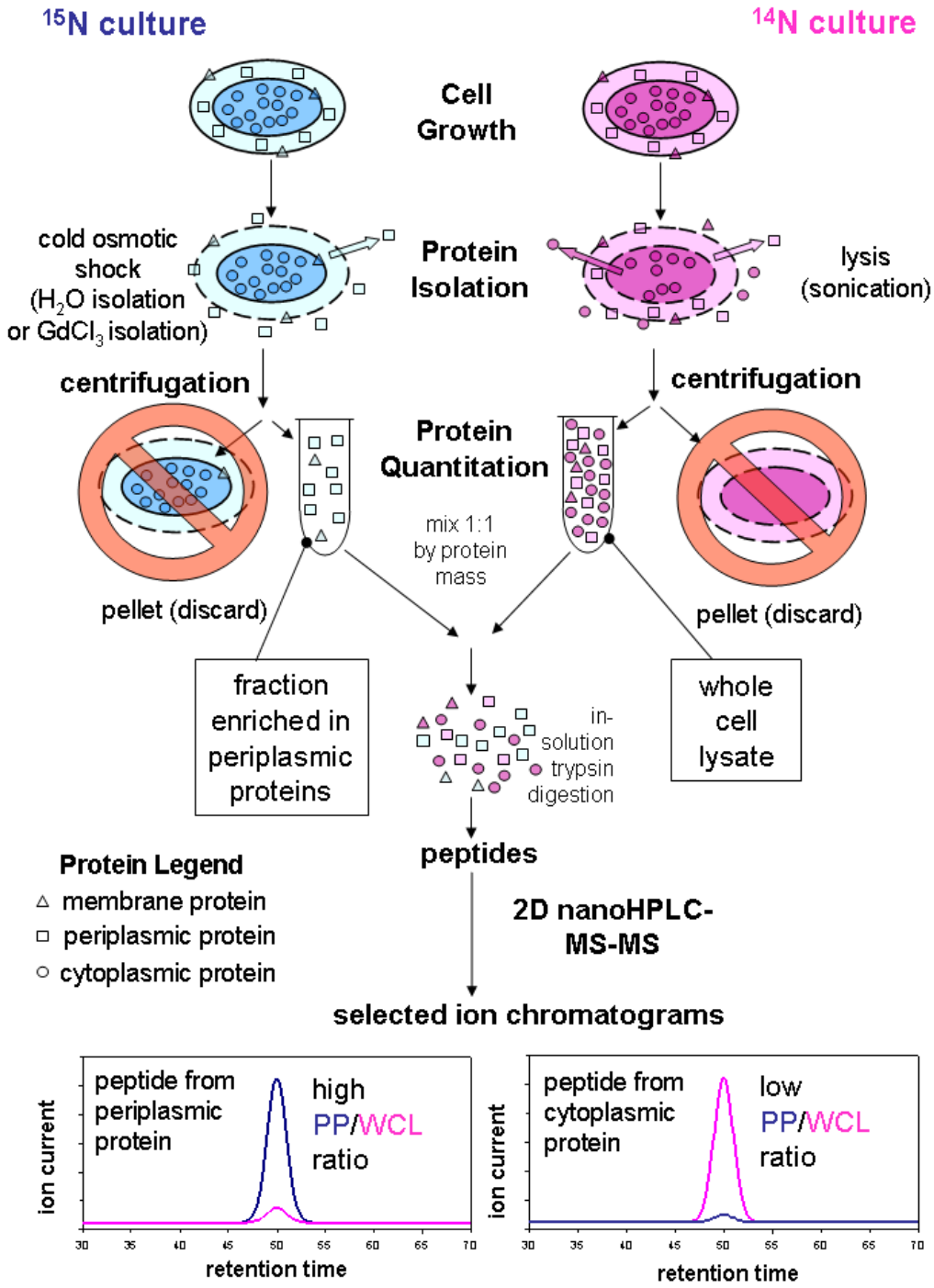



**Figure 5.2.** Experimental overview for measurement of PP:WCL protein isotope ratios.

isotope of nitrogen, $^{14}N$. Upon characterization of the mixture by LC-MS-MS, relative abundance measurement of isotopic signatures provides a means to distinguish peptides enriched from the periplasmic enrichment (PP) from peptides of the whole cell lysate (WCL) fraction. Expressed as the PP:WCL ratio for each protein, higher PP:WCL ratios potentially indicate enrichment in the periplasmic fraction, while lower values indicate greater abundance in the whole cell lysate. We evaluated two separate osmotic shock buffers to isolate periplasmic proteins ($H_2O$ [44] and $GdCl_3$ [45].

For proteins enriched by the $H_2O$ isolation protocol, 394 total PP:WCL protein isotope ratios were measured, 31 of which corresponded to annotated periplasmic proteins indicating enrichment. Similarly, 29 PP:WCL protein isotope ratios corresponded to annotated periplasmic proteins enriched via the $GdCl_3$ isolation protocol, but the total number of measured ratios, 722, differed.

Integration of PP:WCL ratio measurements with an informatics strategy to identify amino-terminal signal peptides resulted in 62 annotated periplasmic proteins containing this post-translational modification. Between the $H_2O$ or $GdCl_3$ protocols, 26 of these proteins were measured among the most enriched in the periplasmic isolation by PP:WCL isotope ratio and were identified bearing the post-translational modification of amino-terminal signal peptide cleavage. Collectively, the measurements performed in our quantitative approach provide increased confidence in studying protein localization to the *E. coli* K12 periplasm.

## *5.2 Experimental Procedures*

**Materials.** Reagents obtained from Sigma Chemical Company (St. Louis, MO)

included Trizma Hydrochloride (UltraPure), Trizma Base (UltraPure), $CH_3CO_2NH_4$ (99.999% purity), guanidine hydrochloride (molecular biology grade), sucrose (99+% molecular biology grade), $KH_2PO_4$ (≥99.0% purity), and disodium ethylenediamine-tetraacetic acid (EDTA) (99+% molecular biology grade). Other reagents included dithiothreitol (DTT) (OmniPur), glucose (USP purity), and formic acid (SupraPur), EMD Chemicals, Darmstadt, Germany); $^{15}NH_4Cl$ (98+ atom % excess $^{15}N$) (Isotec, Miamisburg, OH), $NH_4Cl$ (99.5% purity), $CaCl_2$ (99% purity), and NaCl (molecular biology grade) (Mallinckrodt Baker Chemicals, Philipsburg, NJ); lyophilized sequencing-grade modified trypsin (Promega, Madison, WI), HPLC-grade $CH_3CN$ and $H_2O$ (Burdick and Jackson, Muskegon, MI), $GdCl_3 \cdot 7H_2O$ (99.99% metals basis purity) (Alfa Aesar, Ward Hill, MA); $MgSO_4 \cdot 6H_2O$ (99.9% purity) and $Na_2HPO_4$ (USP) (FisherScientific, Fair Lawn, NJ). Ultrapure 18 MΩ water used for protein isolation buffers was obtained from a Milli-Q system (Millipore, Billerica, MA). HPLC-grade water was used in trypsin buffer (50 mM Trizma-HCl 10 mM $CaCl_2$, pH 7.6). HPLC-grade water, HPLC-grade acetonitrile, formic acid (SupraPur), and $CH_3CO_2NH_4$ were used in preparation of HPLC mobile phases.

**Bacterial Growth.** Identical *Escherichia coli* K12 MG1655 (ATCC number 47076, American Type Culture Collection, Rockville, MD) cultures were grown separately in $^{14}N$- and $^{15}N$-enriched M9 minimal medium [85] with ammonium chloride (1 g/L, w/v) as the sole nitrogen source [40]. Separate 5-mL cultures of *E. coli* K12 MG1655 were grown overnight in $^{15}N$-enriched M9 medium at 37°C for five consecutive days to ensure incorporation of $^{15}NH_4Cl$ into the amino acids of the *E. coli* proteome.

Separate (two $^{14}N$-enriched, two $^{15}N$-enriched) 5 mL liquid cultures were used to inoculate four 500-mL portions of M9 minimal medium in 2 L baffled Erlenmeyer flasks. Growth was conducted under aerobic conditions in a shaking incubator (New Brunswick Innova) at 37°C, with agitation at approximately 1,440 rpm. Cells were harvested in late logarithmic phase ($OD_{600}$ = 1.5).

**Periplasmic Protein Isolation.** Periplasmic protein isolation was performed by one of two cold osmotic shock protocols: one protocol used a hypotonic shock buffer of $H_2O$ [44] (designated as *$H_2O$ isolation*), while the second protocol used a hypotonic shock buffer of 1 mM gadolinium (III) chloride ($GdCl_3$) in HPLC-grade $H_2O$ [45] (designated as *$GdCl_3$ isolation*). Briefly, 20 mL-aliquots of $^{15}N$-enriched *E. coli* K12 cells were harvested by centrifugation at 5000 *g* at 4°C for 25 min, washed twice with 10 mL of 20 mM Trizma-HCl pH 7.6, and again centrifuged at 5000 *g* at 4°C for 25 min. The cell pellet was resuspended in 2.0 mL periplasting buffer (20 mM Trizma-HCl pH 7.6, 20% sucrose, 0.5 mM EDTA) and incubated on ice for 10 min. 2 mL of $H_2O$ or 1 mM $GdCl_3$ hypotonic shock buffer, pre-chilled to 4°C, was added to the cellular suspension, which was incubated on ice for an additional 10 min. The cellular suspension was then centrifuged at 12,000 *g* at 4°C for 15 min. The supernatant liquid was collected as periplasmic protein isolate; a 10 µL aliquot was removed for SDS-PAGE and the remainder was stored at -80°C.

**Whole Cell Lysate Isolation.** Soluble proteome fractions were isolated from cell cultures, grown in M9 minimal medium, as described above. Approximately 2 g of wet cellular paste was resuspended in 10 mL 20 mM Trizma-HCl (pH 7.6, chilled to 4°C),

lysed by five one-minute intervals of pulsating sonication (Branson Sonifier, Danbury, CT) on ice. Briefly, each one-minute sonication interval consisted of disruption of the cellular suspension with a microprobe set to high power for 30 sec. followed by cooling incubation period for 30 sec. on ice. This process was repeated five times to achieve lysis of cellular membranes. The cellular suspension was centrifuged at 12, 000 *g* to sediment cellular debris. A 10 µL aliquot was removed for SDS-PAGE. The remainder of the whole cell lysate (WCL) fraction was aliquotted into 1-mL portions and stored at -80°C until further use.

**SDS-PAGE.** 10-µL aliquots of protein fractions ($^{15}$N-labeled $H_2O$ periplasmic protein isolate, $^{14}$N-labeled soluble proteome fraction, and $^{15}$N-labeled $GdCl_3$ periplasmic protein isolate), along with a Precision Plus DualColor™ molecular weight marker (BioRad), were separated by 1D SDS-PAGE on a precast 4-20% Precise Protein Gel (Pierce) inserted into a Hoeffer Mighty Small Gel Apparatus. SDS-PAGE was conducted at a constant voltage (110-120 V). Following the SDS-PAGE protein separation, the gel was stained with SimplyBlue™ Coumassie stain (Invitrogen), and an image of the gel was acquired with a VersaDoc imaging system (BioRad).

**Protein Fraction Mixing.** Each periplasmic and cellular protein isolate was dialyzed against 18 MΩ ultrapure water for 16 h at 4°C using SnakeSkin Dialysis Tubing, (3500 MWCO, Pierce, Rockford, IL) and quantified by the microBCA assay (Pierce). Mixtures of $^{14}$N and $^{15}$N labeled proteins were created at a 1:1 ratio (w/w) by protein mass according to the results of the microBCA assay.

The protein mixtures were created as described here in more detail. For experimental

protein mixtures, relative abundance measurements of peptide isotopic signatures were expressed for each protein as the PP:WCL ratio (or $^{15}N$:$^{14}N$ ratio). Values of the PP:WCL ratio were used to estimate enrichment of periplasmic proteins in PP isolate. These experimental protein mixtures were created as follows:

a. $^{15}N$-labeled $H_2O$ PP isolate mixed with $^{14}N$-labeled WCL fraction; and,

b. $^{15}N$-labeled $GdCl_3$ PP isolate mixed with $^{14}N$-labeled WCL fraction.

To estimate errors associated with preparing mixtures of $^{15}N$- and $^{14}N$-labeled proteins, the following control mixtures were also created:

a. $^{15}N$-labeled $H_2O$ PP isolate mixed with $^{14}N$-labeled $H_2O$ PP isolate;and,

b. $^{15}N$-labeled $GdCl_3$ PP isolate mixed with $^{14}N$-labeled $GdCl_3$ PP isolate.

In control mixtures, relative abundance measurements of peptide isotopic signatures were expressed for each protein as PP:PP ratios (or $^{15}N$:$^{14}N$ ratios). Barring mixing errors, values of PP:PP ratios should equal 1 (or $\log_2=0$).

**Digestion by Enzymatic Proteolysis.** Guanidine hydrochloride was added to each protein mixture to a final concentration of 6M. Disulfide bonds were reduced with the addition of 10 mM DTT and incubation at 60°C for 1 h. Following incubation, guanidine hydrochloride was diluted to 1M by the addition of 50 mM Trizma-HCl 10 mM $CaCl_2$, pH 7.6. Sequencing-grade modified trypsin (Promega) was added at an enzyme-to-substrate ratio of 1:50 (w/w). Proteolysis was conducted on a platform mixer (VWR) inside a 37°C incubator (VWR) with gentle nutating for 18 h. Proteolytic peptides were subjected to reverse-phase $C_{18}$ solid phase extraction (SepPak Light, Waters, Milford, MA) and exchanged into a solvent of 100% $H_2O$, 0.1% formic acid by centrifugal

evaporation (SpeedVac, Savant Instruments, Holbrook, NY) to a final concentration of 10 μg/μL, on the basis of the amount of protein starting material. Proteolytic peptide samples were filtered with an Ultrafree-MC centrifugal filter device (Durapore PVDF 0.45 μm filter, Millipore, Bedford, MA) at 2000 *g* and stored at -80°C until LC-MS-MS analysis.

**Liquid Chromatography and Tandem Mass Spectrometry.** The resulting mixtures of proteolytic peptides were separated by a two dimensional (strong cation exchange and reverse phase) nanoHPLC system [41], configured for a split-phase 12-step MudPIT separation [26], interfaced via a nanoelectrospray source (Proxeon Biosystems, Odense, Denmark) with a linear ion trap mass spectrometer (ThermoFinnigan LTQ, San Jose, CA), under control of the XCalibur software (version 2.0.9, ThermoFinnigan). Replicate biological and technical (LC-MS-MS) analyses were performed as outlined in Figure 5.3.

The LTQ instrument scanned the mass-to-charge range of 400-1700 in data dependent mode and subjected the five most intense ions to MS-MS analysis at 35% normalized collisional energy. Dynamic exclusion [64] was enabled and conducted under the following settings: repeat count 1, repeat duration 60 sec., exclusion size list 300, exclusion duration 180 sec., exclusion width by mass (low and high) 1.5 Da, with expiration disabled. The default MS-MS charge state was set to 3. Full-scan mass spectra were averaged from four microscans; tandem mass spectra were averaged from two microscans.

**Protein FASTA file construction.** FASTA-formatted protein sequences of the predicted *E. coli* proteome were obtained from SwissProt (HAMAP) [88]. A list of

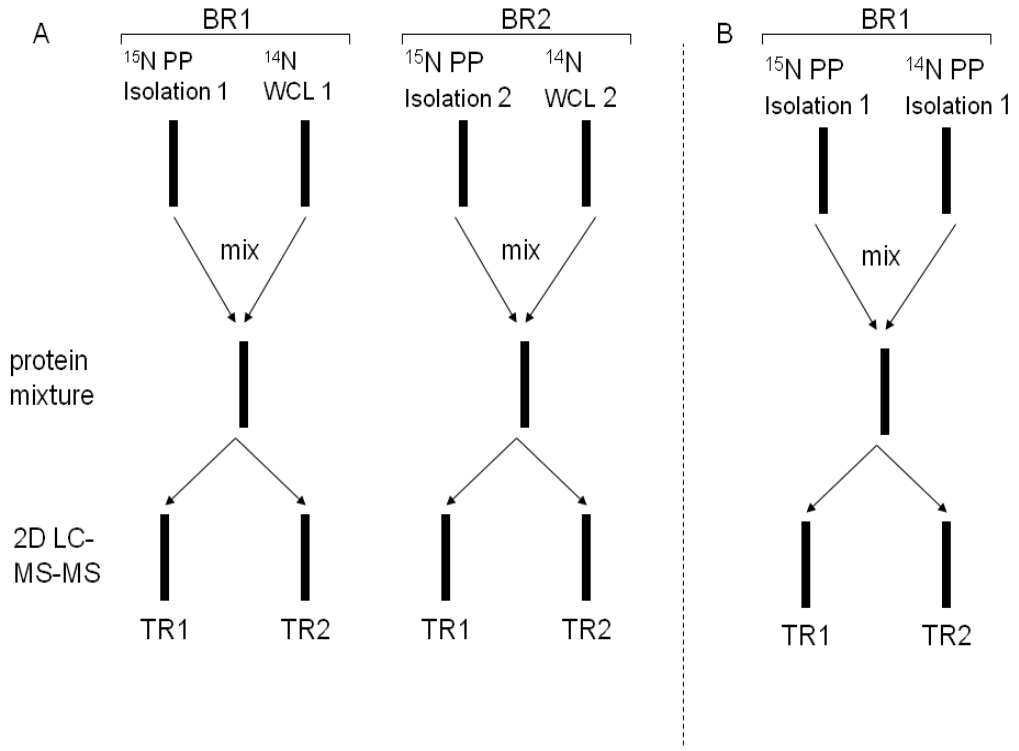



**Figure 5.3.** Replication Scheme of LC-MS-MS Analyses outlined in Figure 2. a. For measurement of PP:WCL isotope ratios, two pairs (1 $^{15}N$ labeled, 1 $^{14}N$ labeled) of cultures were grown and duplicate LC-MS-MS runs performed for each culture after protein mixing. b. For measurement of PP:PP isotope ratios, only a single culture pair (1 $^{15}N$ labeled, 1 $^{14}N$ labeled) was grown and duplicate LC-MS-MS runs performed after protein mixing. BR= biological replicate (i.e., culture and protein isolation), TR = technical replicate (i.e. LC-MS-MS run). Experiments shown above in a. and b. were performed for both the $H_2O$ and $GdCl_3$ isolation protocols.

potential periplasmic proteins was obtained by querying the predicted *E. coli* proteome with the SwissProt Sequence Retrieval System (SRS) [115]. The SRS Query "swiss_prot-Organism: escherichia coli & swiss_prot-Strain: K12 & swiss_prot-Keywords: periplasm*" returned 168 proteins. Nine of the 168 proteins returned from the SRS query contained both transmembrane domains and peripheral periplasmic domains and were thus removed from the targeted list, resulting in a total of 159 potential periplasmic proteins. Annotations for 98 of the 159 potential periplasmic proteins specified an amino-terminal signal peptide cleavage position previously described by experimental methods. For each of these 98 proteins, two additional entries were added to the FASTA sequence file of predicted *E. coli* proteins: one short entry consisting of the amino- terminal signal peptide (approximately 20-30 amino acids) and a second longer entry consisting of the remaining amino acids of the protein. Creation of these two additional entries for each periplasmic protein in the FASTA sequence file permitted identification of the amino-terminal signal peptide cleavage position as a tryptic peptide in our informatics workflow (detailed below). Protein sequences of the remaining 61 potential periplasmic proteins were submitted to both the PrediSi [116] and SignalP [117] signal peptide prediction algorithms. Additional entries to the FASTA sequence file were made for potential periplasmic proteins reflecting the length of the predicted amino-terminal signal peptide and cleavage site according to the following prediction of algorithm scores: PrediSi score $\geq$0.5; SignalP NN D-score $\geq$0.5; and SignalP HMM S-prob $\geq$0.5. An additional 19 FASTA-formatted sequences, representing common contaminant proteins (keratins, in-house standard proteins, and proteolytic enzymes)

were also included. The 4,337 proteins of the *E. coli* predicted proteome, obtained from SwissProt's HAMAP, were reversed and concatenated to the FASTA-formatted protein sequence file by Perl scripting to permit estimation of the false positive identification rate [90]. The final database contained a total of 9,121 FASTA-formatted protein sequences: 4,337 proteins of the SwissProt's HAMAP Proteome, 4,337 "reversed" (or decoy) proteins, 424 entries representing periplasmic proteins, some with processed amino-termini, 2 annotated periplasmic proteins that do not contain a signal peptide, and 19 contaminant proteins.

**Peptide Identification, Protein Quantification, and Data Analysis.** Identification of peptides from tandem mass spectra was performed using the SEQUEST algorithm [10] (version 27) with tryptic enzyme specificity (enzyme number = 1)

Two separate SEQUEST searches were performed, with amino acid molecular masses calculated based on $^{14}N$ for the first, and $^{15}N$ for the second search [41]. Peptide identifications were filtered, organized, and assembled by locus using the DTASelect program [27] (version 2.0.6). DTASelect retained identifications based on singly-, doubly-, and triply-charged peptide ions with $X_{corr}$ scores $\geq$1.8, $\geq$2.5, and $\geq$3.5, respectively, and $\Delta$CN scores $\geq$0.08. Identification at the protein level required identification of two or more non-redundant tryptic peptides. Multiple MS-MS scans for identified peptide sequences were retained by specifying the DTASelect locus filter *–t 0*. DTASelect discarded redundant locus information by default; contaminant proteins (keratins, proteolytic enzymes) were removed using the filter *–l contaminant*. The DTASelect-estimated forward false positive rates were $\leq$0.15% at the peptide level and

≤0.05% at the protein level. The predicted *E. coli* proteome, amino-terminal signal peptide cleavage sites, and predicted amino-terminal signal peptide cleavage sites were sorted into separate protein classes using DTASelect (classifications.txt). Separate classes were defined for protein sequences from the HAMAP proteome and protein sequences containing predicted signal peptide cleavages. The DTASelect classifications function (*--class*) allowed the program to differentiate between these two classes of protein identifications.

Results from these $^{14}N$ and $^{15}N$ SEQUEST searches were merged using the "Merge Directories" tool of the ProRata program [41] GUI (version 1.1). The ProRata program was used with default specifications to extract selected ion chromatograms, estimate peptide abundance ratios, and assemble peptide quantifications into protein quantifications, with the additional requirement of ≥3 peptide quantifications per protein [41]. Biological (cultures/isolations) and technical (LC-MS-MS) replicate data were further processed by the ProRata program to estimate a single $\log_2$ PP:WCL (or $^{15}N$:$^{14}N$) isotope ratio and 90% confidence interval for each protein meeting these criteria.

## *5.3 Results and Discussion*

Knowledge of the spatial distribution of the proteome is highly desirable in elucidating protein function. The dynamic nature of the proteome—simultaneously undergoing synthesis, transport, modification, and degradation of proteins—often complicates the obtaining this valuable information. Protein localization is routinely studied by biochemical separation methods and/or molecular tagging techniques in a "one-by-one" fashion, focusing on examination of a single protein target per study.

A growing number of studies have applied quantitative proteomic techniques to study the localization of large numbers of proteins simultaneously. The combination of molecular tagging via isotopic signatures, biochemical separation of a "target" protein fraction, and measurement of signatures on robust LC-MS-MS platforms greatly increases experimental throughput. Examining a targeted subset of the proteome by integrating these tools may provide insight not attainable through the study of individual proteins. While such targeted, quantitative proteome analyses focusing on protein localization have been performed in eukaryotic organisms [32, 107, 109, 118, 119], prokaryotic organisms also contain distinct cellular structures, such as inner and outer membranes (shown in Figure 1) and multiple mechanisms of protein trafficking, such as the Sec and TAT pathways.

In this study, we wished to apply quantitative proteomics techniques to characterize the spatial distribution of microbial proteins isolated from the periplasm. To this end, we have devised an experimental approach (shown in Figure 5.2) combining stable isotope metabolic labeling, protein fractionation, and isotope ratio mass spectrometry to examine localization of *Escherichia coli* K12 periplasmic proteins. We enriched periplasmic proteins from $^{15}$N-labeled *E. coli* cell cultures by two distinct isolation protocols ($H_2O$ [44] and $GdCl_3$ [45]) and created protein mixtures containing periplasmic protein isolates and a $^{14}$N-labeled whole cell lysate fraction, as shown in Figure 5.3a. LC-MS-MS analysis of tryptic peptides from protein mixtures allowed measurement of PP:WCL protein isotope ratios. The PP:WCL ratio of each protein provides an estimate of periplasmic enrichment during the cold-osmotic shock isolation

protocol. Periplasmic proteins should have larger PP:WCL ratios than non-periplasmic proteins present in the whole cell lysate fraction. Thus, proteins measured with larger PP:WCL ratios indicate enrichment and potential periplasmic localization.

To account for experimental error attributable to protein mixing, a series of control experiments were also performed, composed of $^{15}N$-labeled periplasmic enrichments mixed with $^{14}N$ periplasmic enrichments, shown in Figure 5.3b. Separate $^{15}N$:$^{14}N$ mixtures were created for the $H_2O$ [44] and $GdCl_3$ [45] isolation protocols. LC-MS-MS analysis of this mixture permitted estimation of PP:PP protein isotope ratios for proteins present in the mixture. Barring experimental error from protein mixing, PP:PP isotope ratios should equal 1 ($\log_2=0$) for each protein in these mixtures. Control PP:PP protein isotope ratios of approximately 1 indicate enrichment similar protein abundances between separate $^{15}N$- and $^{14}N$-labeled isolate.

Prior to protein mixing or further downstream sample processing, each of the protein fractions were resolved by SDS-PAGE, shown in Figure 5.4. As expected, $H_2O$ and $GdCl_3$ periplasmic isolates are simpler mixtures than whole cell lysate. Figure 4 also shows that periplasmic isolates from both $H_2O$ and $GdCl_3$ protocols yields similar patterns, suggesting that a similar set of proteins was isolated by the two different cold osmotic shock protocols.

A previous qualitative characterization of periplasmic proteins enriched by cold osmotic shock reported a number of non-periplasmic proteins present among periplasmic proteins [110]. In this previous study, periplasmic localization was inferred by database presence of a Sec-dependent amino-terminal signal peptide for each protein. The TAT

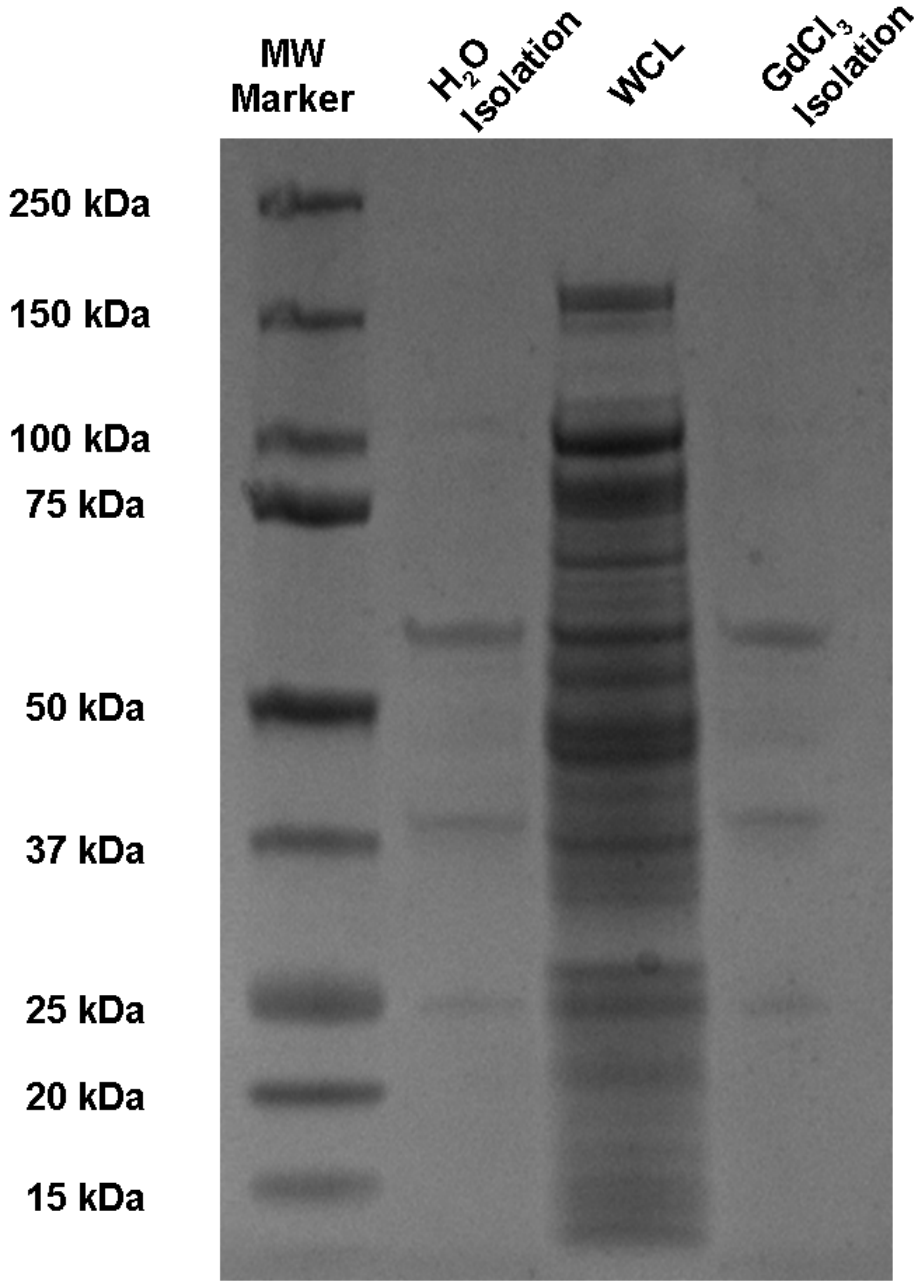



**Figure 5.4.** SDS-PAGE of protein fractions used in this study. Lane descriptions (L-R) are as follows: lane 1, molecular weight marker; lane 2, $H_2O$ periplasmic protein isolate; lane 3, whole cell lysate (WCL); lane 4, $GdCl_3$ periplasmic protein isolate. Each lane represents a 10 µL aliquot of protein sample prior to dialysis, microBCA assay, protein mixing, and further downstream analysis. See experimental procedures for further details.

pathway has revealed a distinct amino-terminal signaling motif for periplasimc localization and shown that co-exported proteins localized to the periplasm may be absent of an amino-terminal motif entirely [114].

Thus, presence of a Sec-dependent motif, as annotated in a database, may not sufficiently implicate all proteins localizing to the periplasm. To complement the series of isotope ratio measurements performed in our study, we also sought to identify tryptic peptides of mature amino-termini resulting from signal peptide cleavage rather than infer modification from database query of primary protein sequence after qualitative identification. A subset database of periplasmic proteins reflecting previously identified signal peptide cleavage site and predicted cleavage sites was appended to the proteome database and included in the search of mass spectrometry data.

**Differentially-Labeled Mixtures of Periplasmic Isolate and Cellular Lysates.**

*$^{15}N$ $H_2O$ Periplasmic Isolate mixed with $^{14}N$ Cellular Lysate.* Fifty-nine of the 394 protein isotope ratios measured by the ProRata program corresponded to proteins annotated as periplasmic proteins. Ratios were measured for proteins identified in $\geq$1 of 4 LC-MS-MS experiments from an average of 819 total protein identifications, with a standard deviation of $\pm$205 identifications, representing averages and standard deviations of 39$\pm$10 annotated periplasmic protein identifications and 780$\pm$215 non-periplasmic protein identifications. The distribution of 394 protein isotope ratios is shown in Figure 5.5, ranked in decreasing order of PP:WCL (or $^{15}N$:$^{14}N$) ratio. Rather than define strict, rigorous statistical cutoffs between "periplasmic" and "non-periplasmic" protein annotations plots showing the cumulative number of identified proteins that are annotated

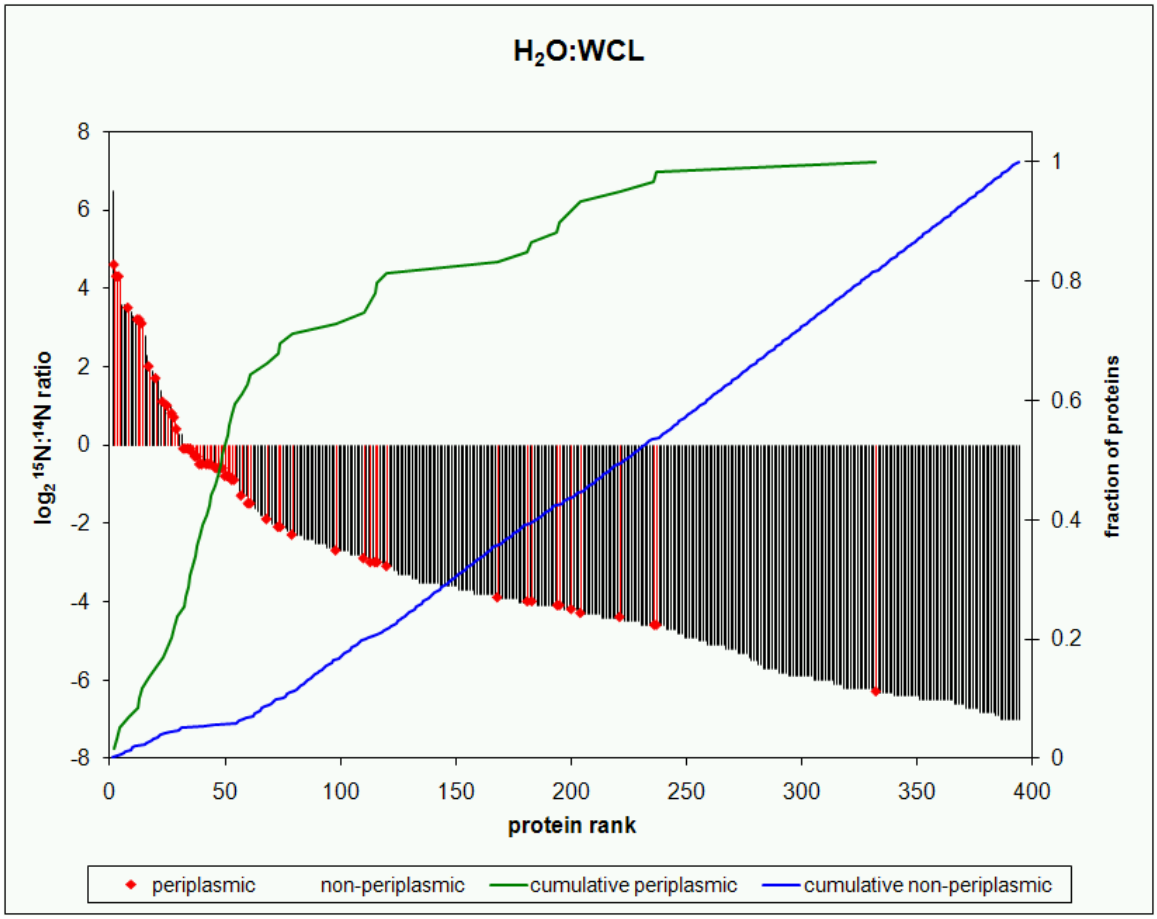



**Figure 5.5:** Distribution of protein isotope ratios from mixtures of $H_2O$ $^{15}N$-periplasmic isolate (PP) and $^{14}N$-whole cell lysate (WCL). PP:WCL isotope ratios are plotted along the y-axis, with proteins ranked according to this value along the x-axis. Periplasmic and non-periplasmic protein annotations were obtained from the SwissProt SRS query (see Experimental Procedures for more details) and are represented as detailed in the legend. Lines representing the cumulative periplasmic and cumulative non-periplasmic proteins as a fraction of the measured isotope ratios are also provided as described in the legend.

as periplasmic (green), or not annotated as periplasmic (blue) are provided in Figure 5.5. The cumulative number of identified periplasmic proteins initially increased at a greater rate than that of the non-periplasmic proteins. At lower values of the PP:WCL isotope ratio, the slope of the cumulative periplasmic distribution is much smaller (approaching 0), indicating that few annotated periplasmic proteins were measured with low PP:WCL isotope ratio values. A complete list of all measured PP:WCL protein ratios is provided in Appendix D.

As expected, several annotated periplasmic proteins were detected among the top 50 ranked PP:WCL isotope ratios over which the slope of the cumulative periplasmic rank is the largest in the distribution. Observed PP:WCL isotope ratios for periplasmic substrate-binding proteins for glutamine [120] (GlnH, 4.6), spermidine/putrescine [120] (PotD, 4.3), glycine/betadine [121] (ProX, 4.3), and thiamine [122] (ThiB, 3.5) are consistent with previous studies reporting periplasmic localization. Each of these proteins have been annotated and experimentally characterized as the substrate-binding component of ATP-binding cassette (ABC) transport systems, which shuttle their respective substrates from the periplasm to an inner membrane-spanning permease, facilitating import into the cytoplasm [123]. Twelve other periplasmic proteins annotated as periplasmic substrate-binding proteins in ABC transport systems were detected among the 50 largest-ranked PP:WCL isotope ratios.

Ratios for periplasmic proteins performing other functions were also detected in this portion of the distribution, such as protease 3 (PrtA, 3.1), the tail-specific protease (Prc, 2), and the blue copper oxidase (CueO, -0.5). The CueO protein has been described

as a member of the TAT pathway [114, 124], which employs a signaling motif distinct from the general secretory (Sec) pathway, shown in Figure 1.

In addition to the above annotated periplasmic proteins, some non-periplasmic proteins were observed with large PP:WCL isotope ratios among the top 50 ranked proteins. Detection of non-periplasmic proteins among enriched periplasmic proteins was expected, as previously reported [110]. Here, we observed large PP:WCL ratios for outer membrane receptors CirA (6.3), FepA (3.6), Fiu (3.6), and flagellin (3.2). A plausible explanation for these specific observed ratios is that the $H_2O$ periplasmic isolation protocol weakens the outer membrane, potentially leading to co-enrichment of these proteins along with periplasmic proteins.

A small fraction of annotated periplasmic proteins was detected with low PP:WCL isotope ratios, such as the osmotically-inducible protein Y (OsmY, -1.5), maltose binding protein (MalE, -1.5), and the vitamin B12 transport protein (BtuE, -4.5), among others. The observed low PP:WCL ratio for the OsmY protein may be attributable to induction of expression during osmotic stress [125], potentially in response to the periplasmic protein isolation protocol and/or the sonication procedure used for obtaining the whole cell lysate.

The ratio observed for BtuE is located along the distribution where the slope of the cumulative periplasmic rank is zero. Interestingly, the BtuE protein has been reported as non-essential for vitamin B12 transport [126] and another protein, BtuF has been implicated as the periplasmic substrate-binding protein for vitamin B12 [127]. These previous findings, combined with our quantitative PP:WCL ratio measurement merit

further investigation of the periplasmic annotation of the BtuE protein.

*$^{15}N$ $GdCl_3$ Periplasmic Isolate mixed with $^{14}N$ Cellular Lysate.* In an identical measurement performed using a different periplasmic isolation protocol, seventy-eight of the 722 protein isotope ratios measured by the ProRata program were proteins annotated as periplasmic proteins. Ratios were measured for proteins identified in $\geq$1 of 4 LC-MS-MS experiments from an average of 546$\pm$44 protein identifications (49$\pm$4 periplasmic proteins, 597$\pm$43 non-periplasmic proteins). The distribution of 722 protein isotope ratios is shown in Figure 6.6. A complete list of all measured PP:WCL protein ratios is provided in Appendix E.

As described above for Figure 5.5, Figure 5.6 also shows the cumulative number of identified periplasmic proteins increased at a greater rate than that of the non-periplasmic proteins, showing that larger $^{15}N$:$^{14}N$ ratios are more likely to be measured for periplasmic proteins. For the $GdCl_3$ periplasmic isolates, this is most evident for the 50 highest-ranked PP:WCL isotope ratios, 29 of which were annotated as periplasmic proteins.

As with the $H_2O$ isolation protocol, a number of proteins detected in this portion of the distribution were also ABC transporter periplasmic substrate-binding proteins, namely those binding thiamine (ThiB, 1.3), glycine/betadine (ProX, -0.3), and spermidine /putrescine (PotD, -1.1). Additional periplasmic binding proteins were detected with similar PP:WCL isotope ratios, such as TolB (0.7) and YobA (0.3), which are involved in the uptake of colicins and copper [88] based on their annotation, respectively. The YcdO protein, belonging to the uncharacterized protein family

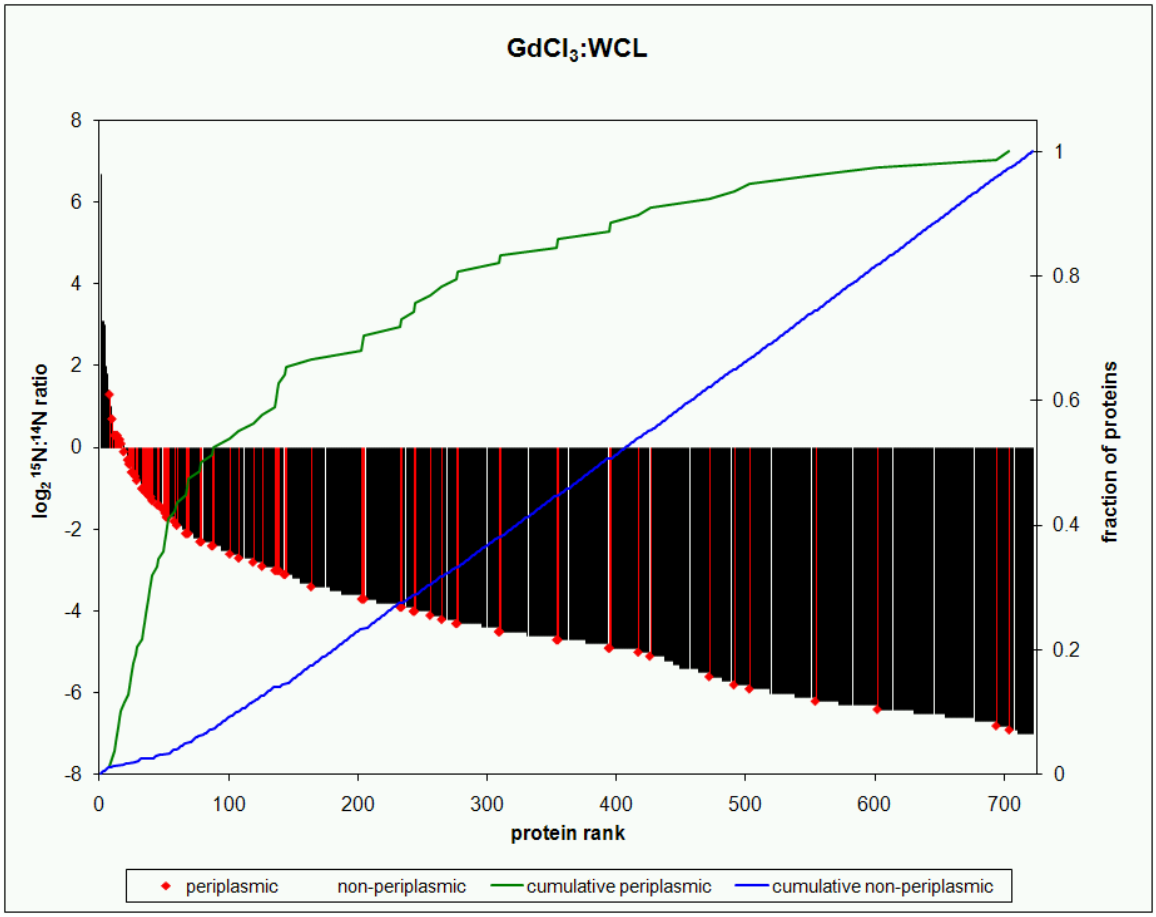



**Figure 5.6:** Distribution of protein isotope ratios from mixtures of $GdCl_3$ $^{15}N$-periplasmic isolate (PP) and $^{14}N$-whole cell lysate. PP:WCL isotope ratios are plotted along the y-axis, with proteins ranked according to this value along the x-axis. Periplasmic and non-periplasmic protein annotations were obtained from the SwissProt SRS query (see Experimental Procedures for more details) and are represented as detailed in the legend. Lines representing the cumulative periplasmic and cumulative non-periplasmic proteins as a fraction of the measured isotope ratios are also provided as described in the legend.

UPF0409 [88], was measured with a PP:WCL isotope ratio of 0.3. Initially suspected to be a TAT substrate protein, export of the YcdO protein to the periplasm has been demonstrated in TAT-deficient strains [128]. Though our measured isotope ratio further implicates periplasmic localization of YcdO it does not clearly pinpoint a function of this unknown protein.

Several non-periplasmic proteins were also observed with large PP:WCL isotope ratios among the top 50 ranked proteins. The largest 7 PP:WCL isotope ratios were obtained for proteins not annotated as periplasmic. As with the $H_2O$ isolation protocol, a number of these proteins were outer membrane-affiliated proteins, such as flagellin (FliC, 3.1), catecholate siderophore receptor (Fiu, 3.1), and outer membrane protein A (OmpA, 1.0), further supporting the notion that outer membrane proteins co-enrich with periplasmic proteins due to compromised membrane rigidity. Interestingly, an uncharacterized protein, YcgK, was observed with a PP:WCL isotope ratio of 3.0, the fourth largest-ranked ratio of all proteins. Though not annotated as a periplasmic protein, sequence analysis of this 14.9 kDa unknown protein contains a potential signal peptide at its amino terminus inferred by electronic annotation [88]. Based upon our measurement of PP:WCL protein isotope ratio, the YcgK protein represents a candidate hypothetical protein whose annotation could be investigated further to implicate periplasmic function.

Closer inspection of Figure 5.6 shows that the values of the 50 highest-ranked PP:WCL protein isotope ratios of $GdCl_3$ isolates are lower than their counterparts from $H_2O$ isolates. This interesting result merits further discussion. In comparison to other methods of periplasmic protein enrichment, it has been hypothesized that the $GdCl_3$

isolation protocol is more stringent in liberating periplasmic proteins versus cytoplasmic proteins due to the presence of gadolinium ion in the osmotic shock buffer, which blocks inner membrane mechanosensitive channel protein MscL [45, 129]. Blockage of the MscL protein would then, in turn, prevent this inner membrane channel from functioning as a conduit for non-specific release of proteins during osmotic downshock. Measurements of the PP:WCL isotope ratios presented here shown that comparable numbers of periplasmic proteins are of the 50-largest PP:WCL ratios for each protocol: 31 ratios by the $H_2O$ protocol and 29 ratios by the $GdCl_3$ protocol. However, numbers of total periplasmic proteins enriched by each protocol differed somewhat, with 59 by the $H_2O$ protocol and 78 by the $GdCl_3$ protocol. Differences among the enriched periplasmic proteins between protocols could be related to the total number of measured PP:WCL protein isotope ratios between protocols, for almost twice as many protein isotope ratios were measured for the $GdCl_3$ isolation protocol (722) than for the $H_2O$ protocol (394).

A plausible explanation for differences between protocols is that the total protein amount differs for each protein mixture of periplasmic isolate and whole cell lysate. Each time enrichment by cold osmotic shock is performed it may yield different amounts of periplasmic isolate. If periplasmic isolate liberates more proteins, it will generally yield smaller PP:WCL protein isotope ratios, for any particular protein will be present in smaller amount in the periplasmic isolate, but will be at the same amount in the whole cell lysate. This is evident between the $H_2O$ and $GdCl_3$ protocols. The $H_2O$ protocol yielded fewer total proteins, yet the PP:WCL protein isotope ratios are larger than the $GdCl_3$ protocol, which yielded more proteins. Consequently, PP:WCL protein isotope

ratios observed for the $GdCl_3$ protocol were smaller than the $H_2O$ protocol.

**Differentially-Labeled Mixtures of Periplasmic Isolate.** Accurate measurement of $^{15}N$:$^{14}N$ protein isotope ratios is critically dependent upon accurate protein mixing in our experimental design, depicted in Figure 5.2. To monitor experimental error in this step, we created control mixtures of isotopologous periplasmic isolate as shown in Figure 5.3b for both $GdCl_3$ and $H_2O$ isolation protocols. Barring experimental error in the protein mixing step, $^{15}N$:$^{14}N$ protein isotope ratios, denoted as PP:PP ratios, should be roughly equivalent between the two isotopologous periplasmic isolates in the mixture, resulting in PP:PP ratios of 1 (or 0 in $\log_2$). Distributions of protein quantifications from differentially-Labeled periplasmic enrichment mixtures are shown in Figures 5.7 and 5.8. Complete lists of all measured PP:PP protein ratios are provided in Appendices F and G for the $H_2O$ and $GdCl_3$ isolation protocols, respectively.

*$^{14}N$ $H_2O$ Periplasmic Isolate mixed with $^{15}N$ $H_2O$ Periplasmic Isolate*. The distribution of PP:PP protein isotope ratios in Figure 5.7 shows the majority of measured ratios did not differ drastically between mixtures of separate $^{15}N$ and $^{14}N$ periplasmic enrichments by the $H_2O$ isolation protocol. Nearly half of the PP:PP protein isotope ratios were measured at $\log_2$=0 (or 1), indicating little change in protein abundance between separate $^{15}N$ and $^{14}N$ periplasmic enrichments. The 296 ratios distributed over this range accounted for 46 annotated periplasmic proteins and 223 annotated non-periplasmic proteins. In fact, 85% of the PP:PP protein isotope ratios ranged from $\log_2$ values of -1 to 1 (representing 0.5 to 2-fold change) between the isotopologous enrichments, which indicate protein mixing in the experimental design shown in Figure

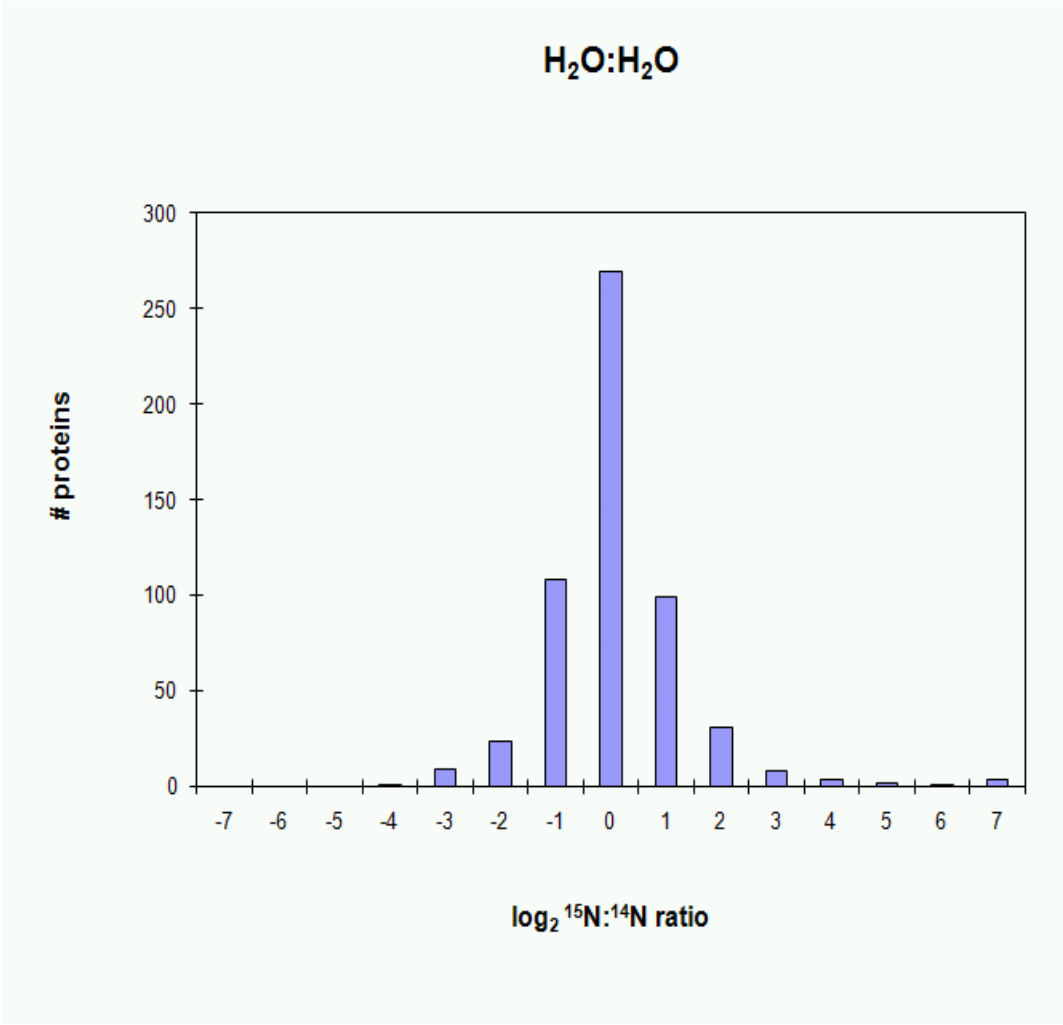



**Figure 5.7:** Distribution of Protein Isotope Ratios from Differentially-Labeled Periplasmic Isolate by the $H_2O$ Isolation Protocol.

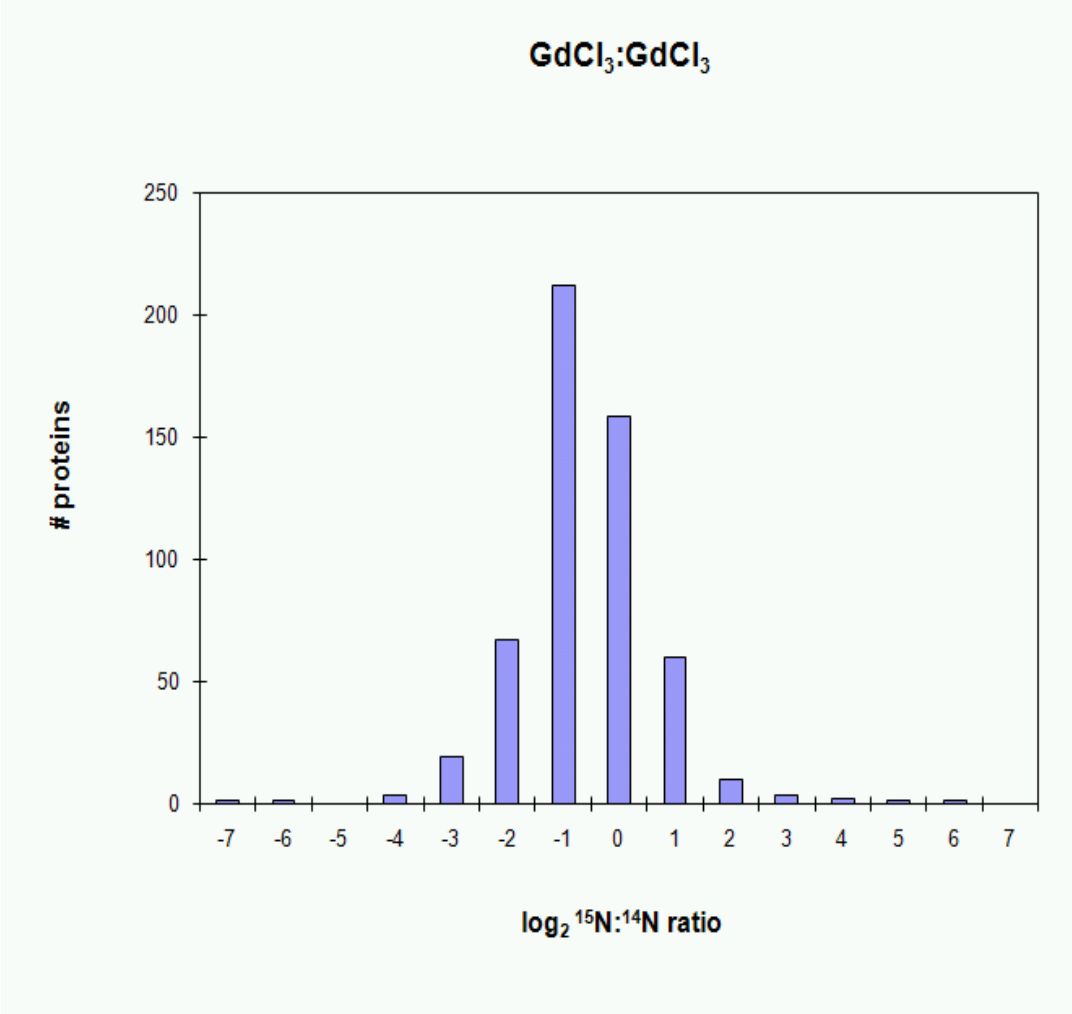



**Figure 5.8:** Distribution of Protein Isotope Ratios from Differentially-Labeled Periplasmic Isolate by the $GdCl_3$ Isolation Protocol.

5.2, is approximately a 1:1 ratio and thus, not a large source of experimental error.

*$^{14}N$ $GdCl_3$ Periplasmic Isolate mixed with $^{15}N$ $GdCl_3$ Periplasmic Isolate*. Though not as striking as with the $H_2O$ isolation protocol, 80% of the 538 measured PP:PP protein isotope ratios for the $GdCl_3$ isolation protocol also ranged from $\log_2$ values of -1 to 1 (representing 0.5 to 2-fold change) between the isotopologous periplasmic enrichments. The 430 total isotope ratios distributed over this range included 60 annotated periplasmic proteins and 370 non-periplasmic annotated proteins.

In contrast to the $H_2O$ isolation protocol, which similar abundances between $^{15}N$ and $^{14}N$ periplasmic enrichments, the highest frequency of proteins isolated by the $GdCl_3$ isolation protocol were distributed near $\log_2$=-1. This indicates a slight (0.5-fold) increase in abundance of the $^{14}N$ $GdCl_3$ isolation relative to the $^{15}N$ $GdCl_3$ isolation, potentially attributable to protein mixing error. Among 212 total PP:PP protein isotope ratios measured in this bin, 39 are annotated periplasmic proteins, potentially indicating more $^{14}N$ $GdCl_3$ protein isolation was included in the mixture relative to the $^{15}N$ $GdCl_3$ protein isolation.

Despite this slight difference among $GdCl_3$ isolates, distributions of PP:PP protein isotope ratios between protocols is still relatively comparable, with 85% of $H_2O$ ratios and 80% of $GdCl_3$ ratios ranging from $\log_2$ -1 to 1 (0.5 to 2-fold). Thus, errors attributable to protein mixing did not drastically skew measurement of $^{15}N$:$^{14}N$ isotope ratios in the experiment.

**Amino-Terminal Signal Peptide Cleavage Identification from Tandem Mass Spectra.** A characteristic feature of the majority of periplasmic proteins is the presence

of an amino-terminal signal peptide, approximately 20-30 amino acids in length, which translocates proteins to the periplasm via the general Sec pathway. Entry into the periplasm results in cleavage of this peptide from the remaining protein sequence. In addition to periplasmic enrichment, measurement of this post-translational modification provides higher confidence evidence in determining protein localization to the periplasm.

In a previous qualitative LC-MS-MS characterization of periplasmic proteins enriched by cold osmotic shock [110], presence of this feature among the primary sequence of identifications was determined by database querying and subsequently used to differentiate periplasmic proteins from other enriched proteins. Identification of the post-translational identification was determined by other experimental methods, such as Edman sequencing [120] , not by mass spectrometry.

To identify the post-translational modification of amino-terminal signal peptide cleavage among enriched periplasmic proteins, we concatenated a subset database of annotated periplasmic proteins to the HAMAP proteome prior to SEQUEST analysis of mass spectrometry data (Experimental Procedures). For each protein in the periplasmic subset database, one entry was created containing the full length of the amino-terminal signal peptide and a second entry was created containing the remaining sequence of the modified periplasmic protein. Cleavage positions for each periplasmic protein was based on previous experimental results listed among SwissProt protein annotations [88, 120], if available, or signal peptide cleavage predictions by the PrediSi [116] and SignalP [117] algorithms. This informatics approach permits identification of tryptic peptides for both cleaved and uncleaved forms of periplasmic proteins. Results are listed in Table 5.1. .

**Table 5.1.** Amino-terminal Signal Peptide Identifications from MS-MS Spectra among Annotated Periplasmic Proteins

| Protein | $H_2O$ PP:WCL Protein Mixtures | | | $GdCl_3$ PP:WCL Protein Mixtures | | | SwissProt Signal Peptide Prediction* (residues in mature chain) | Protein Description |
|---|---|---|---|---|---|---|---|---|
| | PP:WCL Isotope Ratio | replicates | N-term Cleavage Spectrum Count | PP:WCL Isotope Ratio | ∆eplicates | N-term Cleavage Spectrum Count | | |
| Prc | **2 (1.5, 2.6)** | 2 | 8 | | | 7 | Potential (23-682) | Tail-specific protease |
| ProX | **4.3 (4, 4.6)** | 4 | 137 | **-0.3 (-0.5, -0.1)** | 4 | 99 | Confirmed (22-682) | Glycine betaine-binding periplasmic protein |
| PotD | **4.3 (3.8, 4.8)** | 4 | 13 | **-1.1 (-1.3, -0.8)** | 4 | 4 | Confirmed (24-348) | Spermidine/putrescine periplasmic protein |
| PotF | **3.3 (2.7, 3.8)** | 3 | 2 | **-0.8 (-1.3, -0.4)** | 3 | 1 | Confirmed (27-370) | Putrescine-binding periplasmic protein |
| SubI | **1.7 (1.2, 6.9)** | 2 | 8 | **0.2 (-0.2, 0.7)** | 2 | 5 | Confirmed (20-329) | Sulfate-binding protein |
| GlnH | **4.6 (4.3, 5)** | 4 | 3 | -2.9 (-3.1, -2.6) | 4 | 1 | Confirmed (23-248) | Glutamine-binding periplasmic protein |
| LivJ | **0.7 (0.5, 0.8)** | 4 | 174 | **0.3 (0.1, 0.4)** | 4 | 116 | Confirmed (24-367) | Leu/Ile/Val-binding protein |
| FliY | **0.4 (0.1, 0.6)** | 4 | 21 | **0.2 (0.1, 0.4)** | 4 | 18 | Confirmed (30-266) | Cystine-binding periplasmic protein |
| ModA | **1.1 (0.9, 1.4)** | 4 | 145 | **-1.2 (-1.4, -1)** | 4 | 70 | Confirmed (25-257) | Molybdate-binding periplasmic protein |
| HisJ | **-0.1 (-0.3, 0.1)** | 4 | 40 | -0.1 (-0.3, 0.1) | 4 | 20 | Confirmed (23-260) | Histidine-binding periplasmic protein |
| LivK | **-0.1 (-0.3, 0.1)** | 4 | 92 | **-0.4 (-0.5, -0.2)** | 4 | 48 | Confirmed (24-369) | Leucine-specific-binding protein |
| CysP | **-0.1 (-0.3, 0.2)** | 4 | 15 | **-0.4 (-0.6, -0.2)** | 4 | 6 | Confirmed (26-338) | Thiosulfate-binding protein |
| ArtJ | **-0.2 (-0.6, 0.1)** | 4 | 59 | **-0.6 (-0.8, -0.4)** | 4 | 25 | Confirmed (20-243) | Arginine-binding periplasmic protein 2 |
| PpiA | **-0.6 (-1, -0.1)** | 3 | 0 | **-0.6 (-0.9, -0.2)** | 4 | 4 | Confirmed (25-190) | Peptidyl-prolyl cis-trans isomerase A |
| DppA | **-0.3 (-0.4, -0.1)** | 4 | 29 | **-1 (-1.2, -0.9)** | 4 | 41 | Confirmed (29-535) | Periplasmic dipeptide transport protein |
| SurA | **-0.3** | 3 | 27 | **-1** | 4 | 37 | Confirmed | Chaperone surA |

**Table 5.1.** (Continued)

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **(-0.6, 0)** | | | **(-1.3, -0.8)** | | | (21-428) | |
| ArgT | -0.9 (-1.2, -0.7) | 4 | 43 | **-0.8 (-1, -0.6)** | 4 | 35 | Confirmed (23-260) | Lysine-arginine-ornithine-binding periplasmic protein |
| LolA | **-0.6 (-0.9, -0.3)** | 4 | 3 | **-1.1 (-1.4, -0.8)** | 4 | 8 | Confirmed (22-203) | Outer-membrane lipoprotein carrier protein |
| PstS | -0.1 (-0.6, 0.4) | 2 | 7 | -1.6 (-2.1, -1.2) | 3 | 4 | Confirmed (26-346) | Phosphate-binding protein pstS |
| FimC | -0.8 (-1.7, -0.1) | 1 | 2 | **-1.2 (-1.8, -0.7)** | 2 | 0 | Confirmed (37-241) | Chaperone protein fimC |
| TesA | | | 2 | **-1.1 (-1.6, -0.7)** | 2 | 1 | Confirmed (27-208) | Acyl-CoA thioesterase I |
| ArtI | **-0.8 (-1.1, -0.4)** | 4 | 0 | **-1.4 (-1.7, -1.2)** | 4 | 3 | Confirmed (20-243) | Arginine-binding periplasmic protein 1 |
| CueO | **-0.5 (-1.2, 0.1)** | 1 | 5 | -2.1 (-2.4, -1.8) | 2 | 13 | Confirmed (29-516) | Blue copper oxidase cueO |
| OppA | -0.8 (-0.9, -0.7) | 4 | 200 | -1.9 (-2.1, -1.8) | 4 | 43 | Confirmed (27-543) | Periplasmic oligopeptide-binding protein |
| RbsB | **-0.5 (-0.9, 0)** | 4 | 4 | -2.4 (-2.7, -2) | 4 | 14 | Confirmed (26-296) | D-ribose-binding periplasmic protein |
| GsiB | | | 6 | -1.7 (-2, -1.4) | 3 | 5 | Potential (27-512) | Glutathione-binding protein gsiB |
| YnfD | -1.5 (-2.2, -0.9) | 1 | 11 | -2.3 (-2.9, -1.8) | 2 | 12 | Predicted (22-101) | Uncharacterized protein ynfD |
| OsmY | -1.3 (-1.5, -1.1) | 4 | 45 | -2.6 (-2.8, -2.4) | 4 | 30 | Confirmed (29-201) | Osmotically-inducible protein Y |
| OpgD | | | 2 | -2.1 (-2.4, -1.8) | 2 | 0 | Confirmed (33-551) | Glucans biosynthesis protein D |
| GltI | **-0.5 (-0.8, -0.3)** | 4 | 51 | -3.7 (-3.9, -3.4) | 4 | 22 | Confirmed (23-302) | Glutamate/aspartate periplasmic-binding protein |
| ZnuA | -1.5 (-1.7, -1.4) | 4 | 95 | -3.1 (-3.3, -3) | 4 | 41 | Confirmed (27-310) | High-affinity zinc uptake system protein znuA |
| DsbC | -1.9 (-2.4, -1.4) | 1 | 26 | -2.8 (-5.6, -1.3) | 3 | 13 | Confirmed (21-236) | Thiol:disulfide interchange protein dsbC |
| FkbA | **-0.6 (-0.9, -0.3)** | 4 | 61 | -4.5 (-4.8, -4.2) | 4 | 27 | Confirmed (26-270) | FKBP-type peptidyl-prolyl cis-trans isomerase fkpA |
| Skp | -2.7 (-3, -2.4) | 4 | 35 | -3 (-3.3, -2.8) | 3 | 43 | Confirmed (21-161) | Chaperone protein skp |
| SodC | -2.9 (-3.2, -2.5) | 3 | 1 | -3 (-3.3, -2.7) | 4 | 3 | Confirmed (20-173) | Superoxide dismutase [Cu-Zn] |
| DgaL | -3 (-3.3, -2.7) | 4 | 4 | -3 (-3.2, -2.8) | 4 | 4 | Confirmed (24-332) | D-galactose-binding periplasmic protein |

**Table 5.1.** (Continued)

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| UgpB | -3 (-3.4, -2.6) | 2 | 3 | -3 (-3.3, -2.7) | 4 | 9 | Confirmed (24-438) | sn-glycerol-3-phosphate-binding periplasmic protein ugpB |
| CpxP | | | 8 | -3.1 (-3.6, -2.7) | 2 | 8 | Potential (22-166) | Periplasmic protein cpxP |
| GlpQ | -2.3 (-2.8, -1.9) | 4 | 12 | -4.2 (-4.6, -3.9) | 4 | 7 | Confirmed (26-558) | Glycerophosphoryl diester phosphodiesterase |
| Ivy | -2.1 (-2.5, -1.7) | 4 | 18 | -4.5 (-4.8, -4.1) | 3 | 8 | Confirmed (29-157) | Inhibitor of vertebrate lysozyme |
| BglX | | | 13 | -3.7 (-4.2, -3.2) | 2 | 9 | Potential (21-756) | Periplasmic beta-glucosidase |
| YceI | | | 15 | -3.9 (-4.4, -3.3) | 1 | 10 | Confirmed (23-191) | Protein yceI |
| Aspg2 | -4 (-4.4, -3.5) | 2 | 1 | | | 1 | Confirmed (23-348) | L-asparaginase 2 |
| AraF | -4.4 (-5.2, -0.7) | 1 | 4 | -3.7 (-4.3, -3.1) | 1 | 2 | Confirmed (24-329) | L-arabinose-binding periplasmic protein |
| AgP | -3.1 (-7, -2.4) | 2 | 11 | -5.1 (-5.9, -4.3) | 2 | 2 | Confirmed (23-506) | Glucose-1-phosphatase |
| YnhG | -4.6 (-4.9, -4.3) | 3 | 23 | -3.9 (-6.1, -3.5) | 4 | 12 | Potential (24-334) | Uncharacterized protein ynhG |
| EcoT | | | 19 | -4.3 (-7, -3.5) | 1 | 8 | Confirmed (21-162) | Ecotin |
| TreA | -4.1 (-4.9, -3.3) | 1 | 11 | -4.7 (-5.2, -4.3) | 3 | 3 | Confirmed (31-565) | Periplasmic trehalase |
| DegP | -3.9 (-4.4, -3.3) | 3 | 16 | -4.9 (-5.4, -4.5) | 2 | 6 | Confirmed (27-474) | Protease do |
| YdcS | | | 1 | -4.7 (-5.1, -4.3) | 1 | 2 | Potential (23-381) | Putative ABC transporter periplasmic-binding protein ydcS |
| YgiW | -3 (-3.7, -2.5) | 2 | 33 | -6.4 (-6.9, -5.9) | 1 | 13 | Confirmed (21-130) | Protein ygiW |
| OpgG | -4.1 (-4.6, -3.5) | 1 | 3 | -5.9 (-6.5, -5.4) | 2 | 3 | Confirmed (23-511) | Glucans biosynthesis protein G |
| HdeA | -4.6 (-5, -4.2) | 2 | 0 | -5.6 (-5.9, -5) | 4 | 4 | Confirmed (22-110) | Protein hdeA |
| YtfQ | -6.3 (-6.6, -6.1) | 4 | 58 | -5.8 (-6.1, -5.6) | 3 | 37 | Confirmed (22-318) | ABC transporter periplasmic-binding protein ytfQ |
| Spy | | | 4 | -6.2 (-6.5, -5.9) | 2 | 10 | Confirmed (24-161) | Spheroplast protein Y |
| MppA | | | 5 | -6.9 (-7, -6.3) | 2 | 0 | Confirmed (23-537) | Periplasmic murein peptide-binding protein |
| YraP | | | 4 | | | 0 | Potential | Uncharacterized protein yraP |

**Table 5.1.** (Continued)

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | (24-191) | |
| DdpA | | | 2 | | | 3 | Potential (26-516) | Probable D,D-dipeptide-binding periplasmic protein ddpA |
| AmpC | | | 3 | | | 3 | Confirmed (20-377) | Beta-lactamase |
| Ggt | | | 4 | | | 8 | Confirmed (26-580) | Gamma-glutamyltranspeptidase |
| DegQ | | | 5 | | | 5 | Confirmed (28-455) | Protease degQ |
| SufI | | | 3 | | | 1 | Confirmed (28-470) | Protein sufI |

*"Potential" means that the protein sequence potentially contains an amino-terminal signal peptide based on SwissProt annotation. "Confirmed" means that an amino-terminal signal peptide has been confirmed by experimental means based on the SwissProt annotation. The SwissProt annotation for each protein is accessible by clicking the hyperlinked-protein name.

SEQUEST searches identified signal peptide cleavage positions for 62 annotated periplasmic proteins. None of these 62 proteins were identified in the uncleaved form, that is, with an unprocessed signal peptide still attached.

Elucidation of the TAT pathway [114] has shown consensus primary sequence motifs other than the Sec signal peptide lead to periplasmic localization. Database queries of primary sequence analysis for Sec-dependent amino-terminal signal peptide alone may not have inferred periplasmic localization of the TAT protein CueO; however, our experimental measurements of PP:WCL ratio and MS-MS spectra of the cleavage position have (Table 5.1).

The informatics approach provided marginal insight for uncharacterized proteins. Though a total of 23 MS-MS spectra corresponded to a SignalP-predicted signal peptide cleavage site for the uncharacterized YnfD protein, low PP:WCL ratios measured from both isolation protocols suggest that it is enriched at low abundance in the periplasm. Similarly, MS-MS spectra were identified consistent with potential signal peptide cleavages for the YcdS, YraP, and YnhG proteins. However, measurement of each of their respective PP:WCL isotope ratios does not indicate periplasmic enrichment, shown in Table 5.1.

## *5.4 Conclusions*

Knowledge of cellular localization is integral to fully elucidating protein function. Given the dynamic nature of the cell, this insight is difficult to attain on a large scale for multiple proteins. A complication to localization studies combining biochemical isolation of selected protein fractions with qualitative proteomics platforms is that frequently, protein isolates contain a heterogeneous mixture of the desired proteins of interest and co-isolating contaminant proteins.

In this study, we have shown that quantitative proteomics provides higher confidence experimental evidence for localization of *E. coli* periplasmic proteins. Quantitative PP:WCL

protein isotope ratios indicating enrichment were measured for both the $H_2O$ and $GdCl_3$ isolation protocols. The separate protocols yielded comparable numbers in terms of annotated periplasmic proteins: 31 proteins for the $H_2O$ isolation protocol and 29 for the $GdCl_3$ protocol. However, markedly different values were observed in the total number of measured PP:WCL isotope ratios, with 394 measured in the $H_2O$ protocol opposed to 722 in the $GdCl_3$ protocol.

When combined with experimental identifications of the post-translational modification of amino-terminal signal peptide cleavage, a total of 26 annotated periplasmic proteins bearing this modification were among the largest measured PP:WCL protein isotope ratios in either the $H_2O$ or $GdCl_3$ isolation protocols.

The quantitative proteomics approach presented in this study provides more experimental credibility towards inference of protein localization in comparison to qualitative approaches. Adaptation of this experimental design to future proteome studies could potentially increase the scale of protein localization to other protein fractions among the growing number of sequenced microbial species.

# Chapter 6
# Conclusions

The goal of this dissertation was to expand the latitude of biological questions that may be addressed with bottom-up proteomics. To expand this horizon, biochemical and quantitative proteome techniques were tailored together to address specific areas of proteome research. Blending standard biochemical techniques, such as sucrose-density fractionation and affinity isolation, with newer quantitative proteomics tools provides a more targeted focus to proteome study.

Biochemical enrichment methods permit study of specific proteins of interest. Depending on the specific application, quantitative measurements of protein isotope ratios may afford a degree of increased confidence to the measurement. In conjunction, the application of the described biochemical and quantitative tools enables a more direct approach to addressing specific biological questions than qualitative proteome inventory profiling. In general, characterizing targeted protein mixtures of simpler complexity enable greater protein coverage of selected proteins, for the mass spectrometer is not overwhelmed by large numbers of peptides per unit time, as in the case of more complex proteome mixtures.

Chapter 1 highlights five common biological applications of proteomics: proteome inventory profiling, comparative proteome profiling, protein interaction analyses, protein localization, and post-translational modification identification. In the past five years, proteome inventory profiling of microbial species by the bottom-up approach has become routine [6]. After acquiring a large inventory of proteins, a logical "next step" is to infer each of their roles in the proteome. This dissertation has adapted techniques to specifically address the four *other* biological applications of proteomics:

comparative proteome profiling (Chapter 4),

protein interaction analyses (Chapters 3 and 4; Appendices A and B),

protein localization (Chapter 5), and

post-translational modification identification (Chapter 5).

Prior to engaging each of these biological applications, it was necessary to optimize sample preparation methods for studying small quantities of protein samples. Study of the above four applications involve a variety of distinct protein samples that differ in complexity and quantity. Sample preparation methods were in place for the study of complex proteome samples [43]; however, multiprotein complexes presented were technically challenging with respect to enzymatic proteolysis. Excessive handling steps during sample preparation, such as dilution of chaotropic agents and solid phase extraction, introduce potential sources of peptide loss. It is important to minimize sources of loss, especially when working with small quantities of limited protein sample. Chapter 3 addresses this critical, yet often overlooked, area of the bottom-up proteomics workflow. As proteomics is very much an interdisciplinary field, it frequently involves collaborations between biologists, biochemists, and analytical chemists. More often than not, collaborative proteomics studies require several trial experiments before yielding successful results. Biochemists expend significant time and resources to enrich small protein quantities prior to transferring the sample to analytical chemists for LC-MS-MS characterization. A prerequisite for successful collaboration is both parties reaching compromise conditions spanning the entire workflow, from protein enrichment through protein identification. This includes detailed discussion of optimal buffering solutions, steps of protein enrichment, yield of protein enrichment, and the presence of surfactants or protease inhibitors in the final protein sample.

Thus, the work presented in Chapter 3 is of significance, especially with respect to collaborative proteomics studies. First, it demonstrates general efficacy of five enzymatic protocols for small protein quantities. In collaborative studies, having a variety of buffering conditions, chaotropic agents, proteases, and sample cleanup methods is an important resource. These different conditions may be adapted to suit a specific set of compromise conditions between biochemists and analytical chemists. More importantly for the purposes of this dissertation, the work showed trypsin digestions in a solvent of 80% $CH_3CN$ were most effective on enriched samples of microbial proteins, resulting in greater peptide identifications, protein identifications, and sequence coverage.

Interaction analyses combining affinity isolation of proteins with characterization by LC-MS-MS are promising methods to study protein-protein interactions. Technical challenges facing this methodology are identification of non-specific (or "background") interactions amid affinity isolates and possible collateral effects resulting from heterologous expression of affinity-tagged "bait" proteins. In Chapter 4, both of these challenges were addressed by performing local and global protein isotope ratio measurements of the bacterial RNA polymerase protein complex. Local isotope ratio measurements provided a method to distinguish authentic protein interactions from non-specific interactions. This method, adapted from the I-DIRT method described by Tackett *et al.* [82], successfully differentiated RNA polymerase complex components from non-specific interactions. Global isotope ratio measurements of bait-expressing versus wild-type strains yielded a comparative proteome profile. The comparative profile enabled assessment of collateral effects from heterologous protein expression. Interestingly, this showed that the combination of growth on minimal salt medium (required for

$^{15}$N-labeling) and production of the bait protein may increase the metabolic load in both *E. coli* and *R. palustris*.

In Chapter 5, protein localization to the *E. coli* periplasm was studied. In mixtures of cold osmotic shock periplasmic isolate and whole cell lysate, measurement of protein isotope ratios permitted estimation of periplasmic protein enrichment. This measurement, the PP:WCL protein isotope ratio, increases confidence in periplasmic localization relative to previous qualitative LC-MS-MS studies. To further infer localization to the periplasm, a number of amino-terminal signal peptide cleavages were experimentally identified. This specific post-translational modification was found to be present in the majority of periplasmic proteins studied in Chapter 5. Combination of the PP:WCL protein isotope ratio with experimental measurement of this modification increases the overall confidence in periplasmic localization.

In conclusion, tailoring together the number of biochemical and quantitative techniques has provided increased confidence for investigating four other common biological applications of proteomics outside of qualitative proteome profiling. Targeted studies that expand the scope of common applications of proteomics will provide useful biological insight into the ever-increasing number of sequenced microbial genomes. Hopefully, the combinations of biochemical and quantitative techniques used throughout this dissertation will serve as an adaptable framework for future studies by the bottom-up proteomics approach.

# List of References

1. Madigan, M.T., J.M. Martinko, and J. Parker, *Brock Biology of Microorganisms*. 9th ed. 2000, Upper Saddle River, NJ: Prentice Hall.
2. Alberts, B., et al., *Manipulating Proteins, DNA, and RNA*, in *Molecular Biology of the Cell*. 2002, Garland Science: New York, NY. p. 513-524.
3. Ideker, T., T. Galitski, and L. Hood, *A new approach to decoding life: Systems biology.* Annual Review of Genomics and Human Genetics, 2001. **2**: p. 343-372.
4. Ideker, T., *Systems biology 101 - what you need to know.* Nature Biotechnology, 2004. **22**(4): p. 473-475.
5. Yates, J.R., 3rd, *Mass spectrometry and the age of the proteome.* J Mass Spectrom, 1998. **33**(1): p. 1-19.
6. VerBerkmoes, N.C., et al., *Mass spectrometric approaches for characterizing bacterial proteomes.* Expert Review of Proteomics, 2004. **1**(4): p. 433-447.
7. Domon, B. and R. Aebersold, *Review - Mass spectrometry and protein analysis.* Science, 2006. **312**(5771): p. 212-217.
8. Fenn, J.B., et al., *Electrospray ionization for mass spectrometry of large biomolecules.* Science, 1989. **246**(4926): p. 64-71.
9. Eng, J., McCormack AL, and Yates, JR III., *An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database.* Journal Of The American Society For Mass Spectrometry, 1994. **5**: p. 976-989.
10. Yates, J.R., 3rd, et al., *Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database.* Anal Chem, 1995. **67**(8): p. 1426-36.
11. Ram, R.J., et al., *Community proteomics of a natural microbial biofilm.* Science, 2005. **308**(5730): p. 1915-1920.
12. VerBerkmoes, N.C., et al., *Evaluation of "Shotgun" proteomics for identification of biological threat agents in complex environmental matrixes: Experimental simulations.* Analytical Chemistry, 2005. **77**(3): p. 923-932.
13. Link, A.J., et al., *Direct analysis of protein complexes using mass spectrometry.* Nat Biotechnol, 1999. **17**(7): p. 676-82.
14. Gould, K.L., et al., *Tandem affinity purification and identification of protein complex components.* Methods, 2004. **33**(3): p. 239-44.
15. Butland, G., et al., *Interaction network containing conserved and essential protein complexes in Escherichia coli.* Nature, 2005. **433**(7025): p. 531-7.
16. Hillenkamp, F., et al., *Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers.* Anal Chem, 1991. **63**(24): p. 1193A-1203A.
17. Jonscher, K.R. and J.R. Yates, *The quadrupole ion trap mass spectrometer - A small solution to a big challenge.* Analytical Biochemistry, 1997. **244**(1): p. 1-15.
18. Schwartz, J.C., M.W. Senko, and J.E.P. Syka, *A two-dimensional quadrupole ion trap mass spectrometer.* Journal Of The American Society For Mass Spectrometry, 2002. **13**(6): p. 659-669.
19. Wells, J.M. and S.A. McLuckey, *Collision-induced dissociation (CID) of peptides and proteins*, in *Biological Mass Spectrometry*. 2005, Elsevier Academic Press Inc: San Diego. p. 148-185.
20. Hunt, D.F., et al., *Protein sequencing by tandem mass spectrometry.* Proc Natl Acad Sci U S A, 1986. **83**(17): p. 6233-7.

21. Steen, H. and M. Mann, *The ABC's (and XYZ's) of peptide sequencing.* Nature Reviews Molecular Cell Biology, 2004. **5**(9): p. 699-711.
22. Thompson, M.R., et al., *Dosage-dependent proteome response of Shewanella oneidensis MR-1 to acute chromate challenge.* Journal of Proteome Research, 2007. **6**(5): p. 1745-1757.
23. Blackler, A.R., et al., *Quantitative comparison of proteomic data quality between a 2D and 3D quadrupole ion trap.* Analytical Chemistry, 2006. **78**(4): p. 1337-1344.
24. Washburn, M.P., D. Wolters, and J.R. Yates, 3rd, *Large-scale analysis of the yeast proteome by multidimensional protein identification technology.* Nat Biotechnol, 2001. **19**(3): p. 242-7.
25. Wolters, D.A., M.P. Washburn, and J.R. Yates, 3rd, *An automated multidimensional protein identification technology for shotgun proteomics.* Anal Chem, 2001. **73**(23): p. 5683-90.
26. McDonald, W.H., et al., *Comparison of three directly coupled HPLC MS/MS strategies for identification of proteins from complex mixtures: single-dimension LC-MS/MS, 2-phase MudPIT, and 3-phase MudPIT.* International Journal of Mass Spectrometry, 2002. **219**(1): p. 245-251.
27. Tabb, D.L., W.H. McDonald, and J.R. Yates, *DTASelect and contrast: Tools for assembling and comparing protein identifications from shotgun proteomics.* Journal of Proteome Research, 2002. **1**(1): p. 21-26.
28. McDonald, W.H., et al., *MS1, MS2, and SQT-three unified, compact, and easily parsed file formats for the storage of shotgun proteomic spectra and identifications.* Rapid Commun Mass Spectrom, 2004. **18**(18): p. 2162-8.
29. VerBerkmoes, N.C., et al., *SYSTEMS BIOLOGY Functional analysis of natural microbial consortia using community proteomics.* Nature Reviews Microbiology, 2009. **7**(3): p. 196-205.
30. Gingras, A.C., et al., *Analysis of protein complexes using mass spectrometry.* Nature Reviews Molecular Cell Biology, 2007. **8**(8): p. 645-654.
31. Pflieger, D., et al., *Quantitative proteomic analysis of protein complexes.* Molecular & Cellular Proteomics, 2008. **7**(2): p. 326-346.
32. Andersen, J.S. and M. Mann, *Organellar proteomics: turning inventories into insights.* Embo Reports, 2006. **7**(9): p. 874-879.
33. Sadowski, P.G., et al., *Quantitative proteomic approach to study subcellular localization of membrane proteins.* Nature Protocols, 2006. **1**(4): p. 1778-1789.
34. Gerber, S.A., et al., *Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS.* Proc Natl Acad Sci U S A, 2003. **100**(12): p. 6940-5.
35. Kirkpatrick, D.S., S.A. Gerber, and S.P. Gygi, *The absolute quantification strategy: a general procedure for the quantification of proteins and post-translational modifications.* Methods, 2005. **35**(3): p. 265-273.
36. Meinnel, T. and C. Giglione, *Tools for analyzing and predicting N-terminal protein modifications.* Proteomics, 2008. **8**(4): p. 626-649.
37. Ong, S.E. and M. Mann, *Mass spectrometry-based proteomics turns quantitative.* Nature Chemical Biology, 2005. **1**(5): p. 252-262.
38. Putz, S., et al., *Mass spectrometry-based peptide quantification: applications and limitations.* Expert Review of Proteomics, 2005. **2**(3): p. 381-392.

39. Zybailov, B., et al., *Statistical analysis of membrane proteome expression changes in Saccharomyces cerevisiae.* Journal of Proteome Research, 2006. **5**(9): p. 2339-2347.
40. Meselson, M. and F.W. Stahl, *The Replication of DNA in Escherichia coli.* Proceedings of the National Academy of Sciences of the United States of America, 1958. **44**(7): p. 671-682.
41. Pan, C.L., et al., *ProRata: A quantitative proteomics program for accurate protein abundance ratio estimation with confidence interval evaluation.* Analytical Chemistry, 2006. **78**(20): p. 7121-7131.
42. Strader, M.B., et al., *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): p. 965-78.
43. VerBerkmoes, N.C., et al., *Determination and comparison of the baseline proteomes of the versatile microbe Rhodopseudomonas palustris under its major metabolic states.* J Proteome Res, 2006. **5**(2): p. 287-98.
44. Neu, H.C. and L.A. Heppel, *Release of Enzymes from Escherichia coli by Osmotic Shock and during the Formation of Spheroplasts.* Journal of Biological Chemistry, 1965. **240**(9): p. 3685-&.
45. Ewis, H.E. and C.D. Lu, *Osmotic shock: A mechanosensitive channel blocker can prevent release of cytoplasmic but not periplasmic proteins.* Fems Microbiology Letters, 2005. **253**(2): p. 295-301.
46. McCormack, A.L., et al., *Direct analysis and identification of proteins in mixtures by LC/MS/MS and database searching at the low-femtomole level.* Anal Chem, 1997. **69**(4): p. 767-76.
47. Rigaut, G., et al., *A generic protein purification method for protein complex characterization and proteome exploration.* Nat Biotechnol, 1999. **17**(10): p. 1030-2.
48. Gavin, A.C., et al., *Functional organization of the yeast proteome by systematic analysis of protein complexes.* Nature, 2002. **415**(6868): p. 141-7.
49. Ho, Y., et al., *Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry.* Nature, 2002. **415**(6868): p. 180-3.
50. Butland, G., et al., *Interaction network containing conserved and essential protein complexes in Escherichia coli.* Nature, 2005. **433**(7025): p. 531-537.
51. Gingras, A.C., R. Aebersold, and B. Raught, *Advances in protein complex analysis using mass spectrometry.* Journal Of Physiology-London, 2005. **563**(1): p. 11-21.
52. Klibanov, A.M., *Improving enzymes by using them in organic solvents.* Nature, 2001. **409**(6817): p. 241-246.
53. Gupta, M.N. and I. Roy, *Enzymes in organic media - Forms, functions and applications.* European Journal Of Biochemistry, 2004. **271**(13): p. 2575-2583.
54. Welinder, K.G., *Generation of peptides suitable for sequence analysis by proteolytic cleavage in reversed-phase high-performance liquid chromatography solvents.* Anal Biochem, 1988. **174**(1): p. 54-64.
55. Russell, W.K., Z.Y. Park, and D.H. Russell, *Proteolysis in mixed organic-aqueous solvent systems: applications for peptide mass mapping using mass spectrometry.* Anal Chem, 2001. **73**(11): p. 2682-5.
56. Strader, M.B., et al., *Efficient and specific trypsin digestion of microgram to nanogram quantities of proteins in organic-aqueous solvent systems.* Anal Chem, 2006. **78**(1): p. 125-34.

57. http://www.promega.com/tbs/9piv511/9piv511.pdf, Promega Usage Information for Sequencing-Grade Modified Trypsin.
58. Riviere, L.R.T., P., *Enzymatic Digestion of Proteins in Solution*, in *Current Protocols in Protein Science*, J.E. Coligan, Editor. 1995, Wiley: Hoboken.
59. McDonald, W.H., et al., *Comparison of three directly coupled HPLC MS/MS strategies for identification of proteins from complex mixtures: single-dimension LC-MS/MS, 2-phase MudPIT, and 3-phase MudPIT.* Int J Mass Spectrom, 2002. **219**(1): p. 245-251.
60. Klammer, A.A., and MacCoss, M. J., *Effects of Modified Digestion Schemes on the Identification of Proteins from Complex Mixtures.* J Proteome Res, 2006. **5**(3): p. 695-700.
61. Yu, Y.Q., et al., *Enzyme-friendly, mass spectrometry-compatible surfactant for in-solution enzymatic digestion of proteins.* Anal Chem, 2003. **75**(21): p. 6023-8.
62. Yu, Y.Q., M. Gilar, and J.C. Gebler, *A complete peptide mapping of membrane proteins: a novel surfactant aiding the enzymatic digestion of bacteriorhodopsin.* Rapid Commun Mass Spectrom, 2004. **18**(6): p. 711-5.
63. Dehmelt, L. and S. Halpain, *The MAP2/Tau family of microtubule-associated proteins.* Genome Biology, 2004. **6**(1).
64. ThermoFinnigan. *Product Support Bulletin 105*. [cited 2009 February 19]; Available from: http://www.thermo.com/eThermo/CMA/PDFs/Articles/articlesFile_21286.pdf.
65. Yates, J.R., et al., *Method To Correlate Tandem Mass-Spectra Of Modified Peptides To Amino-Acid-Sequences In The Protein Database.* Anal Chem, 1995. **67**(8): p. 1426-1436.
66. Larimer, F.W., et al., *Complete genome sequence of the metabolically versatile photosynthetic bacterium Rhodopseudomonas palustris.* Nat Biotechnol, 2004. **22**(1): p. 55-61.
67. http://www.ebi.ac.uk/IPI/IPIcow.html, *International Protein Index for the Cow.*
68. http://www.ornl.gov/sci/GenomestoLife/index.shtml, ORNL Genome Annotation Pipeline for *R. palustris*.
69. Tabb, D.L., W.H. McDonald, and J.R. Yates, 3rd, *DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics.* J Proteome Res, 2002. **1**(1): p. 21-6.
70. Haynes, P.A., et al., *The wildcat toolbox: a set of perl script utilities for use in peptide mass spectral database searching and proteomics experiments.* J Biomol Tech, 2006. **17**(2): p. 97-102.
71. Lundell, N. and T. Schreitmuller, *Sample preparation for peptide mapping - A pharmaceutical quality-control perspective.* Analytical Biochemistry, 1999. **266**(1): p. 31-47.
72. Wolfe, S.L., *Introduction to Cell and Molecular Biology*. 1995, Belmont: Wadsworth.
73. Batra, R.G.M., *Enhancement of Enzyme Activity in Aqueous-Organic Solvent Mixtures.* Biotechnology Letters, 1994. **16**(10): p. 1059-1064.
74. Olmstead, J.B., *Microtubule-Associated Proteins.* Annual Review of Cell Biology, 1986. **2**: p. 421-457.
75. Hyman, A.D., D.; Kellog, D.; Salser, S.; Sawin, K.; Steffen, P.; Wordeman, L.; and Mitchison, T., *Preparation of Modified Tubulins.* Methods in Enzymology, 1991. **196**: p. 478-485.
76. Berggard, T., S. Linse, and P. James, *Methods for the detection and analysis of protein-protein interactions.* Proteomics, 2007. **7**(16): p. 2833-2842.

77. Sharp, J.L., et al., *Statistically inferring protein-protein associations with affinity isolation LC-MS/MS assays.* Journal of Proteome Research, 2007. **6**(9): p. 3788-3795.
78. Musso, G.A., Z.L. Zhang, and A. Emili, *Experimental and computational procedures for the assessment of protein complexes on a genome-wide scale.* Chemical Reviews, 2007. **107**(8): p. 3585-3600.
79. Zeghouf, M., et al., *Sequential Peptide Affinity (SPA) system for the identification of mammalian and bacterial protein complexes.* Journal of Proteome Research, 2004. **3**(3): p. 463-468.
80. Pelletier, D.A., et al., *A general system for studying protein-protein interactions in gram-negative bacteria.* Journal of Proteome Research, 2008. **7**(8): p. 3319-3328.
81. Glick, B.R., *Metabolic Load and Heterologous Gene-Expression.* Biotechnology Advances, 1995. **13**(2): p. 247-261.
82. Tackett, A.J., et al., *I-DIRT, a general method for distinguishing between specific and nonspecific protein interactions.* Journal of Proteome Research, 2005. **4**(5): p. 1752-1756.
83. Lee, D.J., et al., *Affinity isolation and I-DIRT mass spectrometric analysis of the Escherichia coli O157 : H7 Sakai RNA polymerase complex.* Journal of Bacteriology, 2008. **190**(4): p. 1284-1289.
84. Khalsa-Moyers, G., et al., *Unpublished Work*. 2005-2009, Oak Ridge National Laboratory: Oak Ridge, TN.
85. Sambrook, J. and D.W. Russell, *Appendix A.1: Bacterial Media, Antibiotics, and Bacterial Strains*, in *Molecular Cloning: A Laboratory Manual*. 1989, Cold Spring Harbor Laboratory Press: Plainview, NY.
86. Pan, C., et al., *Characterization of anaerobic catabolism of p-coumarate in Rhodopseudomonas palustris by integrating transcriptomics and quantitative proteomics.* Molecular & Cellular Proteomics, 2008. **7**(5): p. 938-948.
87. Hervey, W.J., M.B. Strader, and G.B. Hurst, *Comparison of digestion protocols for microgram quantities of enriched protein samples.* Journal of Proteome Research, 2007. **6**(8): p. 3054-3061.
88. *HAMAP: Escherichia coli (strain K12) Complete Proteome.* [cited 2008 January 29]; Available from: http://ca.expasy.org/sprot/hamap/ECOLI.html.
89. *Rhodopseudomonas palustris CGA009 Complete Genome Sequence and Annotation*. [cited 2008 January 29]; Available from: http://genome.ornl.gov/microbial/rpal.
90. Elias, J.E. and S.P. Gygi, *Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry.* Nature Methods, 2007. **4**(3): p. 207-214.
91. Chamberlin, M.J., *Introduction to RNA Polymerase*, in *RNA Polymerase*, R. Losick and M.J. Chamberlin, Editors. 1976, Cold Spring Laboratory Press: Cold Spring Harbor, NY. p. 17-67.
92. Wolfe, S.L., *mRNA Transcription in E. coli*, in *Introduction to Cell and Molecular Biology*. 1995, Wadsworth: Belmont, CA.
93. *COGnitor*. [cited 2008 August 28]; Available from: http://www.ncbi.nlm.nih.gov/COG/old/xognitor.html.
94. *Kyoto Encyclopedia of Genes and Genomes*. [cited 2008 September 15]; Available from: ftp://ftp.genome.jp/pub/kegg/pathway/organisms.

95. *Virtual Institute for Microbial Stress and Survival Operon Prediction*. [cited 2008 September 25]; Available from: http://www.microbesonline.org/operons/OperonList.html.
96. Liu, H.B., R.G. Sadygov, and J.R. Yates, *A model for random sampling and estimation of relative protein abundance in shotgun proteomics.* Analytical Chemistry, 2004. **76**(14): p. 4193-4201.
97. Ghosh, P., C. Ramakrishnan, and D. Chatterji, *Inter-subunit recognition and manifestation of segmental mobility in Escherichia coli RNA polymerase: a case study with omega-beta ' interaction.* Biophysical Chemistry, 2003. **103**(3): p. 223-237.
98. Kanehisa, M., et al., *KEGG for linking genomes to life and the environment.* Nucleic Acids Research, 2008. **36**: p. D480-D484.
99. Kanehisa, M. and S. Goto, *KEGG: Kyoto Encyclopedia of Genes and Genomes.* Nucleic Acids Research, 2000. **28**(1): p. 27-30.
100. Kanehisa, M., et al., *From genomics to chemical genomics: new developments in KEGG.* Nucleic Acids Research, 2006. **34**: p. D354-D357.
101. Price, M.N., et al., *A novel method for accurate operon predictions in all sequenced prokaryotes.* Nucleic Acids Research, 2005. **33**(3): p. 880-892.
102. *KEGG Map of Branched-Chain Amino Acid Biosynthesis*. [cited 2008 August 28]; Available from: http://www.genome.ad.jp/kegg/pathway/map00290.html.
103. Birnbaum, S. and J.E. Bailey, *Plasmid Presence Changes the Relative Levels of Many Host-Cell Proteins and Ribosome Components in Recombinant Escherichia-Coli.* Biotechnology and Bioengineering, 1991. **37**(8): p. 736-745.
104. *KEGG Enzyme 4.4.1.8*. [cited 2008 August 28]; Available from: http://www.genome.ad.jp/dbget-bin/www_bget?enzyme+4.4.1.8.
105. *KEGG Enzyme 2.5.1.49*. [cited 2008 August 28]; Available from: http://www.genome.ad.jp/dbget-bin/www_bget?enzyme+2.5.1.49.
106. Dreger, M., *Subcellular proteomics.* Mass Spectrometry Reviews, 2003. **22**(1): p. 27-56.
107. Yates, J.R., et al., *Proteomics of organelles and large cellular structures.* Nature Reviews Molecular Cell Biology, 2005. **6**(9): p. 702-714.
108. Schirmer, E.C., et al., *Nuclear membrane proteins with potential disease links found by subtractive proteomics.* Science, 2003. **301**(5638): p. 1380-1382.
109. Dunkley, T.P.J., et al., *Localization of organelle proteins by isotope tagging (LOPIT).* Molecular & Cellular Proteomics, 2004. **3**(11): p. 1128-1134.
110. Link, A.J., E. Carmack, and J.R. Yates, *A strategy for the identification of proteins localized to subcellular spaces: Application to E-coli periplasmic proteins.* International Journal of Mass Spectrometry and Ion Processes, 1997. **160**(1-3): p. 303-316.
111. Fulda, S., et al., *Isolation of salt-induced periplasmic proteins from Synechocystis sp. strain PCC 6803.* Archives of Microbiology, 1999. **171**(3): p. 214-217.
112. Chi, A., et al., *Periplasmic proteins of the extremophile Acidithiobacillus ferrooxidans.* Molecular & Cellular Proteomics, 2007. **6**(12): p. 2239-2251.
113. Ferguson, S.J., *The Periplasm*, in *Prokaryotic Structure and Function: A New Perspective*, S. Mohan, C. Dow, and J.A. Coles, Editors. 1991, Cambridge University Press: Cambridge. p. 311-339.
114. Berks, B.C., *A common export pathway for proteins binding complex redox cofactors?* Molecular Microbiology, 1996. **22**(3): p. 393-404.

115. SwissProt. *Sequence Retrieval Service (SRS) Version 5*. 2008 October 21, 2008]; Available from: http://ca.expasy.org/srs5bin/cgi-bin/wgetz?-fun+Pagelibinfo+-info+SWISS_PROT.
116. Hiller, K., et al., *PrediSi: prediction of signal peptides and their cleavage positions.* Nucleic Acids Research, 2004. **32**: p. W375-W379.
117. Bendtsen, J.D., et al., *Improved prediction of signal peptides: SignalP 3.0.* Journal of Molecular Biology, 2004. **340**(4): p. 783-795.
118. Jiang, X.S., et al., *A comparative proteomic strategy for subcellular proteome research - ICAT approach coupled with bioinformatics prediction to ascertain rat liver mitochondrial proteins and indication of mitochondrial localization for catalase.* Molecular & Cellular Proteomics, 2005. **4**(1): p. 12-34.
119. Emanuelsson, O., et al., *Locating proteins in the cell using TargetP, SignalP and related tools.* Nature Protocols, 2007. **2**(4): p. 953-971.
120. Link, A.J., K. Robison, and G.M. Church, *Comparing the predicted and observed properties of proteins encoded in the genome of Escherichia coli K-12.* Electrophoresis, 1997. **18**(8): p. 1259-1313.
121. Barron, A., J.U. Jung, and M. Villarejo, *Purification and Characterization of a Glycine Betadine Binding-Protein from Escherichia coli.* Journal of Biological Chemistry, 1987. **262**(24): p. 11841-11846.
122. Webb, E., K. Claas, and D. Downs, *thiBPQ encodes an ABC transporter required for transport of thiamine and thiamine pyrophosphate in Salmonella typhimurium.* Journal of Biological Chemistry, 1998. **273**(15): p. 8946-8950.
123. Linton, K.J. and C.F. Higgins, *The Escherichia coli ATP-binding cassette (ABC) proteins.* Molecular Microbiology, 1998. **28**(1): p. 5-13.
124. Graubner, W., A. Schierhorn, and T. Bruser, *DnaK plays a pivotal role in Tat targeting of CueO and functions beside SlyD as a general Tat signal binding chaperone.* Journal of Biological Chemistry, 2007. **282**(10): p. 7116-7124.
125. Yim, H.H. and M. Villarejo, *OsmY, a New Hyperosmotically Indicible Gene, Encodes a Periplasmic Protein in Escherichia coli.* Journal of Bacteriology, 1992. **174**(11): p. 3637-3644.
126. Rioux, C.R. and R.J. Kadner, *Vitamin-B12 Transport in Escherichia coli K12 Does Not Require the btue Gene of the btuced Operon.* Molecular & General Genetics, 1989. **217**(2-3): p. 301-308.
127. Cadieux, N., et al., *Identification of the periplasmic cobalamin-binding protein BtuF of Escherichia coli.* Journal of Bacteriology, 2002. **184**(3): p. 706-717.
128. Sturm, A., et al., *YcdB from Escherichia coli reveals a novel class of Tat-dependently translocated hemoproteins.* Journal of Biological Chemistry, 2006. **281**(20): p. 13972-13978.
129. Ajouz, B., et al., *Release of thioredoxin via the mechanosensitive channel MscL during osmotic downshock of Escherichia coli cells.* Journal of Biological Chemistry, 1998. **273**(41): p. 26670-26674.

# Appendices

**Appendix A.** Efficient and Specific Trypsin Digestion of Microgram to Nanogram Quantities of Proteins in Organic−Aqueous Solvent Systems.

**Appendix B.** A General System for Studying Protein-Protein Interactions in Gram-Negative Bacteria.

**Appendix C.** Modified Perl Script for conducing Automated SEQUEST searches on *nix Operating Systems

**Appendix D.** $H_2O$ Isolation Protocol, All PP:WCL Protein Isotope Ratios (Chapter 5)

**Appendix E.** $GdCl_3$ Isolation Protocol, All PP:WCL Protein Isotope Ratios (Chapter 5)

**Appendix F.** $H_2O$ Isolation Protocol, All PP:PP Protein Isotope Ratios (Chapter 5)

**Appendix G.** $GdCl_3$ Isolation Protocol, All PP:PP Protein Isotope Ratios (Chapter 5)

Appendices D-G lists all measured $^{15}N$:$^{14}N$ protein isotope ratios described in Chapter 5.

The data are listed in 5 columns. The *Protein* column contains the SwissProt accession number and protein name. The *PP* column designates periplasmic annotation; a 1 in this column indicates periplasmic annotation, a 0 indicates a non-periplasmic annotation. The *$log_2$ ratio* is the measured ProRata protein isotope ratio. The number of chromatographic features (or peaks) the measurement was calculated from is listed in the *Chr. Features* column. The number of experimental replicates is listed in the *Reps* column. The protein description from the SwissProt annotation is shown in the sixth column, *Protein Description*.

# Appendix A
# Efficient and Specific Trypsin Digestion of Microgram to Nanogram Quantities of Proteins in Organic−Aqueous Solvent Systems

Strader *et al. Analytical Chemistry*, 2006, **78** (1), 125–134.
http://pubs.acs.org/doi/abs/10.1021/ac051348l

As a co-author of this publication, W. J. Hervey. IV performed sample preparation, collection of data on the LCQ Deca XP instrument, and critically read the final manuscript prior to and after submission.

# Efficient and Specific Trypsin Digestion of Microgram to Nanogram Quantities of Proteins in Organic-Aqueous Solvent Systems

*Michael Brad Strader[1]†, David L. Tabb[2]¶, W. Judson Hervey†‡, Chongle Pan†‡, and Gregory B. Hurst†**

[†]Organic and Biological Mass Spectrometry Group, Chemical Sciences Division, Oak Ridge National Laboratory, P.O. Box 2008, Oak Ridge, Tennessee 37831-6131

[¶]Genome Analysis and Systems Modeling Group, Life Sciences Division, Oak Ridge National Laboratory

[‡]Graduate School of Genome Science and Technology, University of Tennessee-Oak Ridge National Laboratory

[1]Present address: Laboratory of Neurotoxicology, National Institute of Mental Health, Bethesda, MD 20892

[2]Present address: Department of Biomedical Informatics, Vanderbilt University Medical Center, 465 21st Ave. S., U9211 MRB III, Nashville, TN 37232-8575.

Revised submission to *Analytical Chemistry*.
*Author for Correspondence: Gregory B. Hurst
Phone: (865) 574-6142
Fax: (865) 576-8559
E-mail: hurstgb@ornl.gov

**ABSTRACT**

Mass spectrometry-based identification of the components of multiprotein complexes often involves solution-phase proteolytic digestion of the complex. The affinity purification of individual protein complexes often yields nanogram to low microgram amounts of protein, which poses several challenges for enzymatic digestion and protein identification. We tested different solvent systems to optimize trypsin digestions of samples containing limited amounts of protein for subsequent analysis by LC-MS-MS. Data collected from digestion of 10 μg, 2 μg, 1 μg and 0.2 μg portions of a protein standard mixture indicated that an organic aqueous solvent system containing 80% acetonitrile consistently provided the most complete digestion, producing more peptide identifications than the other solvent systems tested. For example, a one hour digestion in 80% acetonitrile yielded over 52% more peptides than the overnight digestion of 1 μg of a protein mixture in purely aqueous buffer. This trend was also observed for peptides from digested ribosomal proteins isolated from *Rhodopseudomonas palustris.* In addition to improved digestion efficiency, the shorter digestion times possible with the organic solvent also improved trypsin specificity, resulting in smaller numbers of semi-tryptic peptides than an overnight digestion protocol using an aqueous solvent. The technique was also demonstrated for an affinity-isolated protein complex, GroEL. To our knowledge, this report is the first using mass spectrometry data to show a linkage between digestion solvent and trypsin specificity.

## INTRODUCTION

Mass spectrometry (MS) has become a widely used method for studying proteins, protein complexes, and whole proteomes because of innovations in soft ionization techniques, bioinformatics and chromatographic separation techniques.[1-7] An example of a high-throughput mass spectrometry strategy commonly used for this purpose is a variation of the "shotgun" approach, involving in-solution digestion of a protein complex followed by one dimensional (1D) or two dimensional (2D) liquid chromatography (LC) coupled with electrospray ionization (ESI) MS-MS.[6-8] One of the applications of this method is for characterizing multi-protein complexes by identifying large numbers of proteins in a single data acquisition.[9] Large-scale implementations of this strategy have been reported for yeast and *E. coli*.[10-12] In order to achieve a goal of characterizing large numbers of protein complexes[13] isolated by affinity purification from *Rhodopseudomonas palustris*,[14] an efficient protocol for digesting these complexes is required.

The isolation of individual protein complexes by affinity purification often yields small amounts of protein, complicating enzymatic digestion and peptide identification.[15, 16] In many cases, samples may contain less than 100 nanograms of total protein, resulting in low protein concentrations (≤ 10 ng/μl) that are unsuitable for efficient enzymatic digestion. One could keep protein concentrations higher by working in smaller volumes, but unless microfluidic techniques are invoked, a few microliters is a lower practical volume limit for conventional pipetting techniques. Because the reaction rate of digestion is proportional to the concentration[17], dilute protein concentrations may lengthen digestion times. If digestion remains incomplete, large peptide fragments from incompletely digested proteins may not be suitable for identification via MS-MS analysis. For example, these larger peptides may be outside the scan range of the mass

spectrometer. Other characteristics of large peptides, such as higher charge states (>3) or multiple missed enzymatic cleavage sites, may place them beyond the default settings for software used for automatic peptide identification from MS or MS-MS data. To promote more complete digestion, it is useful to identify proteolytic digestion methods that maximize peptide yields from samples containing low microgram to high nanogram amounts of total protein.

Digestion strategies may employ organic solvents, heat, chaotropes, or surfactants to denature proteins before digestion to render more of the protein's structure accessible to the proteolytic enzyme.[18-21] Thermal denaturation often results in sample loss due to precipitation because many proteins are susceptible to aggregation when treated with heat. Chaotropes and surfactants, on the other hand, can inactivate proteases at high concentrations required for denaturation and therefore require dilution of the sample prior to digestion.[22] Dilution permits proteins to refold, reduces proteolytic activity by decreasing substrate concentration, and can increase surface area available for adsorptive loss of peptides on container walls. In addition, chaotropes and surfactants can compete with peptide ions for adsorption on stationary phases during liquid chromatography and for charge during mass spectrometry.[21, 23, 24] To avoid this interference, these chemicals can be removed, although this step can lead to additional loss of peptides. An acid-labile surfactant has recently been introduced that decomposes into insoluble degradation products.[25] These degradation products, however, can co-precipitate with hydrophobic peptides and result in sample loss. To minimize the required amount of starting material while maximizing sensitivity in the MS measurement, one must avoid when possible the drawbacks associated with the described digestion strategies.

Addition of organic solvents such as methanol or acetonitrile to buffers can assist in digestion by unfolding and solubilizing proteins.[26, 27] Trypsin and some other proteases are resistant to

unfolding in organic solvents, retaining activity under conditions that denature other proteins.[28] Organic solvents can be removed after digestion by lyophilization, providing a more efficient "clean-up" step compared to removal of chemical denaturants, etc. Russell *et al.* have reported that addition of organic solvents can accelerate trypsin reactions for proteolysis-resistant proteins and reduce digestion times to less than an hour.[19] These digestions, however, were analyzed using matrix assisted laser desorption ionization (MALDI)-MS, making the benefits of this digestion technique less clear for the LC-MS-MS analysis of low microgram and nanogram amounts of protein. Because addition of organic solvents reduces the potential for sample loss by allowing smaller volumes to be used, obviating chaotrope removal steps, accelerating reaction rates, and enabling protease-resistant proteins to be digested, they seemed an ideal choice for protocols designed to digest samples containing limited amounts of protein, such as those obtained from affinity isolations.

For this report, we systematically tested different solvent systems to optimize trypsin digestion of samples containing low microgram to high nanogram amounts of protein. The resulting digestions were analyzed by LC-MS-MS. MS data collected from digestions of samples containing 200 ng to 10 μg protein indicated that an organic-aqueous solvent system containing 80% acetonitrile consistently resulted in the most complete digestion, producing more peptide identifications than several other solvent systems. In addition, the shorter digestion times possible with addition of organic solvents resulted in smaller numbers of semi-tryptic-peptides (peptides resulting from cleavage at one end at a residue other than lysine or arginine) than an overnight digestion protocol using an aqueous solvent. The use of an 80% acetonitrile solvent achieved both more numerous peptide identifications and reduced nonspecific cleavages. This system was then compared with an overnight digestion in aqueous buffer for digestions of

protein complexes isolated from *Rhodopseudomonas palustris* by sucrose density gradient fractionation (the 70S ribosome) and by affinity purification (GroEL).

## MATERIALS AND METHODS

**Materials.** All proteins, salts, buffers, dithiothreitol (DTT), guanidine HCl, trifluoroacetic acid, diethyl pyrocarbonate (DEPC), phenylmethylsulfonyl fluoride (PMSF), sucrose and RNase-free DNase I were obtained from Sigma Chemical Co. (St. Louis, MO). RNase Away was obtained from Molecular BioProducts (San Diego, CA). Sequencing-grade trypsin was purchased from Promega (Madison, WI). Formic acid was obtained from EM Science (Affiliate of Merck KgaA, Darmstadt, Germany). HPLC grade acetonitrile and water were used for all LC-MS analyses (Burdick & Jackson, Muskegon, MI). Ultrapure 18 MΩ water obtained from a Millipore Milli-Q system (Bedford, MA) was used for sample buffers. Fused silica capillary tubing was purchased from Polymicro Technologies (Phoenix, AZ). BCA assay reagent and standards were obtained from Pierce Chemical Co. (Rockford, IL).

**Construction of Protein Standard Mixture.** Protein standard mixtures were generated using 6 proteins: bovine serum albumin (MW 69 kDa), yeast alcohol dehydrogenase I (MW 37 kDa), bovine carbonic anhydrase II (MW 29 kDa), horse myoglobin (MW 17 kDa), bovine hemoglobin (MW 15 kDa) and chicken egg lysozyme C (MW 14 kDa). Hemoglobin includes α and β polypeptides, and the isomer yeast alcohol dehydrogenase II was found to be a component of yeast alcohol dehydrogenase I, giving a total of 8 polypeptides in the mixture. Mixtures contained equal masses of each protein. Each of the proteins was dissolved in 50 mM Tris-HCl, 10 mM $CaCl_2$ (pH 7.6) and then combined in equal masses to give 10 μg, 2 μg, 1 μg, and 200 ng of total protein in final volumes of 100 μl of one of the digestion solvents described below.

**Proteolytic Digestion of Protein Standard Mixture**. The same amount of trypsin (200 ng) was added to each sample for digestion. For 10 μg, 2 μg, 1 μg and 200 ng samples the enzyme to substrate ratios (wt/wt) were, therefore, 1:50, 1:20, 1:5 and 1:1 respectively. The protein mixture digestions, performed in triplicate, differed also by the use of different solvent conditions (organic or aqueous) and incubation time for digestion. For all amounts of protein standard mixtures, trypsin digestions were performed in 100 μl of each solvent, using conditions listed in Table 1.

A control sample of 10 μg was digested using the manufacturer's protocol for the trypsin used in these experiments, that included denaturation in 6M guanidine HCl, 50 mM Tris-HCl, 10 mM $CaCl_2$ (pH 7.6) for 45 minutes followed by dilution to 0.5M guanidine HCl and overnight digestion with 200 ng trypsin at 37°C. The resultant peptides from this control digestion were desalted using solid phase extraction ($C_{18}$ Zip-Tip, Millipore, Billerica, MA).

After digestion, all peptide samples were treated with DTT (20 mM) for 1 hour at 37°C to reduce disulfide bonds. The reduced peptides were lyophilized and resuspended in 100 μl of 95% $H_2O$/5% acetonitrile/0.1% formic acid. To inhibit further trypsin activity, 2 μl of 10% formic acid were added to each resuspended sample. All samples were stored at -80°C until analysis.

**Isolation and Digestion of *R. palustris* Ribosomal Proteins.** 70S ribosomes from *R. palustris* were purified and fractionated using a high salt sucrose cushion and sucrose density fractionation as previously described.[29] Ribosomal protein extraction and the removal of contaminant rRNA was performed using the acid extraction method.[30] After overnight dialysis in a 3500 molecular weight cutoff dialysis cassette (Slide-A-Lyzer, Pierce, Rockford, IL) against Ultrapure water, the protein samples were concentrated and then quantitated using the BCA

assay according to manufacturer's instructions (Pierce, Rockford, IL). Ribosomal protein samples were concentrated by solvent evaporation, reconstituted in 50 mM Tris-HCl, 10 mM $CaCl_2$ (pH 7.6) and digested with 200 ng of trypsin in solvent systems 1 and 4 (Table 1). Each digestion was performed in duplicate.

**Tandem Affinity Purification and Digestion of *R. palustris* Groel2 Complex.** The *R. palustris* wild type strain (CGA009), harboring the pBBR5-DEST/42 modified Gateway expression plasmid (Invitrogen, Carlsbad, CA) with the GroEL2 open reading frame (ORF) was a generous gift from Dr. Dale Pelletier at the Oak Ridge National Laboratory. The ORF was cloned into the expression plasmid such that V5 and 6xHis affinity tags were fused at the C-terminus of the protein. The presence of both tags allowed the use of a dual affinity purification strategy to "capture" the GroEL complex, using a strategy similar to tandem affinity purification (TAP).[31] The first affinity purification was a Ni-NTA capture in which the C-terminal 6xHis tag chelated the Ni-NTA resin (Qiagen, Valencia, CA). The second affinity purification was a capture using V5 resin (Sigma; anti-V5 antibody conjugated to agarose beads). This is our standard protocol for large-scale isolation of protein complexes from *R. palustris*, in which a large number of strains each bears a plasmid encoding a different affinity-tagged protein.[13]

*R. palustris* cells harboring the expression plasmid were grown anaerobically and harvested at mid-log phase ($O.D._{660}$ ~0.8).[32] Cell pellets were resuspended in NTA binding buffer (50 mM $NaH_2PO_4$ at pH 8, 300 mM NaCl, 10 mM imidazole, 5 mM ATP, 10 mM $MgCl_2$, 10 mM KCl, 100 μg/ml PMSF and 10 μg/ml leupeptin) and then sonicated with a Cell Disruptor 185 (Amphotech, Beverly, MA) using a series of six 15 second pulses separated by 30 second cooling intervals. Cellular debris was removed in an initial centrifugation at 4°C using an SS-34 Sorval rotor at 12,100 × *g* for 30 minutes. The supernatant was centrifuged for an additional 15

minutes at 23,700×*g*. The final resulting supernatant was then immediately used in the first stage of the affinity purification of GroEL.

**Ni-NTA Capture:** After the addition of 100 µl of 50% Ni-NTA bead suspension (previously washed in NTA Binding buffer 4X), the supernatants were incubated on a rotator for one hour at ambient temperature. The beads were then collected by centrifugation at 425×*g*, transferred to new tubes, and washed 4X with NTA wash buffer (50 mM $NaH_2PO_4$ at pH 8, 300 mM NaCl, 20 mM imidazole, 5 mM ATP, 10 mM $MgCl_2$, 10 mM KCl). Afterwards, bound proteins were eluted from the Ni-NTA beads 4 times with 50 µl NTA elution buffer (50 mM $NaH_2PO_4$ at pH 8, 300 mM NaCl, 500 mM imidazole, 5 mM ATP, 10 mM $MgCl_2$, 10 mM KCl). Combined eluents (approximately 200 µl total) were diluted with 400 µl buffer (5 mM ATP, 10 mM $MgCl_2$, 10 mM KCl) and either stored at -80°C or immediately used for the second affinity purification step.

**V5 Capture:** 100 µl of a 50% V5 bead suspension (previously washed in PBS buffer) were added to the combined eluents from the Ni-NTA capture and incubated on a rotator for one hour at ambient temperature. The beads were then centrifuged at 425×*g* and washed 4 times with V5 wash buffer (50 mM Tris-HCl, 10 mM $CaCl_2$ at pH 7.6, 5 mM ATP, 10 mM $MgCl_2$, 10 mM KCl). Afterwards, the bound proteins were eluted three times from the V5 beads with 50 µl V5 elution buffer (80% acetonitrile, 20% 50 mM Tris-HCl, 10 mM $CaCl_2$ at pH 7.6). The combined eluents (approximately 175 µl total) were digested with 200 ng of trypsin in solvent systems 1 and 4 (Table 1). Because of limited sample amounts, only one digestion of GroEL for each of the two solvent conditions was performed. After digestion, peptides were treated with DTT, lyophilized and resuspended in 100 µl of 95% $H_2O$/5% acetonitrile/0.1% formic acid. Aliquots

at each stage of the affinity purification were analyzed by Western blot using anti-V5 antibodies to determine the purification efficiency.

**1D LC-MS-MS Analysis.** For all peptide samples, 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 as previously described.[33] For all 1D LC-MS-MS data acquisition, 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.

**1D LC-FTICR-MS Analysis.** 1D LC-FTICR-MS experiments were performed with an Ultimate HPLC system coupled with a HiResESI Fourier-transform ion cyclotron resonance mass spectrometer (IonSpec, Lake Forest, CA) equipped with a 9.4T magnet (Cryomagnetics Inc., Oak Ridge, TN). Samples were separated with a Vydac 218MS5.30015 C18 column (300 μm id × 15cm, 300Å pore size, 5 μm particles) at a flow rate of 4 ul/min and directly introduced to the FTICR MS with an electrospray source (Analytica, Branford CT).

**Protein Identification from MS Data Analysis.** The SEQUEST algorithm was used to match experimental MS-MS spectra with their counterparts predicted from a protein sequence database.[34] An unconstrained database search was employed so that peptides resulting from cleavage at residues other than lysine or arginine at one end (semi-tryptic peptides) or both ends (non-tryptic) could be identified. The sequence database used for searches in this manuscript consisted of two major elements. The 4833 ORFs of the published *R. palustris* database[14] were search targets for the ribosome and GroEL searches but acted as distractors (indicators of false

positive identifications) during the protein standard mixture searches. Sequences for the eight proteins in the standard mixture were also included in the database; we added the sequence for alcohol dehydrogenase II because this protein was observed as a component in the alcohol dehydrogenase I standard. The total count of proteins in the protein standard mixture database was 4841, yielding a 0.17% chance of randomly hitting one of the eight standard proteins.

DTASelect assembled, filtered and compared the identifications from SEQUEST searches on all data sets. This software sorts peptide identifications by the proteins that contain them.[5] A protein in the mixture was considered successfully identified if at least 2 component peptides passed DTASelect's default SEQUEST score cutoffs. Spectra from singly-charged peptides were required to exceed 1.8 in the SEQUEST parameter XCorr, while XCorr values for doubly- and triply-charged peptides were required to exceed 2.5 and 3.5, respectively.[5] The best matching sequence for each spectrum was required to have an XCorr at least 8% greater than the second best (DeltCN $\geq$0.08).

**MS1PeakFinder Algorithm.** Software created in the C++ programming language analyzed the mass spectra collected during each LC separation to catalog the observed ions. For each ion, the chromatographic profile was reconstructed from the intensities reported at its *m/z* through successive scans. Once this list of eluting ions was inferred, the tandem mass spectra were matched to the list, and those that were confidently identified by SEQUEST were flagged.

Scripts in the R statistical environment can be used to visualize these reports.[35] An image called a "matchmap" segregates the ions into three classes: ions observed only during MS scans, ions for which a tandem mass spectrum was collected but not successfully identified, and ions for which a tandem mass spectrum was both collected and confidently identified. These classes were colored yellow, orange, and red, respectively.

## RESULTS AND DISCUSSION

The goal of this research was to optimize our digestion protocol through a systematic test of different solvent systems for dilute or limited protein samples, such as those typically obtained through affinity isolation of protein complexes. To this end, we compared digestion of 10 μg, 2 μg, 1 μg, and 200 ng amounts of a protein standard mixture in several solvent systems, including 100% aqueous buffer (50 mM Tris at pH 7.6, 10 mM $CaCl_2$) and the same aqueous buffer with different organic additives, including 60% methanol, 60% acetonitrile and 80% acetonitrile. To compare digestion efficiencies we used three criteria. First, we determined the total number of peptide identifications, including the number of semi-tryptic peptides, individual protein sequence coverage, and peptides per protein for each digested sample. Second, we analyzed the completeness of digestion by analyzing each LC-MS-MS run to identify the ion elution profiles for ions that were not confidently identified. Finally, we employed LC-FTICR-MS to identify incompletely digested ions eluting late in chromatographic runs. The two solvent systems that yielded the most efficient digestion of the protein standard mixture with respect to these criteria were then compared for digesting two biologically relevant complexes from *Rhodopseudomonas palustris*: the 70S ribosome and GroEL. The 80% acetonitrile system provided results in one hour that were comparable or superior to those obtained overnight using an aqueous buffer. Because times as short as 5 minutes have been reported for digestion of myoglobin in 80% acetonitrile,[19] we did not investigate digestion times longer than 1 hour in buffers containing methanol or acetonitrile.

**Protein Standard Mixture Digestions and LC-MS-MS Analysis**. The protein standard mixture used in this study included eight proteins with different proteolytic susceptibilities, to

emulate a protein complex isolated by affinity methods. Alcohol dehydrogenase I and II, carbonic anhydrase II, lysozyme C and hemoglobin α/β chains are amenable to proteolysis, while globular proteins such as myoglobin, stabilized predominantly by a hydrophobic interior, and serum albumin, whose tertiary structure is stabilized by 17 disulfide bonds, are resistant to proteolytic digestion.[18, 36]

Using 10 μg samples of the protein standard mixture, we compared a digestion protocol including the chaotrope guanidine HCl and a desalting step against protocols that use no guanidine HCl (see Figure 1). Because sample losses can occur during the dilution and desalting processes, we expected that the number of peptide identifications would be lower for the protocol incorporating a chaotrope. This was indeed the case in our study; the 41 peptide identifications (highest number of three replicates) obtained by the denaturant method were lower than the peptide identification numbers obtained for the other digestion protocols, regardless of the amount of the digested sample (see Table 2). Because sample loss would be an even greater concern with lower sample amounts, the denaturant method was not applied to samples containing less than 10 μg of protein.

The one hour digestion in 80% acetonitrile resulted in the highest number of identifications across the entire range of sample quantities. Table 2 shows the number of peptide identifications for three replicate LC-MS-MS analyses of each protein standard mixture sample. To compare solvent digestion efficiency, we used the value representing the highest number of peptides determined out of the three replicate separations. When 1 μg of protein was digested, the 80% acetonitrile solvent resulted in the identification of 52% more peptides than the overnight aqueous digest. The 200 ng and 10 μg quantities of proteins produced either similar or smaller numbers of peptide identifications than the 1 and 2 μg samples. In the least concentrated sample,

digestions with fewer peptides probably resulted from low signal that is characteristic of smaller samples, while the most concentrated sample had a high ratio of substrate to enzyme, resulting in saturation.

A recent report by Olsen *et al.* suggested that MS data of tryptically digested samples should be analyzed for peptide identifications by configuring search algorithms for strict trypsin specificity.[37] In this configuration, potential peptides resulting from cleavages after residues other than lysine or arginine at one end (semi-tryptic peptides) or both ends (non-tryptic) are not considered in database searches. Their results from in-gel trypsin digests of mouse liver proteins indicated that fully tryptic peptides were exclusively identified except for C-terminal peptides cut by trypsin at either lysine or arginine or semi-tryptic peptides resulting from in source fragmentation products of fully tryptic peptides containing internal proline residues.[37] Our results, on the other hand, establish a link between digestion specificity and the solvent in which the digestion is performed; the proportion of fully tryptic peptides depends upon the digestion solvent system used. Table 2 demonstrates that trypsin specificity was lowest for digestions performed overnight in the aqueous solvent; for example, 82% of the peptide identifications from the 1 μg overnight digestion in aqueous solvent were fully tryptic, compared to 97% from the one hour digestion in 80% acetonitrile. From the overnight aqueous digest, the identified peptides included 107 fully tryptic, 23 semi-tryptic, and 0 non-tryptic peptides. This trend was consistent for the other replicates and for the other protein concentrations (data not shown). Manual inspection of the 23 spectra matched to semi-tryptic peptides from the 1 μg overnight digestions, as exemplified by Figure 2, indicated that these peptides had been successfully identified by SEQUEST. Table 3 lists the 23 semi-tryptic peptides identified from the 1 μg overnight digests in aqueous solvent. Interestingly, 21 of the 23 had chymotryptic ends

(tryptophan, tyrosine, phenylalanine, methionine, leucine, alanine, aspartic acid and glutamic acid, at a terminal position of the peptide), suggesting that the L-(1-tosylamido-2-phenyl)-ethylchloromethyl ketone (TPCK) treatment of the commercially available trypsin may not have completely inhibited all contaminant chymotrypsin activity. Because the amount of active chymotrypsin is likely very small, the overnight digestion may have allowed sufficient time for chymotrypsin activity while the shorter digestion times possible with addition of organic solvents resulted in more specific proteolysis reactions.[38] This is supported by the fact that a one hour digest in the same aqueous solvent produced lower proportions of semi-tryptic peptides than an overnight digestion. If the matches listed in Table 3 were identified to random sequences, there would be a greater probability of obtaining non-tryptic identifications rather than semi-tryptic because a majority of possible sequences are non-tryptic. Instead, semi-tryptic identifications were far more common than non-tryptics; only one non-tryptic was identified for all solvent systems at 1 μg of protein digested. Among the peptides listed in Table 3, five were identified in at least one of the other digestion conditions. In addition, several were identified with other charge states or were represented by more than one spectrum.

To further examine the effect of digestion conditions, we examined the sequence coverage and numbers of peptides identified from each protein individually. Table 4 shows these statistics for the replicate with the highest number of peptide identifications for each solvent condition of the 1 μg mixture; the overall trends observed for these data were consistent for other replicates across the entire range of total protein quantities. While most of the proteins yielded similar sequence coverage in each of the digestion conditions, myoglobin, lysozyme and serum albumin varied considerably in response to the conditions in which they were digested. In the case of the myoglobin's proteolytically resistant structure, an overnight digestion in aqueous solvent

resulted in 75% sequence coverage, while a one hour digestion in either 60% methanol or 80% acetonitrile resulted in nearly 100% sequence coverage. Lysozyme digested poorly in aqueous solvent but digested well in all three organic-containing solvents, most likely because these solvents improved its solubility. In fact, digestion of lysozyme in aqueous solvents was so poor that no identifications were obtained in the mixtures containing 10 and 2 μg total protein. For serum albumin, a one hour digestion in 80% acetonitrile resulted in 61 identified peptides, corresponding to 70% of the amino acid sequence. An overnight digestion of the same sample in aqueous buffer resulted in only 16 identified peptides, corresponding to 23% of the serum albumin sequence. Surprisingly, neither 60% acetonitrile nor 60% methanol improved the digestion efficiency of serum albumin substantially. Taken together, the improved sequence coverage for serum albumin suggests that the 80% acetonitrile solvent serves as an excellent denaturant, exposing a larger number of lysine and arginine residues for proteolysis.

**Assessing Digestion Completeness by Visualization of LC-MS-MS Runs.** We evaluated how well trypsin performed in each solvent by analyzing ion elution profiles from each LC-MS-MS run. Incomplete digestion should result in fewer small peptides, and more numerous larger partially digested protein fragments. These large polypeptides can be expected generally to elute later in reverse phase HPLC gradients due to their larger sizes and hydrophobicities. We visualized the ions observed during each separation using the "MS1PeakFinder" algorithm developed at ORNL to reveal the retention times and *m/z* ratios of identifiable and unknown ions for each sample. Figure 3 shows MS1PeakFinder plots for the overnight digestion in aqueous buffer and the one hour digestion in 80% acetonitrile for the 1 μg protein standard mixture. These plots are not intended to provide intensity or chromatographic peak width information, but rather to indicate whether each parent ion was subjected to CID, and whether the tandem mass

spectrum was confidently identified by SEQUEST and DTASelect as a peptide. The overall trends observed in these two plots are representative of data across the entire range of protein quantities.

For both plots, ions corresponding to confident spectral identifications (red peaks) eluted between 20 minutes and 70 minutes. A dense patch of ions, eluting during a retention time interval of 77 to 82 minutes, was seen in the MS1PeakFinder plot representing the overnight digestion (Figure 3a). These ions correspond to peptides that were among the last to elute from the column. A majority of the ions seen in this dense patch did not result in confident identifications, and may represent undigested or partially digested proteins that produced ions with more than three charges, preventing identification by SEQUEST as configured for this work. For the one hour digestion, a less compact patch of species eluted during retention times of 45 to 60 minutes (Figure 3b). While several red peaks were observed in the 45 to 60 minute retention time interval, a majority of the ions detected during this interval were not confidently identified. Because species producing these unidentified ions eluted sooner than the 77 to 82 minute window observed from the overnight digest, they most likely represent smaller proteolytic fragments that reflect a higher level of digestion. The number of different ions from species eluting during a retention time interval of 77 to 82 minutes appeared to decrease with an increasing number of peptide identifications reported in Table 2. For example, the one hour aqueous digestion produced the fewest peptide identifications, and contained 1072 ions detected between 75 and 82 minutes. On the other hand, only 58 ions detected during the same retention times for the 80% acetonitrile digestion, for which the largest number of peptide identifications was observed. In fact, the majority of the ions in the 80% acetonitrile digestion were detected during the 20-70 minute time interval.

**Assessing Digestion Quality by FTICR.** In order to identify species corresponding to the dense patch in Figure 3a eluting between 77 to 82 minutes, we analyzed protein standard mixtures digested overnight in aqueous buffer or digested for one hour in 80% acetonitrile using 1D-LC Fourier transform ion cyclotron resonance (FTICR) mass spectrometry. FTICR allows isotopic resolution of multiply charged ions, allowing the determination of molecular mass to better than 10 ppm accuracy.[39] By knowing the accurate molecular masses, we were able to determine that species eluting during the time interval corresponding to the dense patch of unidentified ions in the LC-MS-MS results shown in Figure 3a represented intact undigested proteins from the mixture and partially digested protein fragments with charge states too high to be identified by SEQUEST as configured for this work. Figure 4 shows the LC-FTICR total ion chromatograms of the mixture digested overnight in aqueous buffer, and digested for one hour in 80% acetonitrile. The C18 reverse phase column, LC instrumentation, and gradients used for these experiments were identical to those used for the LC-MS-MS analyses described above, which were performed using the quadrupole ion trap mass spectrometer for detection. As illustrated in the insets in Figure 4, intact myoglobin (most abundant isotope mass [MAIM][29] measured = 16,950.954 Da, calculated MAIM = 16,950.993 Da, mass error = 2.3 ppm) and hemoglobin $\alpha$ chain (measured MAIM = 15,052.891 Da, calculated MAIM = 15052.929 Da, mass error = 2.5 ppm) were identified in the 75-82 minute retention time window for the overnight aqueous digestion. In addition, several unidentified species, probably partially digested protein fragments, in the 27- 29 kDa range were observed (data not shown). We did not identify any intact proteins in the 75-82 minute retention time window for the 80% acetonitrile digestion shown in Figure 4, and consistent with the LC-MS-MS results, overall ion signal was

lower in this retention time window. These data collectively show that trypsin cut more efficiently in the 80% acetonitrile digestion.

**Analysis of the 70S ribosome and GroEL Complexes from *R. palustris***. One of our major goals is to develop an efficient protocol for high throughput MS analysis of protein complexes from microbes. [13, 40] Currently, we use dual affinity purification to isolate protein complexes, a method that results in variable amounts of protein for each purified complex. Because limited sample amounts can complicate enzymatic digestion and protein identifications, we have focused on tailoring digestion protocols for samples, obtained through dual affinity purification, that are destined for MS analysis. The intent of this report is to describe a protocol for protein digestion that maximizes the number of protein identifications from MS. Our results indicate that the 80% acetonitrile solvent system was most suitable for digesting a protein standard mixture under conditions similar to those imposed by our overall workflow. Biological samples, however, are often significantly more complex than our model mixture, containing a large number of proteins at different concentrations, with different proteolytic susceptibilities and solubilities, and varying amounts and types of impurities. With this in mind, we chose the 70S ribosome (purified by sucrose density fractionation)[29] and dual affinity purified GroEL from *R. palustris* as representative protein complexes to test the effectiveness of the 80% acetonitrile digestion protocol for biologically relevant samples. As a control, we compared the 80% acetonitrile digestions with our standard protocol for affinity isolation samples, which is overnight digestion in aqueous buffer.

Because the 70S ribosome is composed of over 50 proteins, it contains a wider variety of subunits than most complexes that would be obtained from affinity isolation procedures, and thus offers a significant challenge for our approach. Table 5 shows the peptide and protein

identifications obtained from digesting 1 μg and 200 ng of the *R. palustris* 70S ribosome. For both sample amounts, the 80% acetonitrile digestions gave more peptide and protein identifications than the overnight aqueous digestions. For example, 39 proteins were identified for a one hour digestion of 1 μg of ribosomes compared to 30 proteins for the overnight digestion in aqueous solvent. The identifications obtained for the one hour digestion corresponded to 72% of the proteins comprising the 70S ribosome. In contrast to the protein standard mixtures, no semi-tryptic identifications were obtained for either digestion of the ribosome samples; only fully tryptic peptides were identified. Although these samples contained 6 to 7 times as many protein sequences as the protein standard mixture, the 80% acetonitrile digest continued to outperform the overnight aqueous digest. The lower numbers of peptide and protein identifications observed for the smaller 200 ng sample size suggest that other aspects of our workflow, such as the LC-MS-MS analysis, are limiting factors for sensitivity as sample sizes decrease. Indeed, a more comprehensive characterization of the ribosomal proteins requires larger amounts of protein (and replicate analyses) than used in the current work.[29] However, the increased peptide and protein identifications obtained here by using the 80% acetonitrile digestion solvent for both the 1 μg and 200 ng sample sizes suggest that this solvent represents a significant step toward the use of lower amounts of protein for characterization of ribosomes or other large protein complexes.

Table 5 also shows the peptide and protein identifications for dual affinity purified GroEL. In the two digestions we identified both gene products of GroEL genes expressed by *R. palustris* (GroEL1 and GroEL2). These data indicate that a one hour digest in 80% acetonitrile produced results comparable to the overnight aqueous digests. These results suggest that GroEL may be fairly amenable to proteolysis and therefore the potential advantages of the organic solvent were

not needed to maximize peptide yields. The overnight aqueous digest, however, yielded many semi-tryptic peptides; 40% of the 51 peptide identifications were of this class. In contrast, 25% of the identifications were semi-tryptic for the one hour digestion with acetonitrile. The identification of semi-tryptic peptides from the digestion in acetonitrile may have occurred due to endogenous proteases from *R. palustris*, as this sample would have contained the highest complement of impurities of the systems studied. While the results for GroEL do not show an increase in the peptides identifiable for these proteins by use of organic solvents, other biological complexes may contain components that are more proteolytically resistant and that require the organic additive to serve as a denaturant.

## CONCLUSION

Our results demonstrate that effective trypsin digestions of high nanogram to microgram amounts of proteins can be performed in one hour using 80% acetonitrile as the solvent. This protocol does not require the chemical denaturing steps necessary for protein digestion methods that require chaotropes or surfactants, followed by dilution and/or purification steps that can collectively lead to sample loss and diminished sensitivity. As the data from the dual affinity purified GroEL attest, however, the use of acetonitrile in rapid digestions may not improve peptide recovery relative to traditional overnight digests when the target proteins are amenable to digestion. The benefits appear to be most pronounced in protease-resistant proteins such as bovine serum albumin and chicken lysozyme. Large protein complexes, such as the ribosome, still require tens of micrograms of material for *complete* characterization, although we have presented evidence that digestion in 80% acetonitrile allows detection of more ribosomal proteins than the aqueous digestion.

A further benefit of employing 80% acetonitrile in digestion buffers is the reduction of nonspecific cleavage. Because overnight aqueous digestions can result in increased numbers of semi-tryptic peptides, as observed for the protein standard mixture and GroEL samples, database searching with no protease specificity becomes necessary to maximize peptide identification. These searches, however, generate larger numbers of candidate sequences, take more time to run, and yield higher false positive rates. The use of 80% acetonitrile can produce substantially higher proportions of fully tryptic peptides than overnight digestions. As a result, these spectra can be identified by more efficient searches that consider only tryptic peptides.

This digestion method opens new opportunities to efficiently detect and characterize proteins in amounts typically obtained from affinity isolation of biological complexes. First, the ability to minimize sample loss and improve the number of peptide identifications will ultimately allow more sensitive detection of less abundant components from biological complexes. Second, the one hour incubation time possible with this method is a distinct improvement for high throughput applications. The results presented here suggest further investigation of at least two aspects of protocols for digestion in organic-containing solvents. Some proteins are likely to precipitate in solutions with high organic content; in fact, organic precipitation is an accepted method for purifying proteins. The ability of the protocols presented here for detecting such organic-insoluble proteins should be explored. Also, while we did not test the use of 80% acetonitrile to digest larger sample amounts ($\geq$ 20 μg), it is likely that these digestion conditions will be applicable to more abundant samples simply by increasing the volume of solvent. This study, therefore, presents a digestion method that potentially can be adapted to a variety of different types of samples.

**ACKNOWLEDGEMENTS**

We thank Caroline S. Harwood, Department of Microbiology, University of Iowa for the generously contributing the *R. palustris* strain used in this study. We thank the John Yates laboratory at The Scripps Research Institute for DTASelect and Contrast software. W. Judson Hervey IV and Chongle Pan acknowledge support from the University of Tennessee-Oak Ridge National Laboratory Graduate School of Genome Science and Technology. We would like to thank William Hayes McDonald for technical advice and critical evaluation of this manuscript. This research was funded by the U.S. DOE Office of Biological and Environmental Sciences Genomes to Life Program. Oak Ridge National Laboratory is operated and managed by the University of Tennessee-Battelle, L.L.C. for the U.S. Department of Energy under contract number DE-AC05-00OR22725.

Table 1. Digestion Solvents

| | **Solvent/Buffer*** | **Incubation Time**** |
|---|---|---|
| 1. | 50 mM Tris HCl 10 mM $CaCl_2$ | Overnight |
| 2. | 50 mM Tris HCl 10 mM $CaCl_2$ | One hour |
| 3. | 60% acetonitrile 40% 50 mM Tris HCl 10 mM $CaCl_2$ | One hour |
| 4. | 80% acetonitrile 20% 50 mM Tris HCl 10 mM $CaCl_2$ | One hour |
| 5. | 60% methanol 40% 50 mM Tris HCl 10 mM $CaCl_2$ | One hour |

* All buffers were pH 7.6

** All incubations were performed at 37°C.

Table 2. Peptide Identifications for Different Quantities of the Protein Standard Mixture

| | Total Protein Amount | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 0.2 µg | | 1 µg | | 2 µg | | 10 µg | |
| **Digestion Solvent** | **Peptides Identified** | **Tryptic**** | **Peptides Identified** | **Tryptic**** | **Peptides Identified** | **Tryptic**** | **Peptides Identified** | **Tryptic**** |
| **Aqueous overnight** | high = 105 | 90% | high = 130 | 82% | high = 148 | 68% | high = 138 | 62% |
| | 94 | | 114 | | 146 | | 137 | |
| | low = 84 | | low = 107 | | low = 127 | | low = 133 | |
| **Aqueous one hour** | high = 95 | 91% | high = 94 | 93% | high = 125 | 81% | high = 123 | 70% |
| | 70 | | 91 | | 111 | | 119 | |
| | low = 69 | | low = 87 | | low = 86 | | low = 112 | |
| **60% Acetonitrile** | high = 95 | 97% | high = 113 | 93% | high = 165 | 96% | high = 117 | 91% |
| | 80 | | 100 | | 134 | | 115 | |
| | low = 64 | | low = 99 | | low = 90 | | low = 111 | |
| **80% Acetonitrile** | high = 138 | 97% | high = 197 | 97% | high = 197 | 96% | high = 146 | 91% |
| | 132 | | 187 | | 157 | | 141 | |
| | low = 119 | | low = 169 | | low = 157 | | low = 140 | |
| **60% Methanol** | high = 67 | 97% | high = 134 | 96% | high = 137 | 97% | high = 120 | 91% |
| | 60 | | 107 | | 137 | | 118 | |
| | low = 55 | | low = 90 | | low = 129 | | low = 115 | |
| **Overnight using chaotrope ***** | N/A | | N/A | | N/A | | high = 41 | 68% |
| | | | | | | | 40 | |
| | | | | | | | low = 40 | |

*Equal masses of each protein were combined to give the corresponding quantities.

**Fully tryptic peptide percentages for highest value out of three replicates are shown.

***Digestion protocol included using guanidine HCl prior to digestion and reverse phase extraction prior to sample injection (see Figure 1).

N/A: Not analyzed.

Table 3. Semi-tryptic Peptides from the Overnight Aqueous Digest of a 1 μg Mixture

| Semi-tryptic peptide | Protein | Charge state(s) | Spectrum Count | X-corr |
|---|---|---|---|---|
| K.VGGHAAEYGAE.A | bovine hemoglobin α | +1 | 1 | 2.4 |
| L.SELSDLHAHK.L | bovine hemoglobin α | +2 | 1 | 2.6 |
| A.SHLPSDFTPAVHASLDK.F | bovine hemoglobin α | +3 | 1 | 4.3 |
| W.GKVEADIAGHGQEVLIR.L | horse myoglobin | +3 | 1 | 4.1 |
| K.GHHEAELKPL.A | horse myoglobin | +3, + 2 | 1, 1 | 3.6, 2.6 |
| K.YLEFISDAIIHVL.H | horse myoglobin | +2 | 1 | 4.8 |
| R.LLVVYPW.T | bovine hemoglobin β | +2 | 1 | 3.0 |
| F.GDLSTADAVMNNPK.V | bovine hemoglobin β | +2 | 2 | 4.6 |
| L.STADAVMNNPK.V | bovine hemoglobin β | +2 | 1 | 3.5 |
| K.LLGNVLVVVL.A | bovine hemoglobin β | +2 | 2 | 2.8 |
| K.VVAGVANAL.A | bovine hemoglobin β | +1 | 1 | 2.0 |
| K.AVVQDPALKPL.A* | bovine carbonic anhydrase | +2 | 1 | 3.3 |
| L.ALVYGEATSR.R | bovine carbonic anhydrase | +1, +2 | 1,2 | 2.1, 3.6 |
| W.IVLKEPISVSSQQMLK.F | bovine carbonic anhydrase | +3 | 1 | 3.8 |
| R.TLNFNAEGEPELLML.A | bovine carbonic anhydrase | +1, +2 | 1,2 | 2.9, 3.4 |
| L.ANWRPAQPLK.N* | bovine carbonic anhydrase | +2 | 1 | 2.8 |
| K.SANLMAGHWVAIS.G | yeast alcohol dehydrogenase I | +2 | 1 | 2.6 |
| L.GIDGGEGKEELFR.S* | yeast alcohol dehydrogenase I | +2 | 1 | 3.3 |
| N.GTTVLVGMPAGAK.C* | yeast alcohol dehydrogenase I | +2 | 2 | 3.8 |
| L.STLPEIYEK.M* | yeast alcohol dehydrogenase I | +1 | 1 | 1.9 |

*semi-tryptic peptides identified in multiple digestion data sets.

Table 4. Protein Sequence Coverage and Peptide Identifications of 1 μg Mixture

| | Aqueous Overnight | | Aqueous One Hour | | 60% $CH_3CN$ | | 80% $CH_3CN$ | | 60% $CH_3OH$ | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Protein** | **% seq. cov.** | **number of peptides** | **% seq. cov.** | **number of peptides** | **% seq. cov.** | **number of peptides** | **% seq. cov.** | **number of peptides** | **% seq. cov.** | **number of peptides** |
| yeast alcohol dehydrogenase I | 59 | 26 | 51 | 29 | 40 | 16 | 56 | 25 | 36 | 21 |
| yeast alcohol dehydrogenase II | 27 | 10 | 32 | 14 | 14 | 5 | 20 | 10 | 10 | 6 |
| bovine serum albumin | 23 | 16 | 0 | 0 | 19 | 9 | 70 | 61 | 25 | 17 |
| horse myoglobin | 75 | 24 | 50 | 11 | 70 | 25 | 94 | 30 | 98 | 27 |
| bovine hemoglobin α | 79 | 14 | 62 | 13 | 59 | 13 | 71 | 12 | 54 | 11 |
| bovine hemoglobin β | 73 | 22 | 72 | 17 | 77 | 19 | 62 | 20 | 76 | 23 |
| Chicken lysozyme | 38 | 5 | 22 | 2 | 86 | 20 | 86 | 17 | 95 | 16 |
| bovine carbonic anhydrase | 59 | 22 | 54 | 14 | 34 | 10 | 72 | 20 | 48 | 17 |

Table 5. Peptide and protein identifications for *R. palustris* 70S ribosomal proteins

| | Ribosomal Proteins | | | | GroEL* | |
|---|---|---|---|---|---|---|
| | 0.2 μg | | 1 μg | | | |
| **Digestion Solvent** | **peptide ids** | **ribosomal protein ids** | **peptide ids** | **ribosomal protein ids** | **peptide ids** | **protein ids** |
| Aqueous** | 22 | 9 | 102 | 31 | 57 | 2 |
| 80% Acetonitrile*** | 36 | 13 | 144 | 39 | 54 | 3 |

*We identified both gene products, GroEL1 and GroEL2, from each digestion method. The third identified protein from the GroEL digestion in 80% acetonitrile was apolipoprotein n-acyltransferase, with 2 peptides identified (4.3% sequence coverage).

**Overnight digestion in 50mM Tris buffer (pH 7.6), 10mM $CaCl_2$.

***One hour digestion in 80% acetonitrile.

## FIGURE CAPTIONS

**Figure 1.** Flowchart representing the solvent systems and digestion protocols used in this study.

**Figure 2.** An example MS-MS spectrum identified as a semi-tryptic peptide. This spectrum represents the doubly-charged F.GDLSTADAVMNPK.V, an internal peptide from hemoglobin's beta chain. (The periods in the peptide sequence represent protease cleavage points; the actual peptide detected corresponds to the sequence between the periods.) SEQUEST yielded an XCorr of 4.6, an unlikely score for a false positive.

**Figure 3.** MS1PeakFinder plots for parent ions in LC-MS-MS data of the 1 μg mixture. (a) overnight digestion in aqueous buffer, (b) one hour digestion in 80% acetonitrile. The dense patch of ions resulting from incompletely digested proteins eluting between 77 and 82 minutes is circled in figure 3a. The circled region in figure 3b indicates a more diffuse region at shorter retention times, corresponding to more complete digestion. Peaks for eluting peptides are red if a tandem mass spectrum resulted in a confident identification, orange if a tandem mass spectrum was acquired but no confident identification was obtained, and yellow if no tandem mass spectra were obtained for the ion.

**Figure 4.** LC-FTICR-MS total ion chromatograms of the 1 μg mixture. The blue trace is from overnight digestion in aqueous buffer; the red trace is from a one-hour digestion in 80% acetonitrile. The insets illustrate the isotopic resolution of nominal masses representing undigested myoglobin and hemoglobin α chain. Peptides were separated using identical chromatographic gradients for both the LC-FTICR-MS and LC-MS-MS measurements shown in Figure 3.

Figure 1.

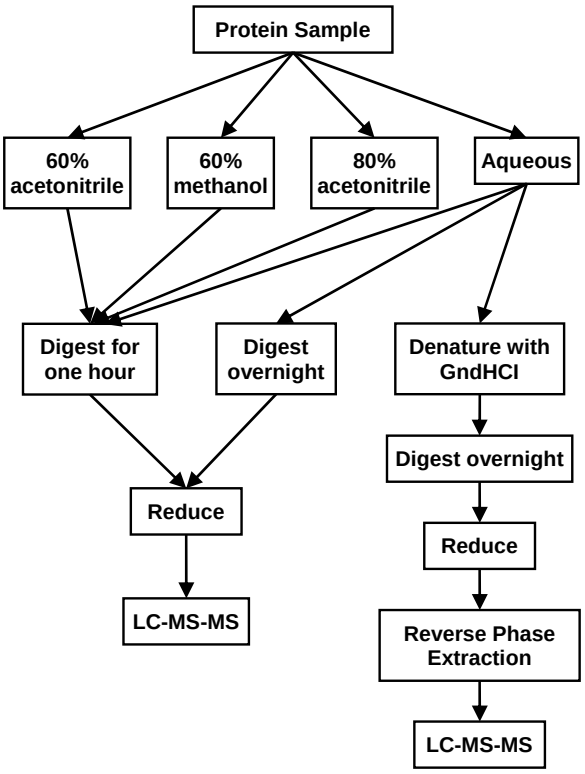

Protein Sample
60% acetonitrile
60% methanol
80% acetonitrile
Aqueous
Digest for one hour
Digest overnight
Denature with GndHCl
Digest overnight
Reduce
Reduce
LC-MS-MS
Reverse Phase Extraction
LC-MS-MS

Figure 2.

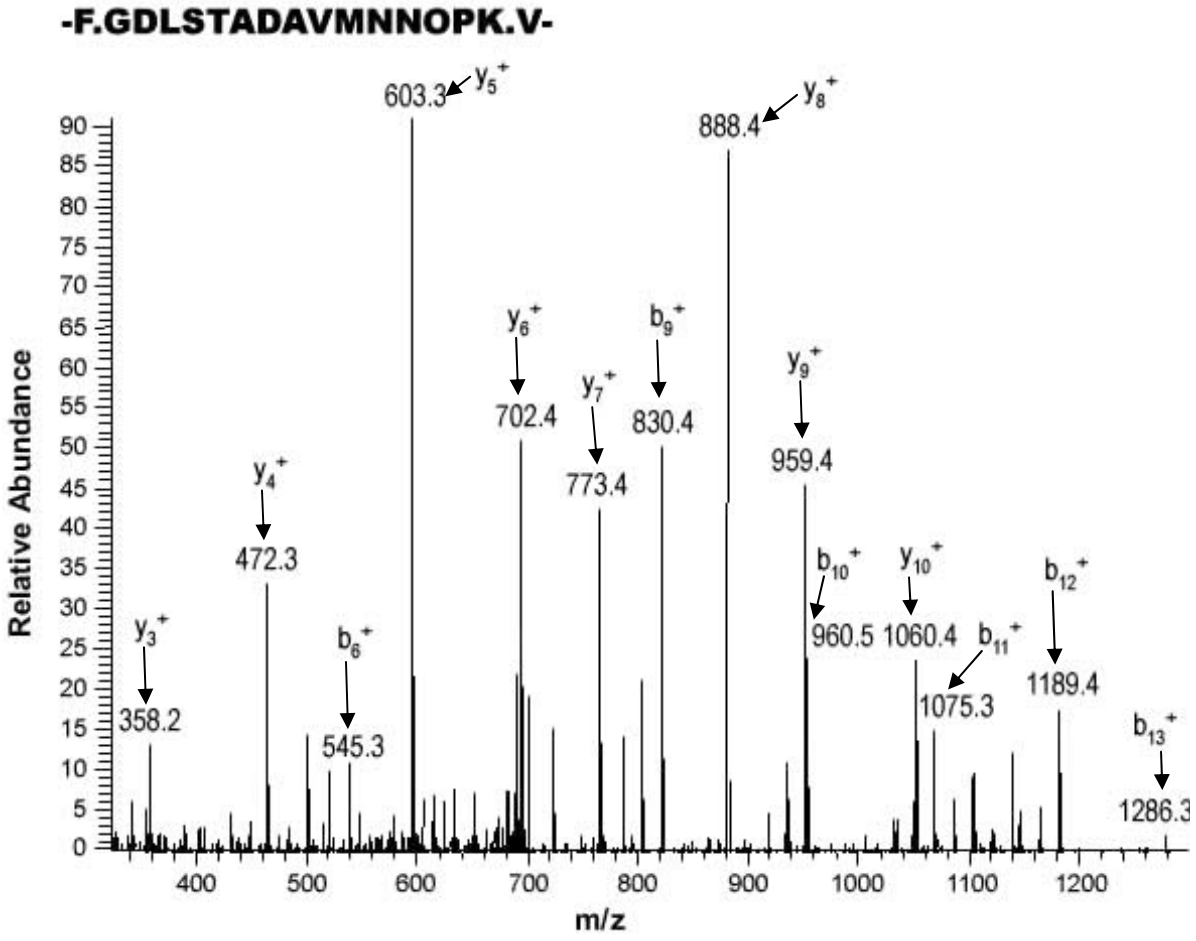

-F.GDLSTADAVMNNOPK.V-
Relative Abundance
m/z
603.3
y5+
888.4
y8+
y6+
702.4
b9+
830.4
y7+
773.4
y9+
959.4
y4+
472.3
y3+
358.2
b6+
545.3
b10+
960.5
y10+
1060.4
b11+
1075.3
b12+
1189.4
b13+
1286.3

Figure 3a

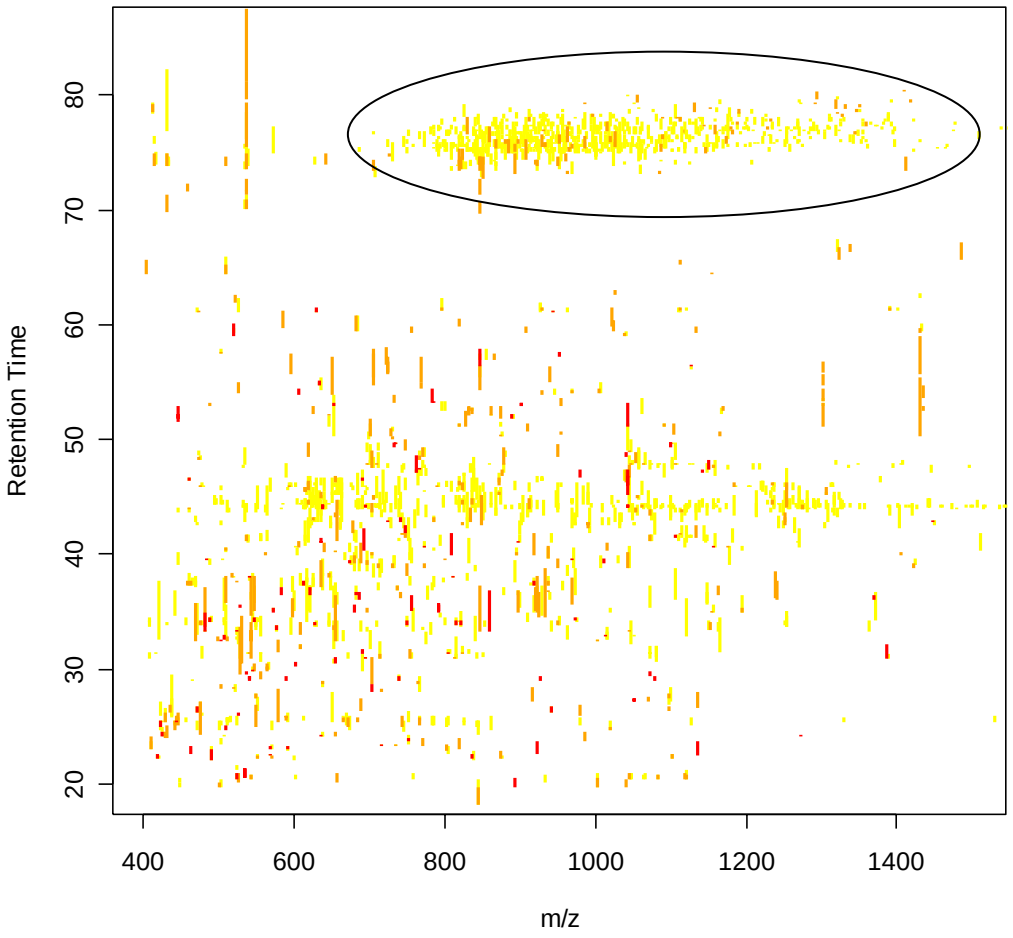

Retention Time
20
30
40
50
60
70
80
400
600
800
1000
1200
1400
m/z

Figure 3b.

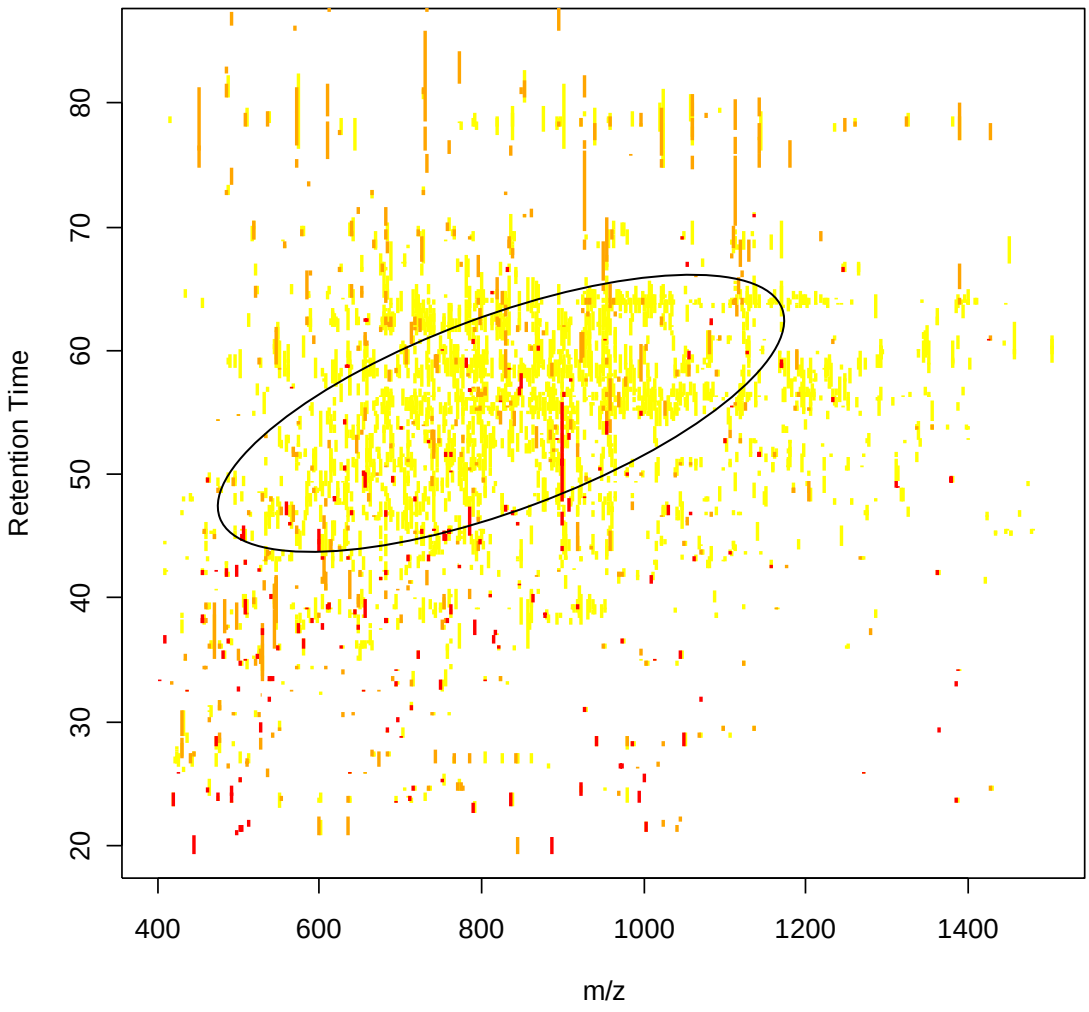
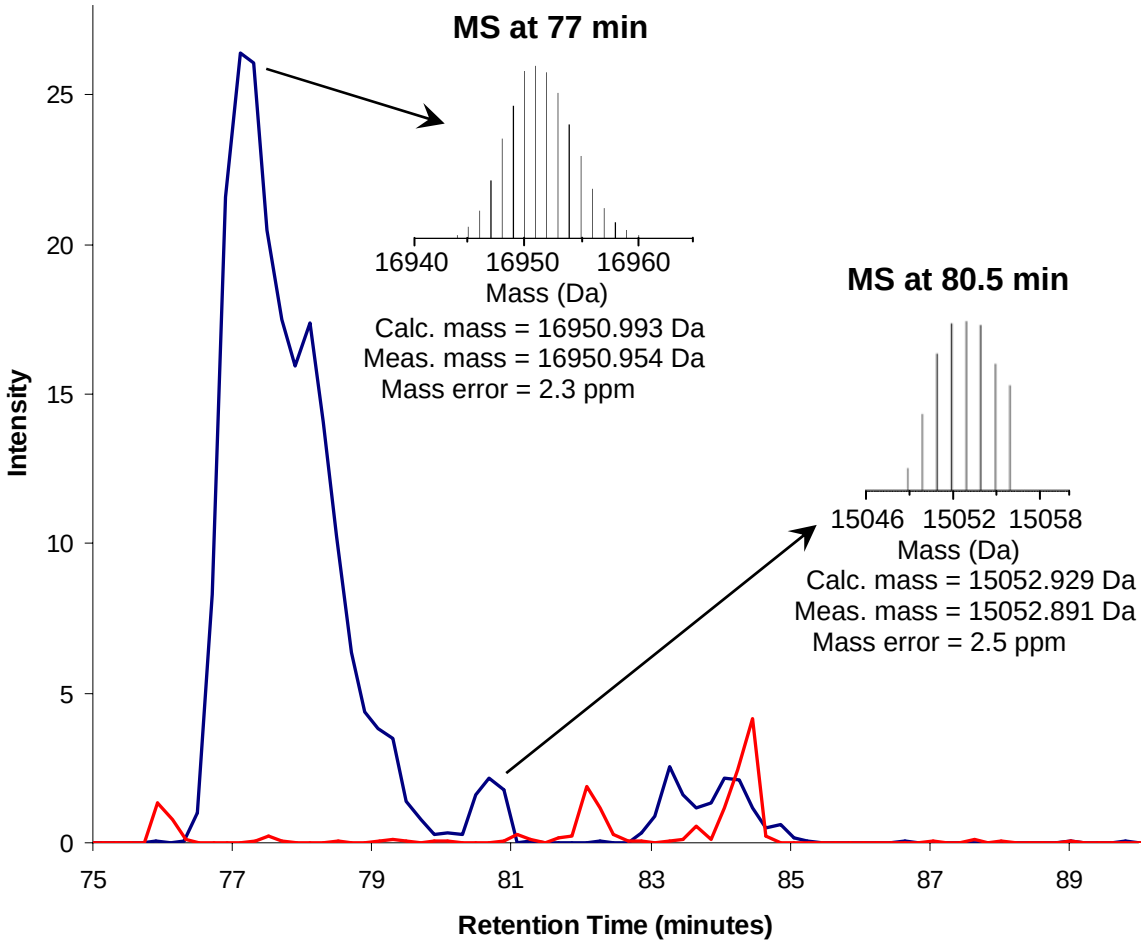

Retention Time
20
30
40
50
60
70
80
400
600
800
1000
1200
1400
m/z

Figure 4.


MS at 77 min
16940
16950
16960
Mass (Da)
Calc. mass = 16950.993 Da
Meas. mass = 16950.954 Da
Mass error = 2.3 ppm
MS at 80.5 min
15046
15052
15058
Mass (Da)
Calc. mass = 15052.929 Da
Meas. mass = 15052.891 Da
Mass error = 2.5 ppm
Intensity
0
5
10
15
20
25
75
77
79
81
83
85
87
89
Retention Time (minutes)

**REFERENCES**

(1) Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Science* **1989**, *246*, 64-71.

(2) Hillenkamp, F.; Karas, M.; Beavis, R. C.; Chait, B. T. *Anal. Chem.* **1991**, *63*, 1193A-1203A.

(3) Yates, J. R., 3rd; Eng, J. K.; McCormack, A. L.; Schieltz, D. *Anal. Chem.* **1995**, *67*, 1426-1436.

(4) Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S. *Electrophoresis* **1999**, *20*, 3551-3567.

(5) Tabb, D. L.; McDonald, W. H.; Yates, J. R., 3rd *J. Proteome Res.* **2002**, *1*, 21-26.

(6) McCormack, A. L.; Schieltz, D. M.; Goode, B.; Yang, S.; Barnes, G.; Drubin, D.; Yates, J. R., 3rd *Anal. Chem.* **1997**, *69*, 767-776.

(7) Wolters, D. A.; Washburn, M. P.; Yates, J. R., 3rd *Anal. Chem.* **2001**, *73*, 5683-5690.

(8) Hunt, D. F.; Yates, J. R., 3rd; Shabanowitz, J.; Winston, S.; Hauer, C. R. *Proc. Natl. Acad. Sci. U. S. A.* **1986**, *83*, 6233-6237.

(9) Link, A. J.; Eng, J.; Schieltz, D. M.; Carmack, E.; Mize, G. J.; Morris, D. R.; Garvik, B. M.; Yates, J. R., 3rd *Nat. Biotechnol.* **1999**, *17*, 676-682.

(10) Gavin, A. C.; Bosche, M.; Krause, R.; Grandi, P.; Marzioch, M.; Bauer, A.; Schultz, J.; Rick, J. M.; Michon, A. M.; Cruciat, C. M.; Remor, M.; Hofert, C.; Schelder, M.; Brajenovic, M.; Ruffner, H.; Merino, A.; Klein, K.; Hudak, M.; Dickson, D.; Rudi, T.; Gnau, V.; Bauch, A.;

Bastuck, S.; Huhse, B.; Leutwein, C.; Heurtier, M. A.; Copley, R. R.; Edelmann, A.; Querfurth, E.; Rybin, V.; Drewes, G.; Raida, M.; Bouwmeester, T.; Bork, P.; Seraphin, B.; Kuster, B.; Neubauer, G.; Superti-Furga, G. *Nature* **2002**, *415*, 141-147.

(11) Ho, Y.; Gruhler, A.; Heilbut, A.; Bader, G. D.; Moore, L.; Adams, S. L.; Millar, A.; Taylor, P.; Bennett, K.; Boutilier, K.; Yang, L.; Wolting, C.; Donaldson, I.; Schandorff, S.; Shewnarane, J.; Vo, M.; Taggart, J.; Goudreault, M.; Muskat, B.; Alfarano, C.; Dewar, D.; Lin, Z.; Michalickova, K.; Willems, A. R.; Sassi, H.; Nielsen, P. A.; Rasmussen, K. J.; Andersen, J. R.; Johansen, L. E.; Hansen, L. H.; Jespersen, H.; Podtelejnikov, A.; Nielsen, E.; Crawford, J.; Poulsen, V.; Sorensen, B. D.; Matthiesen, J.; Hendrickson, R. C.; Gleeson, F.; Pawson, T.; Moran, M. F.; Durocher, D.; Mann, M.; Hogue, C. W.; Figeys, D.; Tyers, M. *Nature* **2002**, *415*, 180-183.

(12) Butland, G.; Peregrin-Alvarez, J. M.; Li, J.; Yang, W.; Yang, X.; Canadien, V.; Starostine, A.; Richards, D.; Beattie, B.; Krogan, N.; Davey, M.; Parkinson, J.; Greenblatt, J.; Emili, A. *Nature* **2005**, *433*, 531-537.

(13) http://www.ornl.gov/sci/GenomestoLife/index.shtml

(14) Larimer, F. W.; Chain, P.; Hauser, L.; Lamerdin, J.; Malfatti, S.; Do, L.; Land, M. L.; Pelletier, D. A.; Beatty, J. T.; Lang, A. S.; Tabita, F. R.; Gibson, J. L.; Hanson, T. E.; Bobst, C.; Torres, J. L.; Peres, C.; Harrison, F. H.; Gibson, J.; Harwood, C. S. *Nat. Biotechnol.* **2004**, *22*, 55-61.

(15) Rigaut, G.; Shevchenko, A.; Rutz, B.; Wilm, M.; Mann, M.; Seraphin, B. *Nat. Biotechnol.* **1999**, *17*, 1030-1032.

(16) Gould, K. L.; Ren, L.; Feoktistova, A. S.; Jennings, J. L.; Link, A. J. *Methods* **2004**, *33*, 239-244.

(17) Fersht, A. *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*, Third ed.; W. H. Freeman and Company: New York, 2000, 103-131.

(18) Park, Z. Y.; Russell, D. H. *Anal. Chem.* **2000**, *72*, 2667-2670.

(19) Russell, W. K.; Park, Z. Y.; Russell, D. H. *Anal. Chem.* **2001**, *73*, 2682-2685.

(20) Ekstrom, S.; Onnerfjord, P.; Nilsson, J.; Bengtsson, M.; Laurell, T.; Marko-Varga, G. *Anal. Chem.* **2000**, *72*, 286-293.

(21) Doucette, A.; Craft, D.; Li, L. *Anal. Chem.* **2000**, *72*, 3355-3362.

(22) Riviere, R.; Fleming, M.; Elicone, C.; Tempst, P. *Techniques in Protein Chemistry II*, 1991.

(23) Ekstrom, S.; Malmstrom, J.; Wallman, L.; Lofgren, M.; Nilsson, J.; Laurell, T.; Marko-Varga, G. *Proteomics* **2002**, *2*, 413-421.

(24) Bosserhoff, A.; Wallach, J.; Frank, R. W. *J. Chromatogr.* **1989**, *473*, 71-77.

(25) Yu, Y. Q.; Gilar, M.; Gebler, J. C. *Rapid Commun. Mass Spectrom.* **2004**, *18*, 711-715.

(26) Griebenow, K.; Kilbanov, A. M. *J. Am. Chem. Soc.* **1996**, *118*, 11695-11700.

(27) Welinder, K. G. *Anal. Biochem.* **1988**, *174*, 54-64.

(28) Simon, L.; Laszlo, K.; Vertesi, A.; Bagi, K.; Szajani, B. *J. Mol. Catal. B: Enzym.* **1998**, *4*, 41-45.

(29) 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. *J. Proteome Res.* **2004**, *3*, 965-978.

(30) Hardy, S. J.; Kurland, C. G.; Voynow, P.; Mora, G. *Biochemistry* **1969**, *8*, 2897-2905.

(31) Puig, O.; Caspary, F.; Rigaut, G.; Rutz, B.; Bouveret, E.; Bragado-Nilsson, E.; Wilm, M.; Seraphin, B. *Methods* **2001**, *24*, 218-229.

(32) Geissler, J. F.; Harwood, C. S.; Gibson, J. *J. Bacteriol.* **1988**, *170*, 1709-1714.

(33) Tabb, D. L.; Narasimhan, C.; Strader, M. B.; Hettich, R. L. *Anal. Chem.* **2005**, *77*, 2464-2474.

(34) Eng, J. K.; McCormack, A. L.; Yates, J. R., 3rd *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 976-989.

(35) Hornik, K., The R FAQ., http://www.ci.tuwien.ac.at/~hornik/**R**/ ; ISBN 3-900051-08-9.

(36) Slysz, G. W.; Schriemer, D. C. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 1044-1050.

(37) Olsen, J. V.; Ong, S. E.; Mann, M. *Mol. Cell Proteomics* **2004**, *3*, 608-614.

(38) Simon, L.; Kotorman, M.; Garab, G.; Laczko, I. *Biochem. Biophys. Res. Commun.* **2001**, *280*, 1367-1371.

(39) Bogdanov, B.; Smith, R. D. *Mass Spectrom. Rev.* **2005**, *24*, 168-200.

(40) Buchanan, M. V.; Larimer, F. W.; Wiley, H. S.; Kennel, S. J.; Squier, T. J.; Ramsey, J. M.; Rodland, K. D.; Hurst, G. B.; Smith, R. D.; Xu, Y.; Dixon, D.; Doktycz, M. J.; Colson, S.; Gesteland, R.; Giometti, C.; Young, M.; Giddings, M. *Omics* **2002**, *6*, 287-303.

# Appendix B
# A General System for Studying Protein-Protein Interactions in Gram-Negative Bacteria

Pelletier and Hurst *et al.* *Journal of Proteome Research*, 2008, **7** (8), 3319–3328.
http://pubs.acs.org/doi/abs/10.1021/pr8001832

As a co-author of this publication, W. J. Hervey. IV performed microbial cell growth, protein isolation, sample preparation and collection of data on the LCQ Deca XP instrument for the *E. coli* tandem affinity isolation (TAP) clones.

A General System for Studying Protein-Protein Interactions in Gram-Negative Bacteria

Dale A. Pelletier[1,*,+], Gregory B. Hurst[2,+], Linda J. Foote[1], Patricia K. Lankford[1], Catherine K. McKeown[1], Tse-Yuan Lu[1], Denise D. Schmoyer[3], Manesh B. Shah[1], W. Judson Hervey IV[2,8], W. Hayes McDonald[2], Brian S. Hooker[5], William R. Cannon[6], Don S. Daly[6], Jason M. Gilmore[6], H. Steven Wiley[7], Deanna L. Auberry[5], Yisong Wang[1], Frank W. Larimer[1], Stephen J. Kennel[1,**], Mitchel J. Doktycz[1], Jennifer L. Morrell-Falvey[1], Elizabeth T. Owens[1], Michelle V. Buchanan[4]

[1]Biosciences Division, [2]Chemical Sciences Division, [3]Computer Science and Mathematics Division, and [4]Physical Sciences Directorate, Oak Ridge National Laboratory, Oak Ridge TN 37831

[5]Biological Sciences Division, [6]Computational Sciences and Mathematics Division, [7]Environmental and Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland WA 99352

[8]UT-ORNL Genome Science and Technology Program, Oak Ridge TN 37831

**Present Address: University of Tennessee Medical Center, Knoxville TN 37920

[+]These two authors contributed equally to the writing of this manuscript.

**Keywords**: protein interactions, protein complexes, RNA polymerase, affinity purification, mass spectrometry

*Corresponding author:
E-mail: dpelletier@ornl.gov
Phone: 865-576-2857
Fax: 865-574-6210

**Abstract**

One of the most promising methods for large-scale studies of protein interactions is isolation of an affinity-tagged protein with its *in vivo* interaction partners, followed by mass spectrometric identification of the co-purified proteins. Previous studies have generated affinity-tagged proteins using genetic tools or cloning systems that are specific to a particular organism. To enable protein-protein interaction studies across a wider range of Gram-negative bacteria, we have developed a methodology based on expression of affinity-tagged "bait" proteins from a medium copy-number plasmid. This construct is based on a broad-host-range vector backbone (pBBR1MCS5). The vector has been modified to incorporate the Gateway DEST vector recombination region, to facilitate cloning and expression of fusion proteins bearing a variety of affinity, fluorescent, or other tags. We demonstrate this methodology by characterizing interactions among subunits of the DNA-dependent RNA polymerase complex in two metabolically versatile Gram-negative microbial species of environmental interest, *Rhodopseudomonas palustris* CGA010 and *Shewanella oneidensis MR-1*. Results compared favorably with those for both plasmid- and chromosomally-encoded affinity-tagged fusion proteins expressed in a model organism, *E. coli*.

## Introduction

Most polypeptide chains function as components of larger protein complexes or "molecular machines" [1-3]. Living organisms gain several advantages from a strategy of assembling machines from smaller components, such as a smaller probability of translation error for small versus large proteins, the increased flexibility and coding economy when using a single subunit protein in several different complexes, more efficient channeling of enzyme substrates, and the ability to modulate the activity or function of a complex by altering its subunit composition [1-4]. Discovering the protein-protein interactions that make up these complexes is an important technique for studying protein function. If a previously unknown protein is observed to interact with other proteins or a complex of known function, then one has gained insight into the function of the unknown [5]. Genome-wide elucidation of protein-protein interactions yields a network structure that may result in higher-level information about interactions within and among protein complexes [6-11].

Various approaches allow identification of physical protein-protein interactions in complexes [3, 12, 13]. Classical "one-at-a-time" experimental approaches focus on proteins that can be linked with a measured activity. Newer techniques seek to elucidate larger sets of interactions that involve a large fraction of the proteome, generally with higher speed, but less rigor, than classical methods. These higher-throughput techniques include the yeast two-hybrid technique [14-17], arrays of immobilized proteins [18, 19], and phage display [20]. Mass spectrometry-based measurements of protein-protein interactions across the proteome in *Saccharomyces cerevisiae* [21, 22] and in *E. coli* [23, 24] have been reported. Desirable attributes of the latter approach include direct detection of proteins (as opposed to *m*RNA, for example), additional experimental tools

for reducing false positive interactions [25], and the potential to study post-translational modifications.

It would be desirable to extend application of a mass spectrometry-based approach for discovery of protein-protein interactions to a wider variety of environmentally relevant Gram-negative bacterial species, expecially those of potential interest for energy production, bioremediation, or carbon sequestration [26]. To address this need, we have developed a straightforward approach that is general to Gram-negative bacteria, and flexible enough to accommodate different affinity tags without the need to amplify and re-clone all genes of interest from genomic DNA. Our goal is a system for efficient *in vivo* expression of a range of affinity tagged fusion proteins across a variety of species, particularly in those strains in which chromosomal integration is difficult or impossible. This report describes the construction and testing of a vector based on pBBR1MCS5, which has been demonstrated to replicate in a number of Gram-negative bacteria [27]. We have modified this vector to incorporate the Gateway™ cloning system [28], which allows facile introduction of a library of cloned genes (entry vectors) into constructs encoding several different affinity (or other) tags (destination vectors). Using the RNA polymerase complex as a test case, we evaluated the performance of the plasmid-encoded hexahistidine/V5 epitope fusion proteins in *R. palustris* [29], *S. oneidensis* MR-1 [30], and *E. coli* [31], and compared this tagging and purification strategy with the chromosome-encoded tandem affinity purification (TAP) tag [32] in *E. coli*.

**Materials and Methods**

*Strains, plasmids and media*

Bacterial strains and plasmids used in this study are listed in Table 1. *Rhodopseudomonas palustris* CGA010 [33] and pBBR1MCS5 broad-host-range vector [27] were gifts from Caroline Harwood, University of Washington-Seattle. *Shewanella oneidensis* MR1 was obtained from ATCC (Manassas, VA). The *E. coli* strains expressing proteins from the tandem affinity purification (TAP) vector [32] integrated into the *E. coli* chromosome were a gift from Andrew Emili, University of Toronto. *E. coli* and *S. oneidensis* strains were cultured on Luria-Bertani (LB) broth or LB agar plates for cloning and strain maintenance. For affinity isolation experiments *E. coli* was cultured on M9 media and *S. oneidensis* was grown in a defined mineral medium [34]. *R. palustris* was grown anaerobically in the light at 30°C in a defined mineral medium containing 10 mM succinate (PMS-10) [35]. Plasmid-containing strains were routinely grown with antibiotics at the following concentrations: ampicillin, 100 μg/ml; chloramphenicol, 30 μg/ml; and gentamycin, 10 μg/ml for *E. coli* and *S. oneidensis* strains, and 100 μg/ml for *R. palustris* strains. The growth medium for *E. coli* B2155 was supplemented with 100 μg/ml diaminopimelic acid (DAP).

*Construction of pBBR5DEST42 broad-host-range destination vector*

Isolated pBBR1MCS5 plasmid DNA was digested with KpnI and PvuI, phenol-extracted, treated with calf intestine alkaline phosphatase, and a 4.5 kb fragment was purified on an agarose gel. The recombination region on pET-DEST42 vector DNA (Invitrogen, Carlsbad, CA) was PCR amplified using the following primers that include KpnI and PvuI restriction sites, respectively: 42F: 5'-cag cgg tac ctc tcg atc ccg cga aat taa ta-3' and 42R: 5'-ctg tcg atc gct gta ggc ata ggc ttg gtt atg-3'. The PCR reactions were done in 100 μl volumes with PfuTurbo DNA polymerase (Stratagene, La Jolla, CA) and 50 ng of pET-DEST42 DNA. PCR products were analyzed on an

agarose gel to verify that the products were the correct size. The PCR product was digested with KpnI and PvuI and agarose gel-purified. 60 ng of PCR product was ligated with 30 ng of 4.5 kb pBBR1MCS5 KpnI/PvuI fragment following standard protocols [36]. Ligation products were transformed into *E. coli* DB3.1 chemically competent cells (Invitrogen) and transformants were selected by plating on LB agar plates containing appropriate antibiotics. Individual colonies were grown overnight in LB containing 30 μg/ml chloramphenicol and 10 μg/ml gentamycin, and plasmid DNA was prepared using QIAprep spin miniprep following the manufacturer's protocol (Qiagen, Valencia, CA). Plasmid DNA was digested with KpnI and PvuI and digestion products were analyzed on an agarose gel to confirm the presence of products of the expected sizes.

*Cloning of Target Genes into Gateway Entry Vector*

Genes of interest were PCR-amplified from *R. palustris* CGA010, *S. oneidensis* MR1 or *E. coli* K12 MG1655 genomic DNA with Expand DNA polymerase (Roche Diagnostics, Indianapolis, IN), and appropriate primers (Table 1). The primers contain appropriate *attB* recombination sequences, Shine-Delgarno sequence, and the ATG start codon; the transcriptional stop codon was removed to allow C-terminal fusion of affinity tag sequences. PCR products were analyzed on an agarose gel and purified with a QIAquick PCR purification kit (Qiagen). PCR product was recombined with pDONR221 DNA using BP Clonase II enzyme mix following the manufacturer's protocol (Invitrogen), and then transformed into chemically competent DH5α cells and plated onto LB with appropriate antibiotic selection. The inserts were confirmed by sequencing using M13 forward and reverse primers (Laragen, Inc; Los Angeles, CA).

*Creation of Expression Clones from Gateway Entry Clones*

Entry clones in the pDONR221 vector were recombined with the pBBR5DEST42 destination vector using LR Clonase II enzyme mix following the manufacturer's protocol (Invitrogen), transformed into BL21(DE3) chemically competent cells, and plated onto selective media. Individual colonies were grown overnight in LB containing gentamycin, and plasmid DNA was purified with a Qiaprep miniprep kit (Qiagen).

*Plasmid transformation of R. palustris and S. oneidensis*

To obtain electrocompetent cells of *R. palustris* CGA010, 50 ml cultures were grown anaerobically in the light at 30°C in PMS-10 to OD of 0.4 at 660 nm. Cells were recovered by centrifugation at 10,000 rpm for 10 min. at 4°C. Three cycles of washing and resuspension in 50 ml of ice cold sterile distilled water were performed followed by final resuspension in 1 ml of chilled 10% glycerol. Aliquots were stored at -80$^{o}$C until needed. For optimal transformation 40 µl of competent cells, along with 1 to 10 ng of plasmid DNA in 80 µl of chilled distilled water, were added to 2 mm cuvettes of a Gene Pulser (Bio-Rad Laboratories, Hercules, CA). Electroporation was performed using the following settings: voltage 2.5 kV; capacitance 25 µF; resistance 100 Ω. After electroporation 1 ml PMS-10 was added and cells were incubated at 30°C for 20 hr. Transformants were selected by plating on PMS-10 agar plates containing 100µg/ml gentamycin and incubated anaerobically in the light at 30°C. For transformation of *S. oneidensis*, purified plasmids were first transformed into chemically competent *E. coli* β-2155 [37]. Plasmids were then transferred from *E. coli* β-2155 to *S. oneidensis* MR-1 by conjugation and transformants were selected on LB agar plates containing appropriate antibiotic.

*Western blots*

To test for protein expression, individual colonies of expression strains were grown in appropriate liquid media as described previously to mid-log phase and cells were harvested by centrifugation. Cell pellets were lysed in phosphate buffered saline (PBS) with 1 µl/ml BugBuster plus Benzonase Nuclease (Novagen, Madison, WI), 10 µg/ml lysozyme, 100 µg/ml phenylmethylsulfonylfluoride (PMSF) and 10 µg/ml leupeptin. After mixing gently for 20 min. at room temperature, lysates were centrifuged at 16,000 X g for 15 min at 4°C. Cleared lysates were separated on 10% SDS-PAGE and transferred to Immobilon P membrane (Millipore, Billerica, MA) in tris-glycine-MeOH buffer for 1 hour at 200 mA. Membranes were washed overnight at 4°C in PBS containing 0.1% Tween 20 (PBS-Tween). Non-specific binding was blocked by incubation of the membrane with gentle rocking for 1 hour in 5% non-fat milk powder, 10% bovine serum in PBS-Tween ("Blotto"). The membrane was then incubated for 2 hours with rocking at room temperature with the primary antibody, mouse anti-V5 epitope (Invitrogen) at 1/5000 dilution in Blotto. After extensive washing in PBS-Tween, the membrane was incubated with rocking for 1.5 hr in horseradish peroxidase-conjugated goat anti-mouse IgG (Bio-Rad, Hercules, CA) at 1/1000 dilution. After washing, color was developed by the addition of ImmunoPure metal enhanced DAB substrate (Pierce, Rockford, IL).

*Affinity isolation*

Cultures of 0.8 L of each bacterial strain were grown in appropriate media to mid-logarithmic phase and cells were harvested by centrifugation and pellets frozen at -80$^{o}$C until needed. Frozen cell pellets were thawed, and cells were lysed using 5 volumes lysis buffer (w/v), prepared with 10X BugBuster™ extraction reagent (Novagen) diluted to 1X in Ni-NTA binding buffer (50 mM $NaH_2PO_4$, 300 mM NaCl, 10 mM imidazole, pH 8), containing 1 µL/mL

benzonase nuclease reagent (Novagen), lysozyme (5 kU/g cell paste), 100 μg/mL PMSF and 10 μg/mL leupeptin. Cells were suspended in the lysis buffer, and incubated with end-over-end rotation at room temperature for 20 min. Lysates were centrifuged for 30 min at 12,000 x *g* for 20 min. The supernatant was transferred to a fresh tube, centrifuged for 15 min at 20,000 x *g* at 4°C, and the resulting supernatant (lysate) was used for subsequent purification steps.

The lysate was added to 200 μL of a 50% suspension of Ni-NTA agarose beads (Invitrogen), which had previously been washed 3-4 times with Ni-NTA binding buffer. The lysate and Ni-NTA beads were incubated with end-over-end tube rotation for 1 hour at room temperature, and then centrifuged at 100 x g for 30s and the supernatant removed. The beads were washed 3 times with 1 mL wash buffer (50 mM $NaH_2PO_4$, 300 mM NaCl, 20 mM imidazole, 100 μg/mL PMSF, 10 μg/mL leupeptin). Proteins were eluted from the beads using four 50 μL aliquots of elution buffer (50 mM $NaH_2PO_4$, 300 mM NaCl, 500 mM imidazole, 100 μg/mL PMSF, 10 μg/mL leupeptin). The eluates were combined and 400 μL water containing 100 μg/mL PMSF and 10 μg/mL leupeptin was added. After a final short clearing spin to pellet any remaining beads, ~90% of the supernatant was transferred to a clean tube.

The second purification step was performed using mouse anti-V5 monoclonal antibody agarose beads (Sigma). Eluate from the Ni-NTA purification was added to 100 μL of 50% slurry of anti-V5 beads, which had previously been washed 4 times with PBS. Sample and beads were incubated for 1 hour with end-over-end rotation of the tube at room temperature. Three cycles of brief centrifugation at 100 x g for 30s followed by addition of 1 mL PBS, gentle shaking, and removal of the supernatant, were performed. Proteins were eluted from the anti-V5 beads with three 50 μL volumes of aqueous 80% acetonitrile containing 1% formic acid. After

centrifugation at 100 x g for 30s to pellet any remaining beads, supernatant was transferred to a fresh tube.

For TAP affinity isolations, *E. coli* cellular lysates were incubated with IgG (Amersham/GE) affinity resin for 1 hour. Elution from the IgG resin was performed by treatment with AcTEV protease (Invitrogen) for 1 hour. A second 1 hr incubation with calmodulin affinity resin (Amersham/GE) further isolated the affinity-tagged proteins. Proteins were eluted from the calmodulin affinity resin as described by Zeghouf *et al*. [38].

*Trypsin digestion*

Proteins eluted from affinity resins were dried by vaccum centrifugation, then reconstituted in 150 μL of 50 mM Tris, 10 mM $CaCl_2$ (pH 7.6) buffer. One μg trypsin (sequencing grade; Promega) was added and the digestion allowed to proceed overnight at 37°C. The sample was acidified with 50 μL of 2.5% trifluoroacetic acid (TFA). Two 65 μL volumes were aliquoted into separate tubes for desalting via solid-phase extraction (OMIX C18, 100 μg size, Varian) using the manufacturer's protocol. Peptides were eluted from the OMIX tip in 100 μL aqueous 50% acetonitrile, 0.1% TFA. A 100 μL portion of aqueous 0.1% TFA was added, and the sample volume was decreased by approximately one half using vacuum centrifugation. Particulates were removed by centrifugation for 15 min at 10,600 x g before transfer to autosampler vials.

*Mass spectrometry and protein identifications*

The resulting peptides were analyzed by automated LC-ESI-MS/MS as reported previously in greater detail [39]. Briefly, an LCPackings (Sunnyvale CA) chromatographic system (Famos autosampler, Switchos flow switching module, and Ultimate HPLC) was interfaced with a ThermoFinnigan (San Jose CA) DecaXP Plus quadrupole ion trap mass spectrometer via a nanospray source. After injection via the autosampler, peptides were concentrated onto a reverse-phase preconcentration column and washed to remove excess salt. The peptides were then eluted onto a 75 μm ID C18 reverse-phase analytical column. Peptides were eluted from the analytical column using a gradient from 95% water/5% acetonitrile to 30% water/70% acetonitrile (with 0.1% formic acid in all mobile phases). Blank chromatographic runs between samples minimized carryover. The mass spectrometer was operated in a data-dependent MS/MS mode, with up to four peptides automatically selected per MS scan for MS/MS analysis. Proteins interacting with the bait proteins were identified from MS/MS spectra of their tryptic peptides via Sequest [40]. The resulting peptide identifications and other relevant statistics were filtered and collated by protein using DTASelect [41]. Peptide identifications required standard minimum XCorr values of ≥1.8 for singly-charged, ≥2.5 for doubly-charged, or ≥3.5 for triply-charged ions, and DeltaCN ≥0.08. Proteins were initially identified based on detection of one or more peptides and subsequently filtered at later stages of data analysis. Sequest searches were performed against separate FASTA protein databases for the three bacterial species, compiled as follows. Sequences for 4833 *R. palustris* proteins were obtained from the ORNL microbial genome annotation pipeline at http://www.ornl.gov/sci/GenomestoLife/index.shtml [29], with an additional 44 common contaminants and standard proteins. The *E. coli* protein sequence file contained 4337 proteins [42], plus their sequence-reversed counterparts and 188 common

contaminants and standard proteins. The *S. oneidensi*s MR-1 file contained 4898 *S. oneidensis* protein sequences [43], plus 44 common contaminants and standard proteins.

*Evaluation of protein-protein interactions*

The BEPro$^3$ algorithm [44] was applied for statistical evalution of observed bait-prey interactions. Two BePro$^3$ results were used: The *Bayes Odds*, a score associated with each bait-prey pair, ranging from 0 (low likelihood of interaction) to 1 (high likelihood of interaction); and the *Ubiquity*, a measure of the tendency of a given prey protein to be detected in association with a large variety of baits (0 indicates a "selective" prey, and 1 indicates a "sticky" prey.) For the *R. palustris* proteins, the data set included 77 LC-MS/MS measurements using 18 different affinity-tagged baits [44]. The *S. oneidensis* data set encompassed 84 LC-MS/MS measurements of affinity isolations obtained using 29 different affinity-tagged baits. BEPro$^3$ calculations were not performed for *E. coli* results due to the small size of the data set.

**Results**

Figure 1 outlines the approach tested in this paper for systematically identifying interactions among proteins across a range of bacterial species. The major steps are PCR amplification of genes of interest, recombination of each gene into a Gateway entry vector, recombination of each entry clone into a pBBR5DEST42 destination vector, transformation of host cells, cell culture and harvest, affinity isolation, mass spectrometric protein identification, and data analysis.

To demonstrate our system, we focused on the protein complex RNA polymerase, which has been suggested as a standard for evaluating methods for measuring protein-protein interactions

[45]. DNA-directed RNA polymerase uses DNA as the template for RNA synthesis during transcription. The "core" bacterial RNA polymerase contains two copies of the α subunit (RpoA) along with β (RpoB), β' (RpoC), and ω (RpoZ in *E. coli*; RnpO in *R. palustris*) subunits. This core can perform transcription of DNA into RNA, but has little specificity. Incorporation of additional protomers, the sigma factors, which join the complex depending on the cell's stress or environment [46], provides specificity in recognition of promoter regions on DNA. As described below, detection of the core components, sigma factors and other proteins known to interact with RNA polymerase provide a basis for evaluating our approach.

*Vectors coding for Affinity-Tagged Proteins*

Figure 2 shows the plasmid map for the pBBR5DEST42 destination vector, created by insertion of pET-DEST42 recombination region into the vector backbone of pBBR1MCS5. This new vector combines the convenience of facile introduction of genes of interest using the Gateway site-specific recombinase system with the medium copy number and broad-host-range conferred by the pBBR5MCS origin of replication. The plasmid also encodes proteins necessary for mobilization of the plasmid by RK2 conjugal transfer from an appropriate host strain, the *lac* and T7 promoter sequences for expression in *E. coli* or other appropriate host strains, and a carboxy-terminal hexahistidine/V5 epitope fusion for affinity isolation. To evaluate the ability of this vector to stably replicate and drive expression we introduced destination plasmids bearing the *rpoA* gene from three phylogenetically diverse bacterial species, the alpha proteobacterium *R. palustris*, the gamma-proteobacteria *S. oneidensis* and *E. coli*, and subsequently transformed the respective vectors into native host strains. Expression of plasmid-encoded fusion proteins was confirmed by Western blot analysis. Figure 3 demonstrates expression of hexahistidine/V5-

tagged RpoA from *E. coli* containing pBBR5DEST42-ECK3282 (Lane 2), from *R. palustris* cells containing pBBR5DEST42-RPA3226 (Lane3) and from *S. oneidensis* cells containing pBBR5DEST42-SO0256 (Lane 4).

*Comparison of chromosomal and plasmid-encoded SPA affinity tags in E. coli*

Table 2 compares results from several different affinity isolation strategies using the α subunit, RpoA, of RNA polymerase as the tagged protein. The data presented include results from large-scale protein-protein interaction studies in *E. coli* using the hexahistidine tag [23], or a dual affinity tag (SPA or TAP) [24], as well as results obtained in the present study using the TAP-tagged chromosomal insert [24] and the plasmid-encoded hexahistidine/V5-tagged proteins as baits. Results from the present study are presented in Table 2 as the number of MS-MS spectra assigned to a particular protein (the "spectrum count"), which is a crude indicator of the amount of that protein in the sample [47]. For comparison, Table 2 also includes predictions from STRING [48] (http://string.embl.de), and from Prolinks (http://mysql5.mbi.ucla.edu/cgi-bin/functionator/pronav) [49]. The data resulting from the four different experimental approaches shown in Table 2 detected RNA polymerase core components β and β' (RpoB and RpoC), as well as the most common sigma factor, $\sigma^{70}$ (RpoD). Results from both tags used in the present study, as well as literature results obtained using the TAP or SPA tag [24], include the ω subunit (RpoZ); RpoZ was not reported using the hexahistidine tag alone [23]. Using the TAP or SPA tag, previous studies reported detection of both NusA and NusG [24], which are known to interact with RNA polymerase [50]. Using our detection system with the chromosomally encoded TAP-tagged strain [24], we detected NusA but not NusG. Neither NusA nor NusG was detected using the hexahistidine tag alone [23] or the hexahistidine/V5 tag (this study). Among other proteins in

Table 2 known to associate with RNA polymerase, and detected in the published studies [23, 24] (HepA, GreB [51], and Rho [52]), we detected Rho with the hexahistidine/V5 tag, but not with the TAP tag. All four approaches led to detection of ribosomal proteins RplB, RpsE, and RpsD; several other ribosomal proteins were detected using one or more approaches. All four approaches also led to detection of several proteins that are not obviously related to RNA polymerase (see also footnotes b and c, Table 2).

*RNA polymerase in S. oneidensis*

Table 3 summarizes protein-protein interactions determined using RpoA, RpoC and RpoZ expressed as affinity-tagged fusions from the pBBR5DEST42 plasmid in *S. oneidensis*. A prey protein is listed in Table 3 only if it fulfills all of the following conditions for at least one bait: BePro[3] *Ubiquity* ≤ 0.5, BePro[3] *Bayes Odds* ≥0.5, with two or more peptides identified. A more complete listing is provided in Supplementary Table I. Bold entries in Table 3 identify bait-prey interactions with BePro[3] *Bayes Odds* ≥0.5. Multiple bold entries among the core components of the RNA polymerase enzyme indicate the specificity of the measurement. No sigma factors were detected with these baits for *S. oneidensis*. The RpoA bait showed the most interactors, including several that are not known components of RNA polymerase, but are abundant proteins in *E. coli* (ribosomal proteins, β subunit of ATP synthase F1). It is noteworthy that the average spectrum count for the RpoA bait protein itself was, by a large margin, the highest among the baits in Table 3, suggesting strong preferential over-expression of this protein from the plasmid. Unlike the other affinity-tagged components of RNA polymerase that were used as bait in *S. oneidensis*, RpoB was not detected when used as bait, based on duplicate LC-MS-MS measurements from a single culture. However, this protein was detected as a prey in

experiments where RpoA, RpoC, or RpoZ was the affinity-tagged bait. While Table 3 lists the more obvious interactors with the RNA polymerase core components, Supplementary Table I contains evidence for additional interactions. For example, prey proteins that were observed with the RpoA bait, but did not meet all the criteria for inclusion in Table 3, include NusA and NusG, which are known interactors of RNA polymerase [50].

*RNA polymerase in R. palustris*

Table 4 summarizes protein-protein interactions for DNA-dependent RNA polymerase in *R. palustris*, using α (RpoA, RPA3226), β (RpoB, RPA3268), and β' (RpoC, RPA3267) subunits of the core RNA polymerase enzyme and the heat shock sigma factor $\sigma^{32}$ (RpoH, RPA0367) expressed as affinity-tagged fusions from the pBBR5DEST42 plasmid. A prey protein was listed in Table 4 only if it fulfilled the conditions described above for Table 3; a more complete listing is provided in Supplementary Table II. Bold entries in the table identify bait-prey interactions with BePro[3] *Bayes Odds* ≥0.5. As noted above for *S. oneidensis*, multiple bold entries suggest robust detection of interactions among the core components of the enzyme. Although the ω subunit of the core complex from this organism, RnpO, was not successfully cloned and expressed as an affinity-tagged fusion, this protein was nonetheless detected in experiments using the other RNA polymerase components as baits. The major sigma factor ($\sigma^{70}$, RpoD) was detected using several baits. Using $\sigma^{32}$ as affinity-tagged bait led to detection of several core components. Also interacting with $\sigma^{32}$ were heat shock proteins GrpE (RPA0331, Table 3) and, less strongly, DnaK (RPA0333, Supplementary Table II); these two proteins have been reported to associate with $\sigma^{32}$ in *E. coli* [53]. A 14.5 kDa acidic conserved protein of unknown function, RPA0060, was detected fairly robustly with the RpoA bait, suggesting a previously unreported

association with the RNA polymerase complex. RPA0060 was also observed, albeit weakly, as a prey for both the RpoC and RpoB baits. Several abundant proteins not necessarily associated with RNA polymerase were observed, including ribosomal proteins, aconitase, and ATP synthase subunits.

**Discussion**

Microbes inhabit a diverse array of habitats, deriving energy from a variety of sources and biochemicals from a wide range of raw materials. The protein complexes and biochemical pathways that microbes have evolved over countless iterations in nature represent an amazing range of activities and strategies for bioremediation and energy production [26]. To gain access to this potentially vast resource, it will be necessary first to develop faster and more accurate methods, such as that described in this report, for studying the protein-protein interactions involved in these complexes and pathways.

To enhance our ability to discover such novel and potentially useful biochemical capabilities in microbes, we have developed a strategy to extend studies of protein-protein interactions to a larger subset of bacterial species. The basis of this strategy is the pBBR5DEST42 vector, which can be introduced into a broad range of Gram-negative bacterial hosts (Figure 2). In addition to broad species applicability, the strategy facilitates expression of a variety of fusion protein types. Because the approach is based on the Gateway system, each protein of interest can be "decorated" with new affinity or other types of tags simply by constructing a new destination vector; each gene of interest need only be cloned and verified once. This system is particularly useful in species where transformation is feasible, but chromosomal integration is not. We

demonstrated that pBBR5DEST42 could be used to express fusion proteins in *R. palustris*, *S. oneidensis* and *E. coli* (Figure 3, Table 2-4).

Initial large-scale studies of protein-protein interactions in yeast used different approaches for producing affinity-tagged protein. In one study, engineered genes encoding each protein of interest and an affinity tag were introduced into a chromosome by homologous recombination [21], while another group employed heterologous expression, with genes for affinity-tagged proteins introduced on plasmids [22]. It has been suggested that the former approach may result in fewer undetected interactions because protein expression is under the control of native promoters and at chromosomal copy number [45]. For this reason, we compared overall results from plasmid encoded versus chromosomally encoded affinity-tagged RpoA expressed in *E. coli*. To evaluate the performance of the pBBR5DEST42 vector in the context of a strategy that also includes affinity isolation and mass spectrometric identification of interacting proteins, we compared results in *E. coli* from the pBBR5DEST42-encoded fusion of RpoA with the dual hexahistidine/V5 tag to the chromosomally encoded TAP-RpoA [24], and a plasmid-encoded hexahistidine tagged RpoA [23]. Both "dual-tag" strategies led to identification of all four components of the core enzyme, the major sigma factor, and Rho; the hexahistidine/V5 tag did not lead to detection of NusA, NusG, HepA, or GreB. The chromosomal TAP/SPA tag thus offers more sensitivity in some cases than the plasmid-borne dual hexahistidine/V5 tag. From our results, it is not possible to determine whether this is due to the location of the tagged gene (chromosomal versus plasmid) or to the properties of the tag and the associated affinity isolation. However, results shown in Table 2 suggest that, in our hands, the overall methodology, when applied to a chromosomally tagged RpoA, provides similar results for proteins expected to be

more abundant in the sample while diverging somewhat for polypeptides that associate conditionally with RNA polymerase and for background proteins. Nonetheless, the plasmid-encoded hexahistidine/V5 tagging strategy used in the current study identifies known interaction partners for RpoA. The results are generally comparable (given the recognized noisy nature of these types of measurements [13]) to previous studies reported in the literature [23, 24] for major interacting proteins, but with some differences in less-abundant or transient interactors and in the background proteins observed. Affinity-tagged baits expressed from plasmid-encoded genes could lead to false interactions that would not occur in the native complex due to expression of the bait protein at inappropriate times, concentrations or locations in the cell.

When we extended the plasmid-encoded hexahistidine/V5 strategy to other components of RNA polymerase in *R. palustris* and *S. oneidensis* MR-1, the expected small "networks" of multiple connections among subunits emerged. As shown in Tables 3 and 4, interactions among the core components of RNA polymerase could be confidently identified in *R. palustris* and *S. oneidensis*. Other interactors besides the core components were also detected. However, some known interactors (sigma factors for *S. oneidensis*; NusA and NusG for *R. palustris*) were not detected. The observed differences between the two species suggest that a comparative analysis may be fruitful. For example, the pBBR5DEST42 vector may cause different species dependent stresses affecting expression and/or assembly of RNA polymerase holoenzyme. The biophysical characteristics (e.g. binding constants) of protomers that conditionally associate with the RNA polymerase core may differ between species. Regulation of RNA polymerase via sigma factors and other components may differ between species. Examination of this complex in additional species may help to focus such a comparative analysis. As with all affinity-MS approaches, a number of proteins were detected that are not expected interactors with RNA polymerase

(potential false positives), especially abundant proteins such as components of the ribosome. The BEPro$^3$ tool [44] proved useful for distinguishing potentially authentic bait-prey interactions from "background."

**Conclusions**

We have described a tool for expressing fusion proteins containing affinity tags via a system that combines a broad-host-range capability [27] with the flexibility to incorporate any of several different affinity tags via the commercial Gateway site-specific recombination system [28]. Although this paper has focused on the C-terminal hexahistidine/V5 affinity tag, other constructs can be readily generated, for example we have evaluated N-terminal glutathione-s-transferase (GST), C-terminal GST, and N-terminal hexahistidine.

Our goal is to perform large-scale experimental studies of protein-protein interactions in microbial species that have potential for environmental applications. We are currently applying this system to large-scale characterization of protein-protein interactions in *R. palustris*. The system we describe will allow us to pursue several refinements to optimize throughput and accuracy in identifying protein-protein interactions. The flexibility of the destination vector design will allow us to explore the use of other affinity tag combinations and promoters, potentially expanding the range of protein-protein interactions that can be characterized through the powerful combination of affinity isolation and mass spectrometry.

**Acknowledgements**

We thank the laboratories of Caroline Harwood (University of Washington) and Andrew Emili (University of Toronto) for generous gifts of materials. This research was sponsored by the Genomics:GTL program, Office of Biological and Environmental Research, U.S. Department of Energy, under contract No. DE-AC05-00OR22725 with Oak Ridge National Laboratory, managed and operated by UT-Battelle, LLC. WJH acknowledges the UT-ORNL Graduate School of Genome Science and Technology for support.

**Supporting Information Available**: Supplementary Table I and Table II. The material is available free at http://pubs.acs.org.

## Figure Captions

**Figure 1**. Overview of our approach for experimental determination of protein-protein interactions in Gram-negative bacteria.

**Figure 2**. Map containing relevant features of pBBR5DEST42 destination vector (see Materials and Methods for details of plasmid construction) *rep*, pBBR1MCS5 replication protein; *mob,* mobilization protein; *ccdB*, death gene; $Cm^r$, chloramphenicol resistance protein; $Gm^r$, gentamicin resistance protein; V5, C-terminal V5 epitope fusion; 6xHis, hexahistidine C-terminal fusion.

**Figure 3**. Western blot analysis of several RpoA hexahistidine/V5-fusion proteins expressed from pBBR5DEST42. Lane 1, MW standards; lane 2, *E. coli* K12 pBBR5DEST42-ECK3282, predicted MW 40768 Da; lane 3, *R. palustris* pBBR5DEST42-RPA3226, predicted MW 41791 Da; lane 4, *S. oneidensis* MR1 pBBR5DEST42-SO0256, predicted MW 40417 Da. Each lane contains ~30 μg total cell lysate. The blot was probed with anti-V5 antibody.

Table 1. Bacterial strains, plasmids and primers used in this study

| Strain, plasmid or primer | Genotype, phenotype or sequence of primer (5'-3') | Reference, origin or description |
|---|---|---|
| *E. coli* | | |
| K12 | K-12 MG1655 Wild-type strain | ATCC |
| DH5α | $F^-$ ϕ80d*lacZ*ΔM15 Δ(*lacZYA-argF*) U169 *rec*A1 *end*A1 *hsd*R17($r_k^-$, $m_k^+$) *pho*A *sup*E44 λ$^-$ *thi*-1 *gyr*A96 *rel*A1 | Novagen |
| DB3.1 | $F^-$ *gyr*A462 *end*A1 Δ(sr1-*rec*A) *mcr*B *mrr hsd*S20($r_B$-, $m_B$-) *sup*E44 *ara*-14 *gal*K2 *lac*Y1 *pro*A2 *rps*L20($Sm^R$) *xyl*-5 λ- l*eu mtl*1 | Invitrogen |
| β2155 | *thrB1004 pro thi strA hsdS lacZDM15* (F9 *lacZDM15 lacI*q *traD36 proA*1 *proB*1) D*dapA*::*erm* (Ermr) *pir*::RP4 [::*kan* (Kmr) from SM10] | [37] |
| BL21(DE3) | $F^-$ *omp*T *hsd*$S_B$($r_B^-m_B^-$) *gal dcm* (DE3) | Invitrogen |
| *R. palustris* CGA010 | hupV repaired derivative of CGA009 [29] | [33] |
| *S. oneidensis* MR1 | Wild-type strain | ATCC 700550 [54] |
| Plasmids | | |
| pET-DEST42 | $Ap^r$, $Cm^r$, C-terminal 6XHis and V5 epitope | Invitrogen |
| pBBR1MCS5 | $Gm^r$, *mob,* broad host range cloning vector | [27] |
| pBBR5DEST42 | $Gm^r$, C-terminal 6XHis and V5 epitope | This study |
| pDONR221 | $Km^r$, gateway entry vector$Gm^r$, N-terminal GST | Invitrogen |
| PCR Primers[a] | | |
| RPA3226-for | gggg aca agt ttg tac aaa aaa gca ggc ttc gaa gga gat agaATGACGATCCAGAAAAATTGGC | *rpoA* |
| RPA3226-rev | ggg gac cac ttt gta caa gaa agc tgg gtcGTAGTGATCTTCGAAGCGCTT | |
| RPA0367-for | gggg aca agt ttg tac aaa aaa gca ggc ttc gaa gga gat agaATGGCCCGTGCTGCTACTCT | *rpoH* |
| RPA0367-rev | ggg gac cac ttt gta caa gaa agc tgg gtcGTGCGCTGCTTCCAGCGCC | |
| RPA3267-for | gggg aca agt ttg tac aaa aaa gca ggc ttc gaa gga gat agaATGGCGATGAACCAGGAAATTA | *rpoC* |
| RPA3267-rev | ggg gac cac ttt gta caa gaa agc tgg gtcCTCTGCCGGCGGCAGCGA | |
| RPA3268-for | gggg aca agt ttg tac aaa aaa gca ggc ttc gaa gga gat agaATGGCGCAGCAGACGTTCAC | *rpoB* |
| RPA3268-rev | ggg gac cac ttt gta caa gaa agc tgg gtcCTCGGCAGCCTCGGCCGG | |
| ECK3282-for | gggg aca agt ttg tac aaa aaa gca ggc ttc gaa gga gat aga ATGCAGGGTTCTGTGACAG | *rpoA* |
| ECK3282-rev | ggg gac cac ttt gta caa gaa agc tgg gtc CTCGTCAGCGATGCTTGCC | |
| SO0256-for | gggg aca agt ttg tac aaa aaa gca ggc ttc gaa gga gat agaATGCAGGGTTCTGTTACAGAAT | *rpoA* |
| SO0256-rev | ggg gac cac ttt gta caa gaa agc tgg gtcTAGGTCGTCTGCTAAACTAGCT | |
| SO0225-for | gggg aca agt ttg tac aaa aaa gca ggc ttc gaa gga gat | *rpoC* |

| | | |
|---|---|---|
| | agaATGGTTTACTCCTATTCTGAAA | |
| SO0225-rev | ggg gac cac ttt gta caa gaa agc tgg gtcTTCCTGATCCAACTCGATATTA | |
| SO0360-for | gggg aca agt ttg tac aaa aaa gca ggc ttc gaa gga gat agaATGGCTCGCGTAACTGTAGAAG | *rpoZ* |
| SO0360-rev | ggg gac cac ttt gta caa gaa agc tgg gtcTAGTGAACGGCCTTCAGCGAT | |

[a]Gene specific sequences are in upper case. Shine-Delgarno sequence is underlined.

Table 2. Proteins co-isolating with affinity-tagged RpoA in *E. coli**

| **Tag, Location** | **6-His/V5,[a,b] plasmid** | **TAP[a,c], chromosomal** | **TAP/SPA, chromosomal** | **6-His, plasmid** | | |
|---|---|---|---|---|---|---|
| **Detection Method** | LC-MS-MS | LC-MS-MS | LC-MS-MS and MALDI | MALDI | | |
| **Source** | This study | This study | Ref. [24d] | Ref. [23e] | STRING[f] | Prolinks[g] |
| **Gene[h]** | | | | | | |
| **rpoC** | 62.5 | 381 | **Y** | y | 0.999 | 0.505 GN |
| **rpoB** | 50.5 | 331 | **Y** | y | 0.999 | 0.53 GN |
| **rpoD** | 6 | 62 | **Y** | y | 0.999 | 0.577 PP |
| rplB | 15.5 | 6 | **Y** | y | 0.909 | 0.807 GN |
| rpsE | 30 | 5 | **Y** | y | 0.961 | 0.877 GN, 0.17 PP |
| rpsD | 5.5 | 4 | **Y** | y | 0.989 | 0.87 PP, 0.853 GN, 0.676 GC |
| **rpoA (bait)** | 107 | 92 | y | | | |
| **rpoZ** | 11 | 19 | **Y** | | 0.998 | N |
| rplD | 3.5 | 5 | **Y** | | 0.926 | 0.799 GN, 0.392 PP |
| rpsG | 4 | 3 | **Y** | | 0.898 | 0.717 GN, 0.354 PP |
| rpsJ | 5 | 2 | y | | 0.923 | 0.795 GN, 0.471 PP |
| rpsM | 4 | 2 | y | | 0.98 | 0.879 GN, 0.637 PP |
| rplO | 6 | | y | y | 0.965 | 0.745GN |
| tufA | 24.5 | 23 | | | 0.85 | 0.686 GN |
| hupA | | 4 | y | | 0.16 | 0.167 PP |
| nusA | | 4 | **Y** | | 0.994 | 0.778 PP |
| rpsO | | 2 | y | | N | 0.248 PP |
| rpsT | | 2 | y | | N | N |
| hepA | | | **Y** | y | 0.579 | N |
| rho | 8 | | y | | 0.535 | 0.274 PP |
| rplA | 5 | | y | | 0.871 | 0.689 PP, 0.367 GN |
| rplE | 18.5 | | y | | 0.977 | 0.846 GN, 0.195 PP |
| rplF | 2 | | | y | 0.948 | 0.867 GN, 0.633 PP |
| rplM | 6.5 | | **Y** | | 0.86 | 0.685 GN, 0.388 PP |
| rplP | 6.5 | | y | | 0.947 | 0.821 GN, 0.633 PP |
| rpsB | 26.5 | | **Y** | | 0.687 | 0.435 PP |
| rpsC | 17.5 | | **Y** | | 0.974 | 0.816 GN |
| greB | | | **Y** | | 0.897 | 0.148 PP |
| nusG | | | **Y** | | 0.899 | 0.566 PP, 0.454 GN |
| rplC | | | **Y** | | 0.899 | 0.798 GN, 0.314 PP |
| rplL | | | **Y** | | 0.919 | 0.159 GN |
| rpsP | | | **Y** | | 0.242 | N |

*Table contains proteins detected by two or more approaches, plus those reported as validated by Butland *et al.* [24]

a. Table entries correspond to number of MS-MS spectra identified as arising from peptides of the corresponding protein. Average of two LC-MS-MS runs for His-V5 tag; single run for TAP tag.
b. Proteins detected only with His-V5 tagged rpoA: atpF (6), crp (5.5), cyoA (7.5), dacB (3), dnaJ (10.5), gatY (9), glpD (36), lldD (3), ppiD (4), rplJ (8), rpsI (4.5), rpsK (4), secG (3), xerD (6), yfgM (2.5) , yqjD (8.5), yqjI (4.5). This list provides gene name followed in parentheses by number of MS-MS spectra (see note a above).
c. Proteins detected only with TAP tagged rpoA in present study: dps (23), hupA (4), hupB (4), pepA (12), rplV (2), rpmC (6), rpoE (2), rpoS (10), seqA (4), rpsO (2), rpsT (2). This list provides gene name followed in parentheses by number of MS-MS spectra (see note a above).
d. Interactions reported by Butland *et al.* [24] denoted by "y", with "**Y"** denoting interactions reported as validated. Other interactions reported:[24] aceE, aceF, aspS, cspC, mreB, pssA, rfaK, rplS, rplU, rplW, rplX, rpmA, rpmB, rpmG, rpsA, rpsF, rpsN, rpsS, rpsU, wecG.
e. Interactions reported by Arifuzzaman *et al.* [23] denoted by "y" in the table. Other interactors reported [23]: fur, greA, hybE, pta, yccC.
f. Combined score predicted from STRING (http://string.embl.de). N: not found.
g. Predicted scores from Prolinks (http://mysql5.mbi.ucla.edu/cgi-bin/functionator/pronav), with prediction methods: PP: phylogenetic profile; GN: gene neighbor; GC: gene cluster. N: not found. [49]
h. Bold entries are components of RNA polymerase core enzyme.

Table 3. Selected[a] proteins detected in affinity isolations with tagged components of RNA polymerase in *S. oneidensis* MR-1

| | | | | Baits | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | SO0225, RpoC | | SO0256, RpoA | | SO0360, RpoZ | |
| | | biological replicates, technical replicates[b] → | | 3, 6 | | 4, 6 | | 3, 6 | |
| Preys | | | | | | | | | |
| **Gene ID** | **Name[c]** | **Product** | ***Ubiquity*** | ***BO*** | **$SC_{av}$** | ***BO*** | **$SC_{av}$** | ***BO*** | **$SC_{av}$** |
| SO0224 | rpoB | DNA-directed RNA polymerase, beta subunit | 0.17 | **1** | **23.33** | **1** | **104** | **1** | **10.83** |
| SO0225 | rpoC | DNA-directed RNA polymerase, beta subunit | 0.21 | **1** | **16.33** | **1** | **81** | **1** | **6.67** |
| SO0256 | rpoA | DNA-directed RNA polymerase, alpha subunit | 0.17 | **1** | **17.17** | **1** | **42.67** | **1** | **7.83** |
| SO0360 | rpoZ | DNA-directed RNA polymerase, omega subunit | 0.09 | **0.59** | **1.33** | **1** | **12** | **1** | **17.33** |
| SO1490 | adhB | alcohol dehydrogenase II | 0.28 | **0.98** | **2** | 0.25 | 8 | **0.96** | **1.83** |
| SO4236 | leuA | 2-isopropylmalate synthase | 0.35 | | | **1** | **4.83** | **1** | **3.67** |
| SO3907 | | conserved hypothetical protein | 0.04 | 0 | 1.17 | 0 | 1.17 | **0.99** | **2** |
| SO4747 | atpD | ATP synthase F1, beta subunit | 0.24 | | | **0.96** | **3.5** | | |
| SO1624 | purU | formyltetrahydrofolate deformylase | 0.48 | | | **1** | **2.33** | **1** | **1** |
| SO1630 | tsf | translation elongation factor Ts | 0.48 | | | **1** | **2.83** | | |
| SO4263 | | conserved hypothetical protein | 0.24 | | | **1** | **2.83** | | |
| SO0227 | rpsG | ribosomal protein S7 | 0.24 | | | **1** | **2.33** | | |
| SO0253 | rpsM | ribosomal protein S13 | 0.37 | | | **1** | **2.17** | | |
| SO0222 | rplJ | ribosomal protein L10 | 0.38 | | | **1** | **2** | | |
| SO2485 | | deoxyguanosinetriphosphate triphosphohydrolase, putative | 0.29 | | | **0.94** | **2** | | |
| SO3652 | rplU | ribosomal protein L21 | 0.36 | | | **0.66** | **1.83** | | |
| SO2299 | thrS | threonyl-tRNA synthetase | 0.19 | | | **0.87** | **1.67** | | |
| SO3348 | hemH-2 | ferrochelatase | 0.12 | | | **0.93** | **1.33** | | |
| SO4105 | mshA | MSHA pilin protein MshA | 0.37 | | | **0.98** | **1.17** | | |

[a] Selection criteria: *Ubiquity* ≤ 0.5, *Bayes Odds* ≥0.5, two or more peptides. See Supplementary Table I for complete listing.
[b] biological replicates = number of different cultures; technical replicates = number of LC-MS/MS experiments.
*BO*: *Bayes Odds*
$SC_{av}$: spectrum count averaged across technical replicates
Bold entries correspond to *Bayes Odds* ≥0.5.
[c] Underlined entries are components of RNA polymerase core enzyme.

Table 4. Selected[a] proteins detected in affinity isolations with tagged components of RNA polymerase in *R. palustris*

| | | | | **Baits** | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | **RPA0367, RpoH** | | **RPA3226, RpoA** | | **RPA3267, RpoC** | | **RPA3268, RpoB** | |
| | | biological replicates, technical replicates[b] → | | 3, 6 | | 4, 4 | | 2, 4 | | 4, 10 | |
| **Preys** | | | | | | | | | | | |
| **Gene ID** | **Name[c]** | **Product** | ***Ubiquity*** | ***BO*** | **$SC_{av}$** | ***BO*** | **$SC_{av}$** | ***BO*** | **$SC_{av}$** | ***BO*** | **$SC_{av}$** |
| RPA3268 | rpoB | RNA polymerase beta subunit | 0.19 | 0.02 | 19.83 | **0.99** | **177.** | **0.99** | **130.25** | **1** | **85.1** |
| RPA3267 | rpoC | RNA polymerase beta' subunit | 0.31 | **1** | **12.17** | **1** | **87.25** | **1** | **137.75** | **1** | **43.2** |
| RPA3226 | rpoA | DNA-directed RNA polymerase alpha subunit | 0.25 | **1** | **8.67** | **1** | **123.5** | **1** | **43.25** | **1** | **31.8** |
| RPA0367 | rpoH | stress response sigma factor | 0.19 | **1** | **38** | **0.98** | **9.5** | **0.98** | **11** | 0 | 2.9 |
| RPA1288 | rpoD | RNA polymerase sigma subunit | 0.19 | | | **1** | **11.5** | **1** | **25.5** | **1** | **12.9** |
| RPA2692 | rnpO | RNA polymerase omega subunit | 0.19 | 0.01 | 1.33 | **1** | **13.5** | **1** | **17.5** | **1** | **11.7** |
| RPA0183 | RPA0183 | unknown protein | 0.19 | **0.81** | **9.67** | 0 | 2.75 | 0.33 | 15. | **0.95** | **7.5** |
| RPA1093 | ribB | possible GTP cyclohydrolase II, riboflavin | 0.07 | **0.87** | **11.33** | 0 | 2.75 | 0.41 | 13.5 | 0.02 | 4.7 |
| RPA0333 | dnaK | heat shock protein DnaK (70) | 0.06 | **1** | **23.5** | | | | | | |
| RPA3254 | rpsG | 30S ribosomal protein S7 | 0.45 | **1** | **5.17** | **0.81** | **2.75** | **0.99** | **10.25** | **0.92** | **2.7** |
| RPA3246 | rpsS | 30S ribosomal protein S19 | 0.38 | **1** | **6.33** | 0.01 | 1 | 0.13 | 3.75 | **1** | **2.6** |
| RPA1548 | puhA | H subunit of photosynthetic reaction center | 0.05 | 0.24 | 5 | | | **0.53** | **6** | 0 | 2.5 |
| RPA0304 | RPA0304 | possible outer membrane lipoprotein GNA33 | 0.49 | **1** | **6.33** | 0.01 | 1 | 0.23 | 1.25 | 0.05 | 1.3 |
| RPA0060 | RPA0060 | conserved unknown protein | 0.06 | | | **0.9** | **7.25** | 0.14 | 2 | 0 | 0.6 |
| RPA1597 | purQ | phosphoribosylformylglycinamide synthetase | 0.12 | 0 | 0.67 | | | **0.95** | **8.75** | | |
| RPA4356 | ctc, rplY, L25 | putative 50S ribosomal protein L25 | 0.12 | 0 | 2.5 | 0 | 0.75 | **0.87** | **5.25** | | |
| RPA0202 | acnA | aconitate hydratase | 0.06 | 0 | 0.5 | | | **0.97** | **7.5** | | |
| RPA1653 | RPA1653 | conserved unknown protein | 0.07 | **0.9** | **3.83** | 0.01 | 2.75 | | | 0 | 1.3 |
| RPA3252 | tufA, EF-Tu | elongation factor Tu | 0.06 | | | | | **0.93** | **6.5** | 0 | 1.1 |
| RPA3283 | tuf, EF-Tu | elongation factor Tu | 0.06 | | | | | **0.93** | **6.5** | 0 | 1.1 |
| RPA3270 | rplJ | 50S ribosomal protein L10 | 0.17 | **0.62** | **2.5** | 0 | 0.5 | **0.73** | **3.5** | 0 | 0.8 |
| RPA1205 | RPA1205 | putative alcohol dehydrogenase | 0.12 | | | | | **0.99** | **5.75** | 0 | 1.2 |
| RPA0844 | atpF2 | putative FoF1 ATP synthase, subunit B' | 0.24 | 0 | 1.67 | 0 | 0.5 | **0.7** | **4.75** | | |
| RPA3238 | rplE | 50S ribosomal protein L5 | 0.06 | 0 | 0.67 | | | **0.98** | **5.25** | | |
| RPA0176 | atpD | putative H+-transporting ATP synthase beta | 0.31 | 0 | 1.5 | | | **0.98** | **4** | 0 | 0.4 |
| RPA0241 | rplS, L19 | 50s ribosomal protein L19 | 0.07 | 0 | 0.67 | | | **0.99** | **4.75** | | |
| RPA0240 | leuC | 3-isopropylmalate dehydratase | 0.06 | 0 | 1.5 | | | **0.95** | **3.5** | | |
| RPA0331 | RPA0331 | possible heat shock protein (HSP-70 cofactor) | 0.06 | **1** | **5** | | | | | | |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| RPA1051 | RPA1051 | pyruvate phosphate dikinase | 0.06 | | | 0.07 | 0.75 | **0.77** | **4** | | |
| RPA2922 | rpsB | 30S ribosomal protein S2 | 0.13 | 0 | 1.17 | | | **0.96** | **2.75** | 0 | 0.5 |
| RPA0217 | sdhA | succinate dehydrogenase flavoprotein subunit | 0.1 | | | | | **0.63** | **2.75** | | |
| RPA0934 | RPA0934 | conserved unknown protein | 0.06 | | | | | **0.68** | **2.5** | | |
| RPA0854 | hemO | 5-aminolevulinic acid synthase (ALAS) | 0.06 | | | | | **0.61** | **2** | | |
| RPA3253 | fusA, EF-G | elongation factor G | 0.06 | | | | | **0.96** | **1.75** | | |
| RPA4636 | RPA4636 | FeoA family | 0.02 | **0.55** | **1.67** | | | | | | |

[a] Selection criteria: *Ubiquity* ≤ 0.5, *Bayes Odds* ≥0.5, two or more peptides. See Supplementary Table II for complete listing.
[b]biological replicates = number of different cultures; technical replicates = number of LC-MS/MS experiments.
*BO*: *Bayes Odds*
$SC_{av}$: spectrum count averaged across technical replicates
Bold entries correspond to *Bayes Odds* ≥0.5.
[c] Underlined entries are components of RNA polymerase core enzyme.

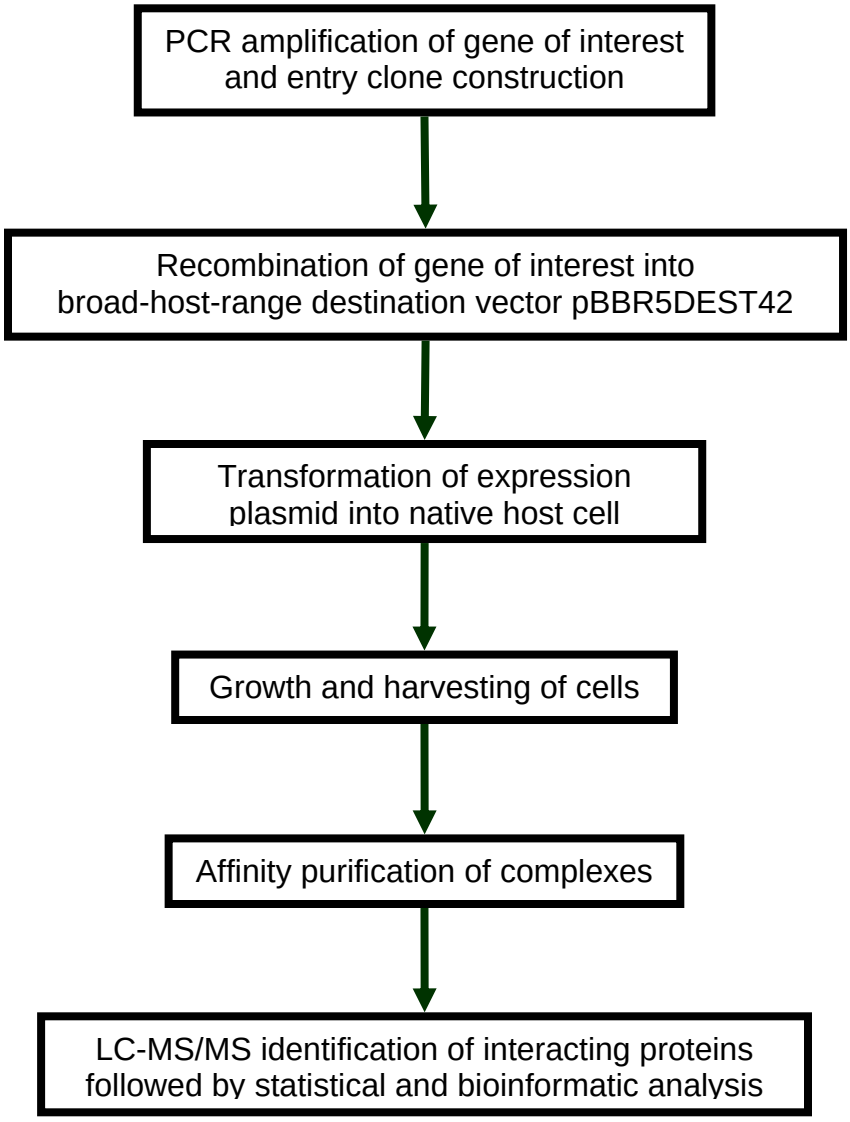

PCR amplification of gene of interest and entry clone construction
Recombination of gene of interest into broad-host-range destination vector pBBR5DEST42
Transformation of expression plasmid into native host cell
Growth and harvesting of cells
Affinity purification of complexes
LC-MS/MS identification of interacting proteins followed by statistical and bioinformatic analysis

Figure 1

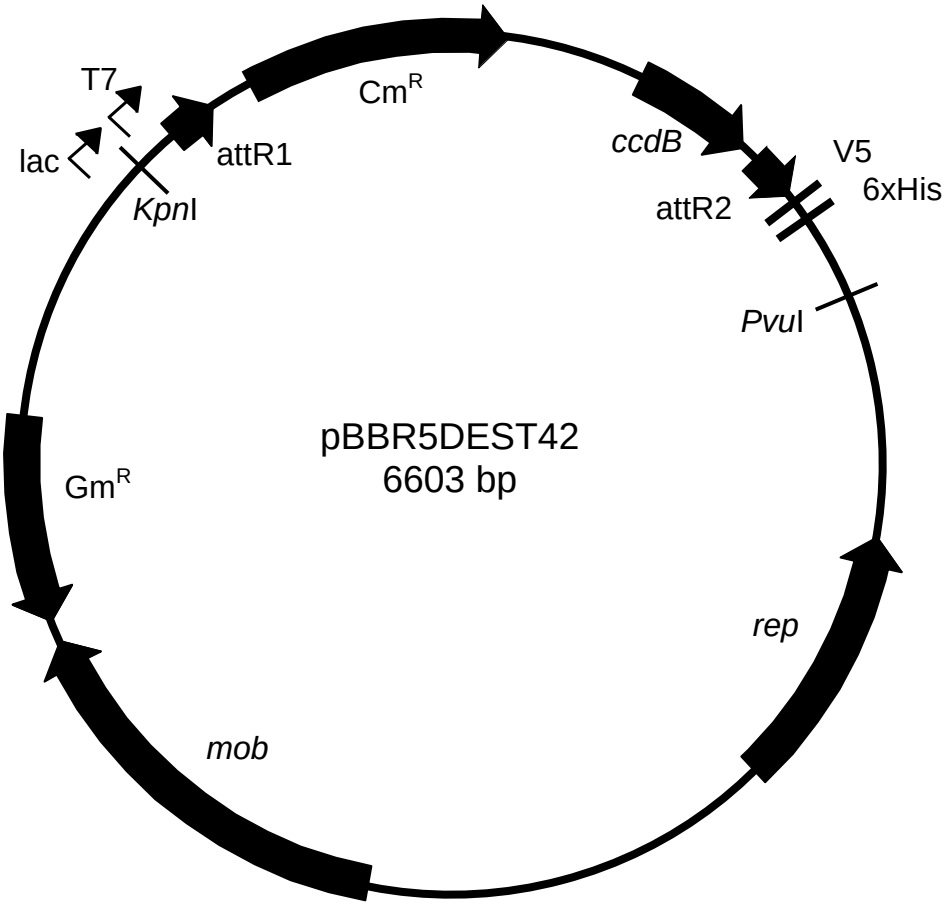

T7
lac
attR1
KpnI
CmR
ccdB
V5
6xHis
attR2
PvuI
pBBR5DEST42
6603 bp
GmR
rep
mob

Figure 2

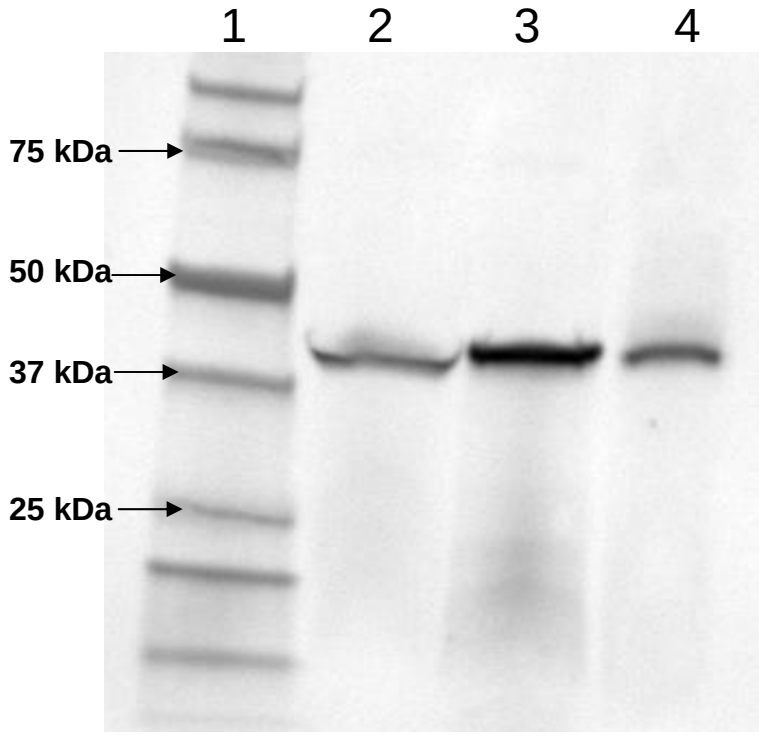



Figure 3

**References**

1. Alberts, B., The cell as a collection of protein machines: Preparing the next generation of molecular biologists. *Cell* **1998,** 92, (3), 291-294.

2. Klotz, I. M.; Darnall, D. W.; Langerman, N. R.; Neurath, H.; Hill, R. L.; Boeder, C. L., Quaternary Structure of proteins. In *The Proteins*, Academic Press: New York, 1975; Vol. 3, pp 293-411.

3. Phizicky, E. M.; Fields, S., Protein-Protein Interactions - Methods for Detection and Analysis. *Microbiological Reviews* **1995,** 59, (1), 94-123.

4. Alberts, B.; Miakelye, R., Unscrambling the Puzzle of Biological Machines - the Importance of the Details. *Cell* **1992,** 68, (3), 415-420.

5. Oliver, S., Proteomics: Guilt-by-association goes global. *Nature* **2000,** 403, (6770), 601-603.

6. de Silva, E.; Stumpf, M. P. H., Complex networks and simple models in biology. *Journal of the Royal Society Interface* **2005,** 2, (5), 419-430.

7. Drewes, G.; Bouwmeester, T., Global approaches to protein-protein interactions. *Current Opinion in Cell Biology* **2003,** 15, (2), 199-205.

8. Ideker, T., A systems approach to discovering signaling and regulatory pathways - or, how to digest large interaction networks into relevant pieces. *Advances in Systems Biology* **2004,** 547, 21-30.

9. Noirot, P.; Noirot-Gros, M. F., Protein interaction networks in bacteria. *Current Opinion In Microbiology* **2004,** 7, (5), 505-512.

10. Schachter, V., Protein-interaction networks: from experiments to analysis. *Drug Discovery Today* **2002,** 7, (11), S48-S54.

11. Xia, Y.; Yu, H. Y.; Jansen, R.; Seringhaus, M.; Baxter, S.; Greenbaum, D.; Zhao, H. Y.; Gerstein, M., Analyzing cellular biochemistry in terms of molecular networks. *Annual Review of Biochemistry* **2004,** 73, 1051-1087.

12. Chen, Y.; Xu, D., Computational analyses of high-throughput protein-protein interaction data. *Current Protein & Peptide Science* **2003,** 4, (3), 159-180.

13. Musso, G. A.; Zhang, Z.; Emili, A., Experimental and Computational Procedures for the Assessment of Protein Complexes on a Genome-wide Scale. *Chemical Reviews* **2007,** 107, (8), 3585-3600.

14. Fields, S.; Song, O. K., A novel genetic system to detect protein-protein interactions. *Nature* **1989,** 340, 245-246.

15. Rain, J. C.; Selig, L.; De Reuse, H.; Battaglia, V.; Reverdy, C.; Simon, S.; Lenzen, G.; Petel, F.; Wojcik, J.; Schachter, V.; Chemama, Y.; Labigne, A.; Legrain, P., The protein-protein interaction map of *Helicobacter pylori*. *Nature* **2001,** 409, (6817), 211-215.

16. Ito, T.; Chiba, T.; Ozawa, R.; Yoshida, M.; Hattori, M.; Sakaki, Y., A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proceedings of The National Academy of Sciences of The United States of America* **2001,** 98, (8), 4569-4574.

17. Uetz, P.; Giot, L.; Cagney, G.; Mansfield, T. A.; Judson, R. S.; Knight, J. R.; Lockshon, D.; Narayan, V.; Srinivasan, M.; Pochart, P.; Qureshi-Emili, A.; Li, Y.; Godwin, B.; Conover, D.; Kalbfleisch, T.; Vijayadamodar, G.; Yang, M. J.; Johnston, M.; Fields, S.; Rothberg, J. M., A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae Nature* **2000,** 403, 623-627.

18. Bertone, P.; Snyder, M., Advances in functional protein microarray technology. *FEBS Journal* **2005,** 272, (21), 5400-5411.

19. Zhu, H.; Bilgin, M.; Bangham, R.; Hall, D.; Casamayor, A.; Bertone, P.; Lan, N.; Jansen, R.; Bidlingmaier, S.; Houfek, T.; Mitchell, T.; Miller, P.; Dean, R. A.; Gerstein, M.; Snyder, M., Global analysis of protein activities using proteome chips. *Science* **2001,** 293, 2101-2105.

20. Sidhu, S. S.; Fairbrother, W. J.; Deshayes, K., Exploring protein-protein interactions with phage display. *Chembiochem* **2003,** 4, (1), 14-25.

21. Gavin, A. C.; B”sche, M.; Krause, R.; Grandi, P.; Marzioch, M.; Bauer, A.; Schultz, J.; Rick, J. M.; Michon, A. M.; Cruciat, C. M.; Remor, M.; H”fert, C.; Schelder, M.; Brajenovic, M.; Ruffner, H.; Merino, A.; Klein, K.; Hudak, M.; Dickson, D.; Rudi, T.;

Gnau, V.; Bauch, A.; Bastuck, S.; Huhse, B.; Leutwein, C.; Heurtier, M. A.; Copley, R. R.; Edelmann, A.; Querfurth, E.; Rybin, V.; Drewes, G.; Raida, M.; Bouwmeester, T.; Bork, P.; Seraphin, B.; Kuster, B.; Neubauer, G.; Superti-Furga, G., Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **2002,** 415, 141-147.

22. Ho, Y.; Gruhler, A.; Heilbut, A.; Bader, G. D.; Moore, L.; Adams, S. L.; Millar, A.; Taylor, P.; Bennett, K.; Boutilier, K.; Yang, L.; Wolting, C.; Donaldson, I.; Schandorff, S.; Shewnarane, J.; Vo, M.; Taggart, J.; Goudreault, M.; Muskat, B.; Alfarano, C.; Dewar, D.; Lin, Z.; Michalickova, K.; Willems, A. R.; Sassi, H.; Nielsen, P. A.; Rasmussen, K. J.; Andersen, J. R.; Johansen, L. E.; Hansen, L. H.; Jespersen, H.; Podtelejnikov, A.; Nielsen, E.; Crawford, J.; Poulsen, V.; Sorensen, B. D.; Matthiesen, J.; Hendrickson, R. C.; Gleeson, F.; Pawson, T.; Moran, M. F.; Durocher, D.; Mann, M.; Hogue, C. W. V.; Figeys, D.; Tyers, M., Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* **2002,** 415, 180-183.

23. Arifuzzaman, M.; Maeda, M.; Itoh, A.; Nishikata, K.; Takita, C.; Saito, R.; Ara, T.; Nakahigashi, K.; Huang, H. C.; Hirai, A.; Tsuzuki, K.; Nakamura, S.; Altaf-Ul-Amin, M.; Oshima, T.; Baba, T.; Yamamoto, N.; Kawamura, T.; Ioka-Nakamichi, T.; Kitagawa, M.; Tomita, M.; Kanaya, S.; Wada, C.; Mori, H., Large-scale identification of protein-protein interaction of *Escherichia coli* K-12. *Genome Research* **2006,** 16, (5), 686-691.

24. Butland, G.; Peregrin-Alvarez, J. M.; Li, J.; Yang, W. H.; Yang, X. C.; Canadien, V.; Starostine, A.; Richards, D.; Beattie, B.; Krogan, N.; Davey, M.; Parkinson, J.; Greenblatt, J.; Emili, A., Interaction network containing conserved and essential protein complexes in *Escherichia coli*. *Nature* **2005,** 433, (7025), 531-537.

25. Tackett, A. J.; DeGrasse, J. A.; Sekedat, M. D.; Oeffinger, M.; Rout, M. P.; Chait, B. T., I-DIRT, a general method for distinguishing between specific and nonspecific protein interactions. *Journal of Proteome Research* **2005,** 4, (5), 1752-1756.

26. Frazier, M. E.; Johnson, G. M.; Thomassen, D. G.; Oliver, C. E.; Patrinos, A., Realizing the Potential of the Genome Revolution: The Genomes to Life Program. *Science* **2003,** 300, (5617), 290-293.

27. Kovach, M. E.; Elzer, P. H.; Hill, D. S.; Robertson, G. T.; Farris, M. A.; Roop, R. M.; Peterson, K. M., 4 new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes. *Gene* **1995,** 166, (1), 175-176.

28. Landy, A., Dynamic, Structural, and Regulatory Aspects of Lambda-Site-Specific Recombination. *Annual Review of Biochemistry* **1989,** 58, 913-949.

29. Larimer, F. W.; Chain, P.; Hauser, L.; Lamerdin, J.; Malfatti, S.; Do, L.; Land, M. L.; Pelletier, D. A.; Beatty, J. T.; Lang, A. S.; Tabita, F. R.; Gibson, J. L.; Hanson, T. E.; Bobst, C.; Torres, J.; Peres, C.; Harrison, F. H.; Gibson, J.; Harwood, C. S., Complete genome sequence of the metabolically versatile photosynthetic bacterium *Rhodopseudomonas palustris*. *Nature Biotechnology* **2004,** 22, (1), 55-61.

30. Heidelberg, J. F.; Paulsen, I. T.; Nelson, K. E.; Gaidos, E. J.; Nelson, W. C.; Read, T. D.; Eisen, J. A.; Seshadri, R.; Ward, N.; Methe, B.; Clayton, R. A.; Meyer, T.; Tsapin, A.; Scott, J.; Beanan, M.; Brinkac, L.; Daugherty, S.; DeBoy, R. T.; Dodson, R. J.; Durkin, A. S.; Haft, D. H.; Kolonay, J. F.; Madupu, R.; Peterson, J. D.; Umayam, L. A.; White, O.; Wolf, A. M.; Vamathevan, J.; Weidman, J.; Impraim, M.; Lee, K.; Berry, K.; Lee, C.; Mueller, J.; Khouri, H.; Gill, J.; Utterback, T. R.; McDonald, L. A.; Feldblyum, T. V.; Smith, H. O.; Venter, J. C.; Nealson, K. H.; Fraser, C. M., Genome sequence of the dissimilatory metal ion-reducing bacterium *Shewanella oneidensis*. *Nature Biotechnology* **2002,** 20, (11), 1118-1123.

31. Blattner, F. R.; Plunkett, G.; Bloch, C. A.; Perna, N. T.; Burland, V.; Riley, M.; ColladoVides, J.; Glasner, J. D.; Rode, C. K.; Mayhew, G. F.; Gregor, J.; Davis, N. W.; Kirkpatrick, H. A.; Goeden, M. A.; Rose, D. J.; Mau, B.; Shao, Y., The complete genome sequence of *Escherichia coli* K-12. *Science* **1997,** 277, (5331), 1453-&.

32. Rigaut, G.; Shevchenko, A.; Rutz, B.; Wilm, M.; Mann, M.; Seraphin, B., A generic protein purification method for protein complex characterization and proteome exploration. *Nature Biotechnology* **1999,** 17, 1030-1032.

33. Rey, F. E.; Oda, Y.; Harwood, C. S., Regulation of uptake hydrogenase and effects of hydrogen utilization on gene expression in *Rhodopseudomonas palustris*. *Journal of Bacteriology* **2006,** 188, (17), 6143-6152.

34. Gorby, Y. A.; Yanina, S.; McLean, J. S.; Rosso, K. M.; Moyles, D.; Dohnalkova, A.; Beveridge, T. J.; Chang, I. S.; Kim, B. H.; Kim, K. S.; Culley, D. E.; Reed, S. B.; Romine, M. F.; Saffarini, D. A.; Hill, E. A.; Shi, L.; Elias, D. A.; Kennedy, D. W.; Pinchuk, G.; Watanabe, K.; Ishii, S. i.; Logan, B.; Nealson, K. H.; Fredrickson, J. K., Electrically conductive bacterial nanowires produced by *Shewanella oneidensis strain* MR-1 and other microorganisms. *PNAS* **2006,** 103, (30), 11358-11363.

35. Kim, M. K.; Harwood, C. S., Regulation of Benzoate-CoA Ligase in *Rhodopseudomonas palustris*. *FEMS Microbiology Letters* **1991,** 83, (2), 199-203.

36. Sambrook, J.; Russell, D. W., *Molecular cloning: a laboratory manual,* . 3rd ed.; Cold Spring Harbor Laboratory Press
Cold Spring Harbor, NY., 2001; p p. 1.84–1.117.

37. Dehio, C.; Meyer, M., Maintenance of broad-host-range incompatibility group P and group Q plasmids and transposition of Tn5 in Bartonella henselae following conjugal plasmid transfer from *Escherichia coli*. *J. Bacteriol.* **1997,** 179, (2), 538-540.

38. Zeghouf, M.; Li, J.; Butland, G.; Borkowska, A.; Canadien, V.; Richards, D.; Beattie, B.; Emili, A.; Greenblatt, J. F., Sequential Peptide Affinity (SPA) system for the identification of mammalian and bacterial protein complexes. *Journal of Proteome Research* **2004,** 3, (3), 463-468.

39. Hervey, W. J.; Strader, M. B.; Hurst, G. B., Comparison of Digestion Protocols for Microgram Quantities of Enriched Protein Samples. *Journal of Proteome Research* **2007,** 6, (8), 3054-3061.

40. Eng, J. K.; McCormack, A. L.; Yates, J. R., III, An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *Journal of the American Society for Mass Spectrometry* **1994,** 5, 976-989.

41. Tabb, D. L.; McDonald, W. H.; Yates, J. R., DTASelect and contrast: Tools for assembling and comparing protein identifications from shotgun proteomics. *Journal of Proteome Research* **2002,** 1, (1), 21-26.

42. Gattiker, A.; Michoud, K.; Rivoire, C.; Auchincloss, A. H.; Coudert, E.; Lima, T.; Kersey, P.; Pagni, M.; Sigrist, C. J. A.; Lachaize, C.; Veuthey, A. L.; Gasteiger, E.; Bairoch, A., Automated annotation of microbial proteomes in SWISS-PROT. *Computational Biology and Chemistry* **2003,** 27, (1), 49-58.

43. Romine, M. F.; Elias, D. A.; Monroe, M. E.; Auberry, K.; Fang, R. H.; Fredrickson, J. K.; Anderson, G. A.; Smith, R. D.; Lipton, M. S., Validation of *Shewanella oneidensis* MR-1 small proteins by AMT tag-based proteome analysis. *Omics-a Journal of Integrative Biology* **2004,** 8, (3), 239-254.

44. Sharp, J. L.; Anderson, K. K.; Hurst, G. B.; Daly, D. S.; Pelletier, D. A.; Cannon, W. R.; Auberry, D. L.; Schmoyer, D. D.; HayesMcDonald, W.; White, A. M.; Hooker, B. S.; Victry, K. D.; Buchanan, M. V.; Kery, V.; Wiley, H. S., Statistically Inferring Protein-Protein Associations with Affinity Isolation LC-MS/MS Assays. *Journal of Proteome Research* **2007,** 6, (9), 3788-3795.

45. Edwards, A. M.; Kus, B.; Jansen, R.; Greenbaum, D.; Greenblatt, J.; Gerstein, M., Bridging structural biology and genomics: assessing protein interaction data with known complexes. *Trends In Genetics* **2002,** 18, (10), 529-536.

46. Hiratsu, K.; Amemura, M.; Nashimoto, H.; Shinagawa, H.; Makino, K., The *rpoE* Gene of *Escherichia coli*, Which Encodes Sigma(E), Is Essential for Bacterial-Growth at High-Temperature. *Journal of Bacteriology* **1995,** 177, (10), 2918-2922.

47. Liu, H. B.; Sadygov, R. G.; Yates, J. R., A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Analytical Chemistry* **2004,** 76, (14), 4193-4201.

48. von Mering, C.; Jensen, L. J.; Kuhn, M.; Chaffron, S.; Doerks, T.; Kruger, B.; Snel, B.; Bork, P., STRING 7--recent developments in the integration and prediction of protein interactions. *Nucleic Acids Research* **2007,** 35, (suppl_1), D358-D362.

49. Bowers, P. M.; Pellegrini, M.; Thompson, M. J.; Fierro, J.; Yeates, T. O.; Eisenberg, D., Prolinks: a database of protein functional linkages derived from coevolution. *Genome Biology* **2004,** 5, (5), Article #R35.

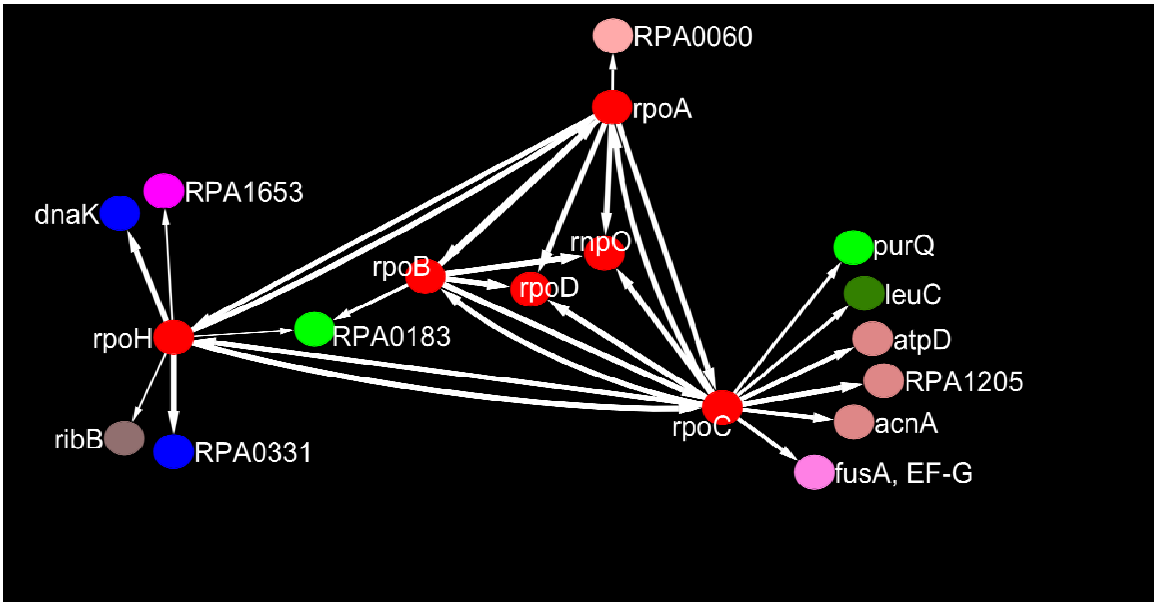
50. Mah, T. F.; Kuznedelov, K.; Mushegian, A.; Severinov, K.; Greenblatt, J., The alpha subunit of *E. coli* RNA polymerase activates RNA binding by NusA. *Genes & Development* **2000,** 14, (20), 2664-2675.

51. Rouby, J.; Pugniere, M.; Mani, J. C.; Granier, C.; Monmouton, P.; Saint Germain, S. T.; Leonetti, J. P., Characterization of monoclonal antibodies against *Escherichia coli* core RNA polymerase. *Biochemical Journal* **2002,** 361, 347-354.

52. Banerjee, S.; Chalissery, J.; Bandey, I.; Sen, R. J., Rho-dependent transcription termination: More questions than answers. *Journal of Microbiology* **2006,** 44, (1), 11-22.

53. Gamer, J.; Bujard, H.; Bukau, B., Physical Interaction Between Heat-Shock Proteins DnaK, DnaJ, and GrpE and the Bacterial Heat-Shock Transcription Factor-Sigma(32). *Cell* **1992,** 69, (5), 833-842.

54. Venkateswaran, K.; Moser, D. P.; Dollhopf, M. E.; Lies, D. P.; Saffarini, D. A.; MacGregor, B. J.; Ringelberg, D. B.; White, D. C.; Nishijima, M.; Sano, H.; Burghardt, J.; Stackebrandt, E.; Nealson, K. H., Polyphasic taxonomy of the genus *Shewanella* and description of *Shewanella oneidensis* sp. nov. *Int J Syst Bacteriol* **1999,** 49, (2), 705-724.




To enable protein-protein interaction studies across a wider range of Gram-negative bacteria, we have developed a methodology based on expression of affinity-tagged "bait" proteins from a medium copy-number plasmid, followed by affinity isolation and mass spectrometry identification of interacting proteins. The vector has been modified to incorporate the Gateway DEST vector recombination region, to facilitate cloning and expression of various fusion proteins. We demonstrate this methodology by characterizing interactions among subunits of the DNA-dependent RNA polymerase complex in two metabolically versatile Gram-negative microbial species of environmental interest, *Rhodopseudomonas palustris* CGA010 and *Shewanella oneidensis MR-1*.

# Appendix C
# Modified Perl Script for conducing Automated SEQUEST searches on *nix Operating Systems

```
#!/usr/bin/perl
##########################################################################
# 07/28/04 GitRSeq v0.95 - Hayes McDonald, Oak Ridge National Lab       #
# A sister script to GitRDug that automates the process of running a    #
# SEQUEST search on a Windows box.  It creates the folders and calls a  #
# series of programs to do the extraction, charge-state selection,      #
# search, and then pulls the data back together into SQT and MS2 files. #
##########################################################################
#
# Some modifications to get it working on the Linux boxes by WJH
#
print "\nGitRSeq v0.95\n";
print "Hayes McDonald, Oak Ridge National Lab\n";
print "\nHello and welcome GitRSeq, a program to take the pain\n";
print "and effort out of doing a SEQUEST search on a Windows box.\n";
print "\nBefore we get started, make certain you have everything you\n";
print "OS Conversion to Unix/Linux by Judson Hervey\n";
print "UT-ORNL Graduate School of Genome Science\n";
print "\nHello and welcome GitRSeq, a program to automate SEQUEST\n";
print "searches under Unix/Linux on Bombur and Balin.\n";
print "\nBefore we get started, make certain you have everything you\n";
print "need:\n";
print " - a folder of MS2 files\n";
print " - a sequest.params file for Unix\n";
print " - and an appropriate fasta file specified in the params file\n";
print "\nKick back and relax for a minute or two - this won't take as long as\n";
print "it does under Windoze...\n";

#####################
# read in raw files #
#####################
use Cwd;

$current_dir = cwd;

#print "\n $current_dir\n\n";
opendir (DIR, "$current_dir");

@filelist = readdir(DIR);

#################
```

```
# program paths #
#################
#$raw2ms2 = "raw2ms2.exe";
# **  We need to comment raw2ms2.exe out because it is Windoze based.  Thus, we need
# **   to start our workflow on Unix machines with MS2 files.  The SendRaw.pl script
# **  that I have written takes care of this issue if you use it to transfer your files.
$ms2zassign = "MS2ZAssign";
$sequest = "sequest27.linux";
$unitemare = "/usr/bin/perl /home/wj7/bin/Unitemare-win-H.pl";
$DTASelect = "DTASelect";

################################################
#this sets up all the flags - not in use yet #
################################################


#############################################################################
# this makes the "searches" directory because people would like the subfolders in #
# different folder than the RAW files                                       #
#############################################################################

$searches_dir = "$current_dir";
print "\nMaking folders and copying sequest.params ...\n";

foreach $file (@filelist)
        {
        if ($file =~ /(\S+)\.MS2/)
                {
                $sub_dir = $1;
                print "making $sub_dir";
                mkdir $sub_dir;
                chdir $sub_dir;
                print "\n\nExtracting MS/MS data - please be patient ...\n";
                system "cp ../sequest.params sequest.params";
                system "cp ../$file .";
                print "\nAssigning charge states ...\n";
                system "$ms2zassign --DTAs";
                ###################
                # sequest search #
                ###################
                system "find . -name \"\*\.dta\" | xargs $sequest -S" ;
                ####################################
                # modified unitemare and cleanup #
                ####################################
                print "\nCleaning up and getting rid of those pesky .DTAs and .OUTs\n";
```

```
                print "\nThank you for your continued patience (think of how much\n";
                print "space you'll be saving)\n";
                system "$unitemare";
                #system "COPY *.MS2 ..";
                if (-e "$sub_dir.sqt")
                        {
                        system "cp ./*.sqt ..";
                        my @currentlist = ();
                        opendir (CURRENT, ".");
                        @currentlist = readdir(CURRENT);
                        closedir CURRENT;
                        $filesremoved = unlink @currentlist;
                        print "\n$filesremoved\n";
                        }
                chdir $searches_dir;
                rmdir $sub_dir;
                #system "RD $sub_dir";
                }
        }

#############
# DTASelect #
#############

system "DTASelect";
```

# Appendix D
# H2O Isolation Protocol, All PP:WCL Protein Isotope Ratios

| Protein | PP | log2ratio | Chr. Features | Reps | Protein Description |
|---|---|---|---|---|---|
| P0AEQ3_GLNH_ECOLI | 1 | 4.6 | 126 | 4 | Glutamine-binding periplasmic protein |
| P0AFK9_POTD_ECOLI | 1 | 4.3 | 45 | 4 | Spermidine/putrescine-binding periplasmic protein |
| P0AFM2_PROX_ECOLI | 1 | 4.3 | 68 | 4 | Glycine betaine-binding periplasmic protein |
| P31550_THIB_ECOLI | 1 | 3.5 | 4 | 1 | Thiamine-binding periplasmic protein |
| P08331_CN16_ECOLI | 1 | 3.2 | 35 | 3 | 2',3'-cyclic-nucleotide 2'-phosphodiesterase |
| P0A855_TOLB_ECOLI | 1 | 3.2 | 15 | 2 | Protein tolB |
| P05458_PTRA_ECOLI | 1 | 3.1 | 7 | 1 | Protease 3 |
| P23865_PRC_ECOLI | 1 | 2 | 15 | 2 | Tail-specific protease |
| P0AG78_SUBI_ECOLI | 1 | 1.7 | 17 | 2 | Sulfate-binding protein |
| P37329_MODA_ECOLI | 1 | 1.1 | 323 | 4 | Molybdate-binding periplasmic protein |
| P0AEG4_DSBA_ECOLI | 1 | 1 | 9 | 2 | Thiol:disulfide interchange protein dsbA |
| P0ADV1_YHBN_ECOLI | 1 | 0.8 | 7 | 2 | Protein yhbN |
| P0AD96_LIVJ_ECOLI | 1 | 0.7 | 432 | 4 | Leu/Ile/Val-binding protein |
| P0AEM9_FLIY_ECOLI | 1 | 0.4 | 141 | 4 | Cystine-binding periplasmic protein |
| P16700_CYSP_ECOLI | 1 | -0.1 | 68 | 4 | Thiosulfate-binding protein |
| P0AEU0_HISJ_ECOLI | 1 | -0.1 | 135 | 4 | Histidine-binding periplasmic protein |
| P04816_LIVK_ECOLI | 1 | -0.1 | 204 | 4 | Leucine-specific-binding protein |
| P0AG82_PSTS_ECOLI | 1 | -0.1 | 18 | 2 | Phosphate-binding protein pstS |
| P30860_ARTJ_ECOLI | 1 | -0.2 | 164 | 4 | Arginine-binding periplasmic protein 2 |
| P23847_DPPA_ECOLI | 1 | -0.3 | 321 | 4 | Periplasmic dipeptide transport protein |
| P0ABZ6_SURA_ECOLI | 1 | -0.3 | 108 | 3 | Chaperone surA |
| P37902_GLTI_ECOLI | 1 | -0.5 | 72 | 4 | Glutamate/aspartate periplasmic-binding protein |
| P02925_RBSB_ECOLI | 1 | -0.5 | 78 | 4 | D-ribose-binding periplasmic protein |
| P0AE22_APHA_ECOLI | 1 | -0.5 | 5 | 1 | Class B acid phosphatase |
| P36649_CUEO_ECOLI | 1 | -0.5 | 3 | 1 | Blue copper oxidase cueO |
| P0AB24_YCDO_ECOLI | 1 | -0.5 | 4 | 1 | UPF0409 protein ycdO |
| P45523_FKBA_ECOLI | 1 | -0.6 | 61 | 4 | FKBP-type peptidyl-prolyl cis-trans isomerase fkpA |
| P61316_LOLA_ECOLI | 1 | -0.6 | 38 | 4 | Outer-membrane lipoprotein carrier protein |
| P0AAX8_YBIS_ECOLI | 1 | -0.6 | 40 | 4 | Uncharacterized protein ybiS |
| P0AFL3_PPIA_ECOLI | 1 | -0.6 | 14 | 3 | Peptidyl-prolyl cis-trans isomerase A |

| | | | | | |
|---|---|---|---|---|---|
| P30859_ARTI_ECOLI | 1 | -0.8 | 47 | 4 | Arginine-binding periplasmic protein 1 |
| P23843_OPPA_ECOLI | 1 | -0.8 | 976 | 4 | Periplasmic oligopeptide-binding protein |
| P31697_FIMC_ECOLI | 1 | -0.8 | 7 | 1 | Chaperone protein fimC |
| P09551_ARGT_ECOLI | 1 | -0.9 | 66 | 4 | Lysine-arginine-ornithine-binding periplasmic protein |
| P07024_USHA_ECOLI | 1 | -0.9 | 53 | 4 | Protein ushA |
| P0AFH8_OSMY_ECOLI | 1 | -1.3 | 231 | 4 | Osmotically-inducible protein Y |
| P0AEX9_MALE_ECOLI | 1 | -1.5 | 31 | 4 | Maltose-binding periplasmic protein |
| P39172_ZNUA_ECOLI | 1 | -1.5 | 335 | 4 | High-affinity zinc uptake system protein znuA |
| P0AEG6_DSBC_ECOLI | 1 | -1.9 | 22 | 1 | Thiol:disulfide interchange protein dsbC |
| P0AD59_IVY_ECOLI | 1 | -2.1 | 33 | 4 | Inhibitor of vertebrate lysozyme |
| P37648_YHJJ_ECOLI | 1 | -2.1 | 6 | 1 | Protein yhjJ |
| P09394_GLPQ_ECOLI | 1 | -2.3 | 24 | 4 | Glycerophosphoryl diester phosphodiesterase |
| P0AEU7_SKP_ECOLI | 1 | -2.7 | 75 | 4 | Chaperone protein skp |
| P0AGD1_SODC_ECOLI | 1 | -2.9 | 47 | 3 | Superoxide dismutase [Cu-Zn] |
| P0AEE5_DGAL_ECOLI | 1 | -3 | 89 | 4 | D-galactose-binding periplasmic protein |
| P0AG80_UGPB_ECOLI | 1 | -3 | 17 | 2 | sn-glycerol-3-phosphate-binding periplasmic protein ugpB |
| P0ADU5_YGIW_ECOLI | 1 | -3 | 18 | 2 | Protein ygiW |
| P19926_AGP_ECOLI | 1 | -3.1 | 14 | 2 | Glucose-1-phosphatase |
| P0C0V0_DEGP_ECOLI | 1 | -3.9 | 31 | 3 | Protease do |
| P00805_ASPG2_ECOLI | 1 | -4 | 19 | 2 | L-asparaginase 2 |
| P0A9L3_FKBB_ECOLI | 1 | -4 | 4 | 1 | FKBP-type 22 kDa peptidyl-prolyl cis-trans isomerase |
| P33136_OPGG_ECOLI | 1 | -4.1 | 12 | 1 | Glucans biosynthesis protein G |
| P13482_TREA_ECOLI | 1 | -4.1 | 15 | 1 | Periplasmic trehalase |
| P06610_BTUE_ECOLI | 1 | -4.2 | 31 | 2 | Vitamin B12 transport periplasmic protein btuE. |
| P0A862_TPX_ECOLI | 1 | -4.3 | 75 | 4 | Thiol peroxidase |
| P02924_ARAF_ECOLI | 1 | -4.4 | 7 | 1 | L-arabinose-binding periplasmic protein |
| P76193_YNHG_ECOLI | 1 | -4.6 | 89 | 3 | Uncharacterized protein ynhG |
| P0AES9_HDEA_ECOLI | 1 | -4.6 | 26 | 2 | Protein hdeA |
| P39325_YTFQ_ECOLI | 1 | -6.3 | 95 | 4 | ABC transporter periplasmic-binding protein ytfQ |
| P17315_CIRA_ECOLI | 0 | 6.5 | 5 | 1 | Colicin I receptor |
| P05825_FEPA_ECOLI | 0 | 3.6 | 6 | 1 | Ferrienterobactin receptor |
| P75780_FIU_ECOLI | 0 | 3.6 | 13 | 1 | Catecholate siderophore receptor fiu |
| P76177_YDGH_ECOLI | 0 | 3.5 | 92 | 4 | Protein ydgH |

| | | | | | |
|---|---|---|---|---|---|
| P0ABD8_BCCP_ECOLI | 0 | 3.4 | 14 | 1 | Biotin carboxyl carrier protein of acetyl-CoA carboxylase |
| P31133_POTF_ECOLI | 0 | 3.3 | 31 | 3 | Putrescine-binding periplasmic protein |
| P04949_FLIC_ECOLI | 0 | 3.2 | 91 | 4 | Flagellin. |
| P76002_YCGK_ECOLI | 0 | 2.8 | 18 | 4 | Uncharacterized protein ycgK |
| P31828_PQQL_ECOLI | 0 | 2.3 | 6 | 1 | Probable zinc protease pqqL |
| P09169_OMPT_ECOLI | 0 | 1.9 | 6 | 1 | Protease 7 |
| P66948_YFGC_ECOLI | 0 | 1.8 | 8 | 2 | TPR repeat-containing protein yfgC |
| P0ADB1_OSME_ECOLI | 0 | 1.7 | 12 | 2 | Osmotically-inducible lipoprotein E |
| P0A910_OMPA_ECOLI | 0 | 1.4 | 73 | 4 | Outer membrane protein A |
| P76116_YNCE_ECOLI | 0 | 1 | 278 | 4 | Uncharacterized protein yncE |
| P0A7G2_RBFA_ECOLI | 0 | 1 | 8 | 1 | Ribosome-binding factor A |
| P76344_YODA_ECOLI | 0 | 0.3 | 220 | 3 | Metal-binding protein yodA |
| P00954_SYW_ECOLI | 0 | 0.3 | 8 | 1 | Tryptophanyl-tRNA synthetase |
| P0ADN2_YIFE_ECOLI | 0 | -0.5 | 26 | 3 | UPF0438 protein yifE. |
| P0A9D2_GST_ECOLI | 0 | -0.5 | 3 | 1 | Glutathione S-transferase |
| P0A6D7_AROK_ECOLI | 0 | -0.9 | 4 | 1 | Shikimate kinase 1 |
| P00509_AAT_ECOLI | 0 | -1.2 | 29 | 3 | Aspartate aminotransferase |
| P0A8F8_UVRB_ECOLI | 0 | -1.3 | 6 | 1 | UvrABC system protein B |
| P78067_YNJE_ECOLI | 0 | -1.3 | 3 | 1 | Putative thiosulfate sulfurtransferase ynjE |
| P76172_YNFD_ECOLI | 0 | -1.5 | 6 | 1 | Uncharacterized protein ynfD |
| P0A6X7_IHFA_ECOLI | 0 | -1.6 | 5 | 1 | Integration host factor subunit alpha |
| P77499_SUFC_ECOLI | 0 | -1.7 | 8 | 1 | Probable ATP-dependent transporter sufC. |
| P33362_YEHZ_ECOLI | 0 | -1.8 | 17 | 2 | Uncharacterized protein yehZ |
| P06999_K6PF2_ECOLI | 0 | -1.8 | 31 | 1 | 6-phosphofructokinase isozyme 2 |
| P0A877_TRPA_ECOLI | 0 | -1.9 | 23 | 2 | Tryptophan synthase alpha chain |
| P37330_MASZ_ECOLI | 0 | -1.9 | 7 | 1 | Malate synthase G |
| P00934_THRC_ECOLI | 0 | -2 | 39 | 3 | Threonine synthase |
| P23857_PSPE_ECOLI | 0 | -2 | 6 | 2 | Phage shock protein E |
| P0AAT6_YBEB_ECOLI | 0 | -2 | 3 | 1 | Uncharacterized protein ybeB. |
| P13029_CATA_ECOLI | 0 | -2.1 | 13 | 1 | Peroxidase/catalase HPI |
| P0A6H1_CLPX_ECOLI | 0 | -2.1 | 4 | 1 | ATP-dependent Clp protease ATP-binding subunit clpX. |
| P32132_TYPA_ECOLI | 0 | -2.2 | 5 | 1 | GTP-binding protein typA/BipA |
| P76143_YNEB_ECOLI | 0 | -2.2 | 35 | 1 | Uncharacterized aldolase yneB |

| | | | | | |
|---|---|---|---|---|---|
| Q59385_ATCU_ECOLI | 0 | -2.3 | 4 | 1 | Copper-transporting P-type ATPase |
| P06968_DUT_ECOLI | 0 | -2.3 | 3 | 1 | Deoxyuridine 5'-triphosphate nucleotidohydrolase |
| P0AF12_MTNN_ECOLI | 0 | -2.3 | 8 | 1 | MTA/SAH nucleosidase |
| P16659_SYP_ECOLI | 0 | -2.3 | 11 | 1 | Prolyl-tRNA synthetase |
| P37191_GATZ_ECOLI | 0 | -2.4 | 20 | 4 | Putative tagatose 6-phosphate kinase gatZ |
| P76558_MAO2_ECOLI | 0 | -2.4 | 11 | 2 | NADP-dependent malic enzyme |
| P00579_RPOD_ECOLI | 0 | -2.4 | 12 | 2 | RNA polymerase sigma factor rpoD |
| P0AEJ2_ENTC_ECOLI | 0 | -2.4 | 7 | 1 | Isochorismate synthase entC |
| P10371_HIS4_ECOLI | 0 | -2.4 | 5 | 1 | 1-(5-phosphoribosyl)-5-[(5-phosphoribosylamino)methylideneam |
| P0AED0_USPA_ECOLI | 0 | -2.5 | 26 | 4 | Universal stress protein A. |
| P0ADX7_YHHA_ECOLI | 0 | -2.5 | 52 | 3 | Uncharacterized protein yhhA |
| P0ABB4_ATPB_ECOLI | 0 | -2.5 | 8 | 1 | ATP synthase subunit beta |
| P67910_HLDD_ECOLI | 0 | -2.5 | 5 | 1 | ADP-L-glycero-D-manno-heptose-6-epimerase |
| P0AGE6_YIEF_ECOLI | 0 | -2.5 | 9 | 1 | Uncharacterized protein yieF. |
| P33219_YEBF_ECOLI | 0 | -2.6 | 15 | 3 | Protein yebF |
| P69503_APT_ECOLI | 0 | -2.6 | 6 | 1 | Adenine phosphoribosyltransferase |
| P0ACR9_MPRA_ECOLI | 0 | -2.6 | 19 | 1 | Transcriptional repressor mprA |
| P02930_TOLC_ECOLI | 0 | -2.6 | 9 | 1 | Outer membrane protein tolC |
| P39377_IADA_ECOLI | 0 | -2.7 | 9 | 2 | Isoaspartyl dipeptidase |
| P0A9S5_GLDA_ECOLI | 0 | -2.7 | 9 | 1 | Glycerol dehydrogenase |
| P0AB77_KBL_ECOLI | 0 | -2.7 | 4 | 1 | 2-amino-3-ketobutyrate coenzyme A ligase |
| P0AG30_RHO_ECOLI | 0 | -2.7 | 14 | 1 | Transcription termination factor rho |
| P37440_UCPA_ECOLI | 0 | -2.7 | 8 | 1 | Oxidoreductase ucpA |
| P0A6A6_LEU2_ECOLI | 0 | -2.8 | 24 | 4 | 3-isopropylmalate dehydratase large subunit |
| P63284_CLPB_ECOLI | 0 | -2.8 | 21 | 2 | Chaperone protein clpB |
| P0A6N8_EFPL_ECOLI | 0 | -2.8 | 21 | 2 | Elongation factor P-like protein. |
| P61714_RISB_ECOLI | 0 | -2.8 | 11 | 2 | 6,7-dimethyl-8-ribityllumazine synthase |
| P25526_GABD_ECOLI | 0 | -2.8 | 14 | 1 | Succinate-semialdehyde dehydrogenase [NADP+] |
| P0AFW2_RMF_ECOLI | 0 | -2.8 | 22 | 1 | Ribosome modulation factor |
| P0ACE7_HINT_ECOLI | 0 | -2.9 | 3 | 1 | HIT-like protein hinT. |
| P39173_YEAD_ECOLI | 0 | -2.9 | 4 | 1 | UPF0010 protein yeaD |
| P0A7R5_RS10_ECOLI | 0 | -3 | 38 | 3 | 30S ribosomal protein S10. |
| P0A8R4_SLYX_ECOLI | 0 | -3 | 16 | 2 | Protein slyX. |

| | | | | | |
|---|---|---|---|---|---|
| P00550_PTM3C_ECOLI | 0 | -3 | 7 | 1 | PTS system mannitol-specific EIICBA component |
| P75746_YBGL_ECOLI | 0 | -3 | 3 | 1 | UPF0271 protein ybgL. |
| P77735_YAJO_ECOLI | 0 | -3.1 | 14 | 1 | Uncharacterized oxidoreductase yajO |
| P0A9W3_YJJK_ECOLI | 0 | -3.1 | 9 | 1 | Uncharacterized ABC transporter ATP-binding protein yjjK. |
| P0A6J8_DDLA_ECOLI | 0 | -3.2 | 3 | 1 | D-alanine--D-alanine ligase A |
| P10408_SECA_ECOLI | 0 | -3.2 | 3 | 1 | Protein translocase subunit secA. |
| P33221_PURT_ECOLI | 0 | -3.3 | 16 | 4 | Phosphoribosylglycinamide formyltransferase 2 |
| P0AFG3_ODO1_ECOLI | 0 | -3.3 | 11 | 2 | 2-oxoglutarate dehydrogenase E1 component |
| P0ABD3_BFR_ECOLI | 0 | -3.3 | 7 | 1 | Bacterioferritin |
| P24186_FOLD_ECOLI | 0 | -3.3 | 4 | 1 | Bifunctional protein folD |
| P0AB43_YCGL_ECOLI | 0 | -3.3 | 4 | 1 | Uncharacterized protein ycgL. |
| P32157_YIIM_ECOLI | 0 | -3.3 | 7 | 1 | Protein yiiM. |
| P0C0L2_OSMC_ECOLI | 0 | -3.4 | 55 | 3 | Peroxiredoxin osmC |
| P30125_LEU3_ECOLI | 0 | -3.4 | 13 | 2 | 3-isopropylmalate dehydrogenase |
| P15047_ENTA_ECOLI | 0 | -3.4 | 9 | 1 | 2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase |
| P0A6N1_EFTU_ECOLI | 0 | -3.5 | 289 | 4 | Elongation factor Tu |
| P23869_PPIB_ECOLI | 0 | -3.5 | 51 | 4 | Peptidyl-prolyl cis-trans isomerase B |
| P42616_YQJC_ECOLI | 0 | -3.5 | 51 | 4 | Protein yqjC |
| P0AG63_RS17_ECOLI | 0 | -3.5 | 28 | 2 | 30S ribosomal protein S17. |
| P0A6T3_GAL1_ECOLI | 0 | -3.5 | 10 | 1 | Galactokinase |
| P09831_GLTB_ECOLI | 0 | -3.5 | 8 | 1 | Glutamate synthase [NADPH] large chain |
| P37689_GPMI_ECOLI | 0 | -3.5 | 30 | 1 | 2,3-bisphosphoglycerate-independent phosphoglycerate mutase |
| P09151_LEU1_ECOLI | 0 | -3.5 | 7 | 1 | 2-isopropylmalate synthase |
| P0ADZ4_RS15_ECOLI | 0 | -3.5 | 7 | 1 | 30S ribosomal protein S15. |
| P36683_ACON2_ECOLI | 0 | -3.6 | 44 | 4 | Aconitate hydratase 2 |
| P0A9D8_DAPD_ECOLI | 0 | -3.6 | 35 | 4 | 2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-succinyltrans |
| P0ADZ0_RL23_ECOLI | 0 | -3.6 | 23 | 2 | 50S ribosomal protein L23. |
| P21177_FADB_ECOLI | 0 | -3.6 | 9 | 1 | Fatty acid oxidation complex subunit alpha |
| P0AEQ1_GLCG_ECOLI | 0 | -3.6 | 21 | 1 | Protein glcG. |
| P30126_LEUD_ECOLI | 0 | -3.6 | 12 | 1 | 3-isopropylmalate dehydratase small subunit |
| P76243_YEAO_ECOLI | 0 | -3.6 | 4 | 1 | Uncharacterized protein yeaO. |
| Q46868_YQIC_ECOLI | 0 | -3.6 | 4 | 1 | Uncharacterized protein yqiC. |
| P0A9M0_LON_ECOLI | 0 | -3.7 | 30 | 3 | ATP-dependent protease La |

| | | | | | |
|---|---|---|---|---|---|
| P68191_SRA_ECOLI | 0 | -3.7 | 25 | 3 | Stationary-phase-induced ribosome-associated protein |
| P75694_YAHO_ECOLI | 0 | -3.7 | 22 | 3 | UPF0379 protein yahO |
| P62620_ISPG_ECOLI | 0 | -3.7 | 11 | 2 | 4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase |
| P12281_MOEA_ECOLI | 0 | -3.7 | 21 | 2 | Molybdopterin biosynthesis protein moeA. |
| P0ADY7_RL16_ECOLI | 0 | -3.7 | 13 | 2 | 50S ribosomal protein L16. |
| P75691_YAHK_ECOLI | 0 | -3.7 | 33 | 1 | Zinc-type alcohol dehydrogenase-like protein yahK |
| P0AC59_GLRX2_ECOLI | 0 | -3.8 | 55 | 4 | Glutaredoxin-2 |
| P69441_KAD_ECOLI | 0 | -3.8 | 116 | 4 | Adenylate kinase |
| P21151_FADA_ECOLI | 0 | -3.8 | 28 | 2 | 3-ketoacyl-CoA thiolase |
| P18843_NADE_ECOLI | 0 | -3.8 | 12 | 2 | NH(3)-dependent NAD(+) synthetase |
| P0AFU8_RISA_ECOLI | 0 | -3.8 | 14 | 2 | Riboflavin synthase alpha chain |
| P76402_YEGP_ECOLI | 0 | -3.8 | 15 | 2 | UPF0339 protein yegP. |
| P0ADS9_YGGN_ECOLI | 0 | -3.8 | 25 | 2 | Uncharacterized protein yggN. |
| P00962_SYQ_ECOLI | 0 | -3.8 | 11 | 1 | Glutaminyl-tRNA synthetase |
| P0C037_YAIE_ECOLI | 0 | -3.8 | 5 | 1 | UPF0345 protein yaiE. |
| P0AAI9_FABD_ECOLI | 0 | -3.9 | 55 | 4 | Malonyl CoA-acyl carrier protein transacylase |
| P0A9Y6_CSPC_ECOLI | 0 | -3.9 | 28 | 3 | Cold shock-like protein cspC |
| P0AAB6_GALF_ECOLI | 0 | -3.9 | 63 | 3 | UTP--glucose-1-phosphate uridylyltransferase |
| P0ACC7_GLMU_ECOLI | 0 | -3.9 | 15 | 3 | Bifunctional protein glmU |
| P25665_METE_ECOLI | 0 | -3.9 | 22 | 2 | 5-methyltetrahydropteroyltriglutamate--homocysteine methyltr |
| P52061_RDGB_ECOLI | 0 | -3.9 | 14 | 2 | Nucleoside-triphosphatase rdgB |
| P0A7L3_RL20_ECOLI | 0 | -3.9 | 18 | 2 | 50S ribosomal protein L20. |
| P0ABB0_ATPA_ECOLI | 0 | -3.9 | 11 | 1 | ATP synthase subunit alpha |
| P31663_PANC_ECOLI | 0 | -3.9 | 17 | 1 | Pantothenate synthetase |
| P08997_MASY_ECOLI | 0 | -4 | 44 | 4 | Malate synthase A |
| P0A7D7_PUR7_ECOLI | 0 | -4 | 51 | 4 | Phosphoribosylaminoimidazole-succinocarboxamide synthase |
| P0A8T7_RPOC_ECOLI | 0 | -4 | 75 | 4 | DNA-directed RNA polymerase subunit beta' |
| P0A7J7_RL11_ECOLI | 0 | -4 | 61 | 3 | 50S ribosomal protein L11. |
| P0AA25_THIO_ECOLI | 0 | -4 | 37 | 2 | Thioredoxin-1 |
| P25516_ACON1_ECOLI | 0 | -4 | 12 | 1 | Aconitate hydratase 1 |
| P69797_PTNAB_ECOLI | 0 | -4.1 | 27 | 4 | PTS system mannose-specific EIIAB component |
| P0A8A0_YEBC_ECOLI | 0 | -4.1 | 40 | 4 | UPF0082 protein yebC. |
| P00370_DHE4_ECOLI | 0 | -4.1 | 27 | 3 | NADP-specific glutamate dehydrogenase |

| | | | | | |
|---|---|---|---|---|---|
| P0A8E7_YAJQ_ECOLI | 0 | -4.1 | 81 | 3 | UPF0234 protein yajQ. |
| P0AB55_YCII_ECOLI | 0 | -4.1 | 26 | 3 | Protein yciI. |
| P0A6K6_DEOB_ECOLI | 0 | -4.1 | 14 | 2 | Phosphopentomutase |
| P0A6R0_FABH_ECOLI | 0 | -4.1 | 21 | 2 | 3-oxoacyl-[acyl-carrier-protein] synthase 3 |
| P0A7E5_PYRG_ECOLI | 0 | -4.1 | 8 | 2 | CTP synthase |
| P0A9P4_TRXB_ECOLI | 0 | -4.1 | 34 | 2 | Thioredoxin reductase |
| P0AAB8_USPD_ECOLI | 0 | -4.1 | 8 | 1 | Universal stress protein D. |
| P00350_6PGD_ECOLI | 0 | -4.2 | 74 | 4 | 6-phosphogluconate dehydrogenase, decarboxylating |
| P0A9Q9_DHAS_ECOLI | 0 | -4.2 | 110 | 3 | Aspartate-semialdehyde dehydrogenase |
| P0A858_TPIS_ECOLI | 0 | -4.2 | 51 | 3 | Triosephosphate isomerase |
| P0A7J0_RIBB_ECOLI | 0 | -4.2 | 7 | 2 | 3,4-dihydroxy-2-butanone 4-phosphate synthase |
| P0AFG0_NUSG_ECOLI | 0 | -4.2 | 17 | 1 | Transcription antitermination protein nusG. |
| P0AFJ1_PHNA_ECOLI | 0 | -4.2 | 7 | 1 | Protein phnA. |
| P0A6L0_DEOC_ECOLI | 0 | -4.3 | 61 | 4 | Deoxyribose-phosphate aldolase |
| P0A6P1_EFTS_ECOLI | 0 | -4.3 | 79 | 4 | Elongation factor Ts |
| P45578_LUXS_ECOLI | 0 | -4.3 | 41 | 4 | S-ribosylhomocysteine lyase |
| P0AFF6_NUSA_ECOLI | 0 | -4.3 | 30 | 4 | Transcription elongation protein nusA |
| P07012_RF2_ECOLI | 0 | -4.3 | 77 | 4 | Peptide chain release factor 2 |
| P21599_KPYK2_ECOLI | 0 | -4.3 | 32 | 3 | Pyruvate kinase II |
| P0A7F3_PYRI_ECOLI | 0 | -4.3 | 14 | 1 | Aspartate carbamoyltransferase regulatory chain. |
| P0AD24_YEJL_ECOLI | 0 | -4.3 | 3 | 1 | UPF0352 protein yejL. |
| P0A6M8_EFG_ECOLI | 0 | -4.4 | 257 | 4 | Elongation factor G |
| P0AET2_HDEB_ECOLI | 0 | -4.4 | 102 | 4 | Protein hdeB |
| P0AB91_AROG_ECOLI | 0 | -4.4 | 22 | 3 | Phospho-2-dehydro-3-deoxyheptonate aldolase, Phe-sensitive |
| P0A707_IF3_ECOLI | 0 | -4.4 | 37 | 3 | Translation initiation factor IF-3. |
| P0A9X4_MREB_ECOLI | 0 | -4.4 | 26 | 3 | Rod shape-determining protein mreB. |
| P38489_NFNB_ECOLI | 0 | -4.4 | 38 | 3 | Oxygen-insensitive NAD(P)H nitroreductase |
| P12758_UDP_ECOLI | 0 | -4.4 | 40 | 3 | Uridine phosphorylase |
| P0A8W8_YFBU_ECOLI | 0 | -4.4 | 23 | 2 | UPF0304 protein yfbU. |
| P0A6F1_CARA_ECOLI | 0 | -4.4 | 19 | 1 | Carbamoyl-phosphate synthase small chain |
| P27550_ACSA_ECOLI | 0 | -4.5 | 39 | 3 | Acetyl-coenzyme A synthetase |
| P0ABA6_ATPG_ECOLI | 0 | -4.5 | 17 | 2 | ATP synthase gamma chain |
| P0A6H5_HSLU_ECOLI | 0 | -4.5 | 72 | 2 | ATP-dependent hsl protease ATP-binding subunit hslU |

| | | | | | |
|---|---|---|---|---|---|
| P0A763_NDK_ECOLI | 0 | -4.5 | 48 | 2 | Nucleoside diphosphate kinase |
| P0A7U7_RS20_ECOLI | 0 | -4.5 | 33 | 2 | 30S ribosomal protein S20. |
| P0AFM6_PSPA_ECOLI | 0 | -4.5 | 27 | 1 | Phage shock protein A. |
| P15640_PUR2_ECOLI | 0 | -4.5 | 13 | 1 | Phosphoribosylamine--glycine ligase |
| P0AE08_AHPC_ECOLI | 0 | -4.6 | 121 | 4 | Alkyl hydroperoxide reductase subunit C |
| P0ABK5_CYSK_ECOLI | 0 | -4.6 | 258 | 4 | Cysteine synthase A |
| P0AEK2_FABG_ECOLI | 0 | -4.6 | 89 | 4 | 3-oxoacyl-[acyl-carrier-protein] reductase |
| P69783_PTGA_ECOLI | 0 | -4.6 | 135 | 4 | Glucose-specific phosphotransferase enzyme IIA component |
| P0A7K2_RL7_ECOLI | 0 | -4.6 | 47 | 4 | 50S ribosomal protein L7/L12 |
| P29217_YCEH_ECOLI | 0 | -4.6 | 38 | 4 | UPF0502 protein yceH |
| P0ACX3_YDHR_ECOLI | 0 | -4.6 | 23 | 2 | Protein ydhR |
| P62768_YAEH_ECOLI | 0 | -4.6 | 17 | 1 | UPF0325 protein yaeH. |
| P39831_YDFG_ECOLI | 0 | -4.6 | 6 | 1 | NADP-dependent L-serine/L-allo-threonine dehydrogenase ydfG |
| P02359_RS7_ECOLI | 0 | -4.7 | 126 | 4 | 30S ribosomal protein S7. |
| P0A8M6_YEEX_ECOLI | 0 | -4.7 | 38 | 4 | UPF0265 protein yeeX. |
| P77695_GNSB_ECOLI | 0 | -4.7 | 12 | 2 | Protein gnsB. |
| P0A786_PYRB_ECOLI | 0 | -4.7 | 20 | 2 | Aspartate carbamoyltransferase catalytic chain |
| P00547_KHSE_ECOLI | 0 | -4.7 | 5 | 1 | Homoserine kinase |
| P0AA04_PTHP_ECOLI | 0 | -4.8 | 60 | 4 | Phosphocarrier protein HPr |
| P37192_GATY_ECOLI | 0 | -4.8 | 17 | 2 | Tagatose-1,6-bisphosphate aldolase gatY |
| P00957_SYA_ECOLI | 0 | -4.8 | 15 | 1 | Alanyl-tRNA synthetase |
| P0A805_RRF_ECOLI | 0 | -4.9 | 111 | 4 | Ribosome recycling factor |
| P0AE52_BCP_ECOLI | 0 | -4.9 | 39 | 3 | Putative peroxiredoxin bcp |
| P06959_ODP2_ECOLI | 0 | -4.9 | 58 | 2 | Dihydrolipoyllysine-residue acetyltransferase component of p |
| P0ABA4_ATPD_ECOLI | 0 | -4.9 | 4 | 1 | ATP synthase delta chain |
| P63020_GNTY_ECOLI | 0 | -4.9 | 7 | 1 | Protein gntY. |
| P0AGE0_SSB_ECOLI | 0 | -4.9 | 12 | 1 | Single-stranded DNA-binding protein |
| P0ACG1_STPA_ECOLI | 0 | -5 | 57 | 4 | DNA-binding protein stpA |
| P0A9Z1_GLNB_ECOLI | 0 | -5 | 25 | 2 | Nitrogen regulatory protein P-II 1. |
| P0AG18_PUR6_ECOLI | 0 | -5 | 14 | 2 | Phosphoribosylaminoimidazole carboxylase catalytic subunit |
| P0AEK4_FABI_ECOLI | 0 | -5 | 47 | 1 | Enoyl-[acyl-carrier-protein] reductase [NADH] |
| P08200_IDH_ECOLI | 0 | -5.1 | 154 | 4 | Isocitrate dehydrogenase [NADP] |
| P62399_RL5_ECOLI | 0 | -5.1 | 62 | 4 | 50S ribosomal protein L5. |

| | | | | | |
|---|---|---|---|---|---|
| P0A870_TALB_ECOLI | 0 | -5.1 | 67 | 4 | Transaldolase B |
| P69428_TATA_ECOLI | 0 | -5.1 | 16 | 3 | Sec-independent protein translocase protein tatA. |
| P0A8B5_YBAB_ECOLI | 0 | -5.1 | 37 | 3 | UPF0133 protein ybaB. |
| P0AC38_ASPA_ECOLI | 0 | -5.1 | 10 | 1 | Aspartate ammonia-lyase |
| P0A9C5_GLNA_ECOLI | 0 | -5.1 | 18 | 1 | Glutamine synthetase |
| P0ADE6_YGAU_ECOLI | 0 | -5.1 | 19 | 1 | Uncharacterized protein ygaU. |
| P0A7M6_RL29_ECOLI | 0 | -5.2 | 27 | 3 | 50S ribosomal protein L29. |
| P0A850_TIG_ECOLI | 0 | -5.2 | 175 | 3 | Trigger factor |
| P0A9P0_DLDH_ECOLI | 0 | -5.2 | 19 | 2 | Dihydrolipoyl dehydrogenase |
| P16456_SELD_ECOLI | 0 | -5.2 | 13 | 2 | Selenide, water dikinase |
| P26646_YHDH_ECOLI | 0 | -5.2 | 16 | 1 | Protein yhdH. |
| P0A9M8_PTA_ECOLI | 0 | -5.3 | 53 | 4 | Phosphate acetyltransferase |
| P0AG44_RL17_ECOLI | 0 | -5.3 | 82 | 3 | 50S ribosomal protein L17. |
| P0A7R1_RL9_ECOLI | 0 | -5.3 | 107 | 3 | 50S ribosomal protein L9. |
| P0AA16_OMPR_ECOLI | 0 | -5.3 | 26 | 2 | Transcriptional regulatory protein ompR. |
| P46837_YHGF_ECOLI | 0 | -5.3 | 10 | 2 | Protein yhgF. |
| P0A8F0_UPP_ECOLI | 0 | -5.4 | 38 | 3 | Uracil phosphoribosyltransferase |
| P0AG67_RS1_ECOLI | 0 | -5.5 | 77 | 4 | 30S ribosomal protein S1. |
| P05042_FUMC_ECOLI | 0 | -5.5 | 10 | 1 | Fumarate hydratase class II |
| P05791_ILVD_ECOLI | 0 | -5.5 | 17 | 1 | Dihydroxy-acid dehydratase |
| P0AFG6_ODO2_ECOLI | 0 | -5.6 | 23 | 3 | Dihydrolipoyllysine-residue succinyltransferase component of |
| P0AF93_YJGF_ECOLI | 0 | -5.6 | 34 | 3 | UPF0076 protein yjgF. |
| P0A6P9_ENO_ECOLI | 0 | -5.7 | 484 | 4 | Enolase |
| P0ADY3_RL14_ECOLI | 0 | -5.7 | 66 | 4 | 50S ribosomal protein L14. |
| P0A9Q7_ADHE_ECOLI | 0 | -5.7 | 36 | 3 | Aldehyde-alcohol dehydrogenase |
| P0A7A9_IPYR_ECOLI | 0 | -5.7 | 38 | 3 | Inorganic pyrophosphatase |
| P0A7V8_RS4_ECOLI | 0 | -5.7 | 91 | 3 | 30S ribosomal protein S4. |
| P07650_TYPH_ECOLI | 0 | -5.7 | 29 | 3 | Thymidine phosphorylase |
| P00935_METB_ECOLI | 0 | -5.7 | 9 | 1 | Cystathionine gamma-synthase |
| P0A9Q1_ARCA_ECOLI | 0 | -5.8 | 39 | 4 | Aerobic respiration control protein arcA |
| P0A6F9_CH10_ECOLI | 0 | -5.8 | 93 | 4 | 10 kDa chaperonin |
| P62707_GPMA_ECOLI | 0 | -5.8 | 270 | 4 | 2,3-bisphosphoglycerate-dependent phosphoglycerate mutase |
| P0ADS6_YGGE_ECOLI | 0 | -5.8 | 46 | 3 | Uncharacterized protein yggE. |

| | | | | | |
|---|---|---|---|---|---|
| P0ACF8_HNS_ECOLI | 0 | -5.9 | 131 | 4 | DNA-binding protein H-NS |
| P61889_MDH_ECOLI | 0 | -5.9 | 307 | 4 | Malate dehydrogenase |
| P0C018_RL18_ECOLI | 0 | -5.9 | 76 | 4 | 50S ribosomal protein L18. |
| P0A991_ALF1_ECOLI | 0 | -5.9 | 41 | 3 | Fructose-bisphosphate aldolase class 1 |
| P0AEZ3_MIND_ECOLI | 0 | -5.9 | 36 | 3 | Septum site-determining protein minD |
| P0A7V3_RS3_ECOLI | 0 | -5.9 | 35 | 3 | 30S ribosomal protein S3. |
| P05793_ILVC_ECOLI | 0 | -5.9 | 76 | 2 | Ketol-acid reductoisomerase |
| P0A7Z0_RPIA_ECOLI | 0 | -5.9 | 25 | 2 | Ribose-5-phosphate isomerase A |
| P02413_RL15_ECOLI | 0 | -5.9 | 72 | 1 | 50S ribosomal protein L15. |
| P07118_SYV_ECOLI | 0 | -5.9 | 12 | 1 | Valyl-tRNA synthetase |
| P0AF36_YIIU_ECOLI | 0 | -5.9 | 13 | 1 | Uncharacterized protein yiiU. |
| P0A825_GLYA_ECOLI | 0 | -6 | 243 | 4 | Serine hydroxymethyltransferase |
| P68066_GRCA_ECOLI | 0 | -6 | 210 | 4 | Autonomous glycyl radical cofactor. |
| P08839_PT1_ECOLI | 0 | -6 | 184 | 4 | Phosphoenolpyruvate-protein phosphotransferase |
| P0A7V0_RS2_ECOLI | 0 | -6 | 47 | 4 | 30S ribosomal protein S2. |
| P0A7W1_RS5_ECOLI | 0 | -6 | 131 | 4 | 30S ribosomal protein S5. |
| P09373_PFLB_ECOLI | 0 | -6 | 66 | 3 | Formate acetyltransferase 1 |
| P0AA10_RL13_ECOLI | 0 | -6 | 32 | 2 | 50S ribosomal protein L13. |
| P0A7N4_RL32_ECOLI | 0 | -6 | 9 | 2 | 50S ribosomal protein L32. |
| P0A6L2_DAPA_ECOLI | 0 | -6 | 24 | 1 | Dihydrodipicolinate synthase |
| P0A6Y8_DNAK_ECOLI | 0 | -6.1 | 383 | 4 | Chaperone protein dnaK |
| P0A799_PGK_ECOLI | 0 | -6.1 | 189 | 4 | Phosphoglycerate kinase |
| P0A6A3_ACKA_ECOLI | 0 | -6.1 | 35 | 2 | Acetate kinase |
| P00961_SYGB_ECOLI | 0 | -6.1 | 35 | 2 | Glycyl-tRNA synthetase beta subunit |
| P00864_CAPP_ECOLI | 0 | -6.2 | 43 | 4 | Phosphoenolpyruvate carboxylase |
| P0A9B2_G3P1_ECOLI | 0 | -6.2 | 369 | 4 | Glyceraldehyde-3-phosphate dehydrogenase A |
| P0ADG7_IMDH_ECOLI | 0 | -6.2 | 156 | 4 | Inosine-5'-monophosphate dehydrogenase |
| P0AD61_KPYK1_ECOLI | 0 | -6.2 | 177 | 4 | Pyruvate kinase I |
| P0A7L0_RL1_ECOLI | 0 | -6.2 | 71 | 4 | 50S ribosomal protein L1. |
| P60438_RL3_ECOLI | 0 | -6.2 | 57 | 4 | 50S ribosomal protein L3. |
| P0A8L1_SYS_ECOLI | 0 | -6.2 | 35 | 4 | Seryl-tRNA synthetase |
| P0A7M2_RL28_ECOLI | 0 | -6.2 | 22 | 3 | 50S ribosomal protein L28. |
| P60422_RL2_ECOLI | 0 | -6.2 | 55 | 3 | 50S ribosomal protein L2. |

| | | | | | |
|---|---|---|---|---|---|
| P0AG55_RL6_ECOLI | 0 | -6.2 | 119 | 3 | 50S ribosomal protein L6. |
| P0AG59_RS14_ECOLI | 0 | -6.2 | 42 | 3 | 30S ribosomal protein S14. |
| P02358_RS6_ECOLI | 0 | -6.2 | 34 | 3 | 30S ribosomal protein S6 |
| P0A7T3_RS16_ECOLI | 0 | -6.2 | 16 | 2 | 30S ribosomal protein S16. |
| P0AG51_RL30_ECOLI | 0 | -6.2 | 11 | 1 | 50S ribosomal protein L30. |
| P0A9G6_ACEA_ECOLI | 0 | -6.3 | 168 | 4 | Isocitrate lyase |
| P0A6E4_ASSY_ECOLI | 0 | -6.3 | 44 | 4 | Argininosuccinate synthase |
| P0A6F5_CH60_ECOLI | 0 | -6.3 | 563 | 4 | 60 kDa chaperonin |
| P0A705_IF2_ECOLI | 0 | -6.3 | 58 | 3 | Translation initiation factor IF-2. |
| P0A8V2_RPOB_ECOLI | 0 | -6.3 | 36 | 3 | DNA-directed RNA polymerase subunit beta |
| P69913_CSRA_ECOLI | 0 | -6.3 | 21 | 2 | Carbon storage regulator. |
| P0A953_FABB_ECOLI | 0 | -6.3 | 44 | 2 | 3-oxoacyl-[acyl-carrier-protein] synthase 1 |
| P0ACF0_DBHA_ECOLI | 0 | -6.4 | 64 | 4 | DNA-binding protein HU-alpha |
| P60723_RL4_ECOLI | 0 | -6.4 | 75 | 4 | 50S ribosomal protein L4. |
| P0A836_SUCC_ECOLI | 0 | -6.4 | 96 | 4 | Succinyl-CoA synthetase beta chain |
| P0ACF4_DBHB_ECOLI | 0 | -6.4 | 174 | 3 | DNA-binding protein HU-beta |
| P0AC62_GLRX3_ECOLI | 0 | -6.4 | 45 | 3 | Glutaredoxin-3 |
| P06988_HISX_ECOLI | 0 | -6.4 | 69 | 3 | Histidinol dehydrogenase |
| P0A6Y1_IHFB_ECOLI | 0 | -6.4 | 30 | 3 | Integration host factor subunit beta |
| P68679_RS21_ECOLI | 0 | -6.4 | 23 | 3 | 30S ribosomal protein S21. |
| P0AEP3_GALU_ECOLI | 0 | -6.4 | 26 | 2 | UTP--glucose-1-phosphate uridylyltransferase |
| P0A8G9_EX7S_ECOLI | 0 | -6.4 | 12 | 1 | Exodeoxyribonuclease 7 small subunit |
| P0A8I5_TRMB_ECOLI | 0 | -6.4 | 6 | 1 | tRNA |
| P0AGE9_SUCD_ECOLI | 0 | -6.5 | 47 | 4 | Succinyl-CoA ligase [ADP-forming] subunit alpha |
| P0AB14_YCCJ_ECOLI | 0 | -6.5 | 39 | 4 | Uncharacterized protein yccJ. |
| P0AB71_ALF_ECOLI | 0 | -6.5 | 50 | 3 | Fructose-bisphosphate aldolase class 2 |
| P00968_CARB_ECOLI | 0 | -6.5 | 74 | 3 | Carbamoyl-phosphate synthase large chain |
| P0A6Z3_HTPG_ECOLI | 0 | -6.5 | 38 | 3 | Chaperone protein htpG |
| P05055_PNP_ECOLI | 0 | -6.5 | 77 | 3 | Polyribonucleotide nucleotidyltransferase |
| P0A7J3_RL10_ECOLI | 0 | -6.5 | 129 | 3 | 50S ribosomal protein L10 |
| P0A7Z4_RPOA_ECOLI | 0 | -6.5 | 76 | 3 | DNA-directed RNA polymerase subunit alpha |
| P0A7T7_RS18_ECOLI | 0 | -6.5 | 41 | 3 | 30S ribosomal protein S18. |
| P64463_YDFZ_ECOLI | 0 | -6.5 | 36 | 3 | Putative selenoprotein ydfZ. |

| | | | | | |
|---|---|---|---|---|---|
| P21179_CATE_ECOLI | 0 | -6.5 | 18 | 2 | Catalase HPII |
| P0A6F3_GLPK_ECOLI | 0 | -6.5 | 29 | 2 | Glycerol kinase |
| P09832_GLTD_ECOLI | 0 | -6.5 | 52 | 2 | Glutamate synthase [NADPH] small chain |
| P0A9A6_FTSZ_ECOLI | 0 | -6.5 | 32 | 1 | Cell division protein ftsZ. |
| P0A7L8_RL27_ECOLI | 0 | -6.5 | 6 | 1 | 50S ribosomal protein L27. |
| P61175_RL22_ECOLI | 0 | -6.6 | 63 | 4 | 50S ribosomal protein L22. |
| P0A7N1_RL31B_ECOLI | 0 | -6.6 | 23 | 3 | 50S ribosomal protein L31 type B. |
| P0A867_TALA_ECOLI | 0 | -6.6 | 42 | 3 | Transaldolase A |
| P0A7R9_RS11_ECOLI | 0 | -6.6 | 20 | 2 | 30S ribosomal protein S11. |
| P0A7S9_RS13_ECOLI | 0 | -6.6 | 16 | 2 | 30S ribosomal protein S13. |
| P68206_YJBJ_ECOLI | 0 | -6.7 | 104 | 3 | UPF0337 protein yjbJ. |
| P39177_USPG_ECOLI | 0 | -6.7 | 31 | 2 | Universal stress protein G. |
| P0ACW6_YDCH_ECOLI | 0 | -6.7 | 18 | 2 | Uncharacterized protein ydcH. |
| P0A9K3_PHOL_ECOLI | 0 | -6.7 | 7 | 1 | PhoH-like protein. |
| P60624_RL24_ECOLI | 0 | -6.7 | 17 | 1 | 50S ribosomal protein L24. |
| P0ADE8_YGFZ_ECOLI | 0 | -6.7 | 17 | 1 | tRNA-modifying protein ygfZ. |
| P04036_DAPB_ECOLI | 0 | -6.8 | 36 | 3 | Dihydrodipicolinate reductase |
| P0A715_KDSA_ECOLI | 0 | -6.8 | 25 | 3 | 2-dehydro-3-deoxyphosphooctonate aldolase |
| P0A794_PDXJ_ECOLI | 0 | -6.8 | 14 | 3 | Pyridoxine 5'-phosphate synthase |
| P31658_HCHA_ECOLI | 0 | -6.8 | 24 | 2 | Chaperone protein hchA |
| P0A7W7_RS8_ECOLI | 0 | -6.8 | 22 | 2 | 30S ribosomal protein S8. |
| P0A9T0_SERA_ECOLI | 0 | -6.8 | 17 | 2 | D-3-phosphoglycerate dehydrogenase |
| P21513_RNE_ECOLI | 0 | -6.8 | 30 | 1 | Ribonuclease E |
| P0A7D4_PURA_ECOLI | 0 | -6.9 | 44 | 3 | Adenylosuccinate synthetase |
| P27302_TKT1_ECOLI | 0 | -6.9 | 28 | 3 | Transketolase 1 |
| P22256_GABT_ECOLI | 0 | -7 | 29 | 3 | 4-aminobutyrate aminotransferase |
| Q46857_DKGA_ECOLI | 0 | -7 | 18 | 2 | 2,5-diketo-D-gluconic acid reductase A |
| P0A817_METK_ECOLI | 0 | -7 | 20 | 2 | S-adenosylmethionine synthetase |
| P07004_PROA_ECOLI | 0 | -7 | 21 | 2 | Gamma-glutamyl phosphate reductase |
| P15639_PUR9_ECOLI | 0 | -7 | 41 | 2 | Bifunctional purine biosynthesis protein purH |
| P09372_GRPE_ECOLI | 0 | -7 | 32 | 1 | Protein grpE |
| P0A8U6_METJ_ECOLI | 0 | -7 | 13 | 1 | Met repressor |
| P69786_PTGCB_ECOLI | 0 | -7 | 10 | 1 | PTS system glucose-specific EIICB component |

| | | | | | |
|---|---|---|---|---|---|
| P64581_YQJD_ECOLI | 0 | -7 | 14 | 1 | Uncharacterized protein yqjD. |

# Appendix E
# GdCl3 Isolation Protocol, All PP:WCL Protein Isotope Ratios

| Protein | PP | log2ratio | Chr. Features | Reps | Protein Description |
|---|---|---|---|---|---|
| P31550_THIB_ECOLI | 1 | 1.3 | 16 | 1 | Thiamine-binding periplasmic protein |
| P0A855_TOLB_ECOLI | 1 | 0.7 | 31 | 2 | Protein tolB |
| P0AA57_YOBA_ECOLI | 1 | 0.3 | 5 | 1 | Protein yobA |
| P0AB24_YCDO_ECOLI | 1 | 0.3 | 21 | 2 | UPF0409 protein ycdO |
| P0AD96_LIVJ_ECOLI | 1 | 0.3 | 486 | 4 | Leu/Ile/Val-binding protein |
| P0AEM9_FLIY_ECOLI | 1 | 0.2 | 161 | 4 | Cystine-binding periplasmic protein |
| P0AG78_SUBI_ECOLI | 1 | 0.2 | 14 | 2 | Sulfate-binding protein |
| P0AEG4_DSBA_ECOLI | 1 | 0.1 | 27 | 3 | Thiol:disulfide interchange protein dsbA |
| P0AEU0_HISJ_ECOLI | 1 | -0.1 | 199 | 4 | Histidine-binding periplasmic protein |
| P0AFM2_PROX_ECOLI | 1 | -0.3 | 123 | 4 | Glycine betaine-binding periplasmic protein |
| P04816_LIVK_ECOLI | 1 | -0.4 | 203 | 4 | Leucine-specific-binding protein |
| P16700_CYSP_ECOLI | 1 | -0.4 | 147 | 4 | Thiosulfate-binding protein |
| P0AFL3_PPIA_ECOLI | 1 | -0.6 | 34 | 4 | Peptidyl-prolyl cis-trans isomerase A |
| P30860_ARTJ_ECOLI | 1 | -0.6 | 235 | 4 | Arginine-binding periplasmic protein 2 |
| P0C0T5_MEPA_ECOLI | 1 | -0.7 | 4 | 1 | Penicillin-insensitive murein endopeptidase |
| P09551_ARGT_ECOLI | 1 | -0.8 | 139 | 4 | Lysine-arginine-ornithine-binding periplasmic protein |
| P0ABZ6_SURA_ECOLI | 1 | -1 | 87 | 4 | Chaperone surA |
| P0ADV1_YHBN_ECOLI | 1 | -1 | 17 | 1 | Protein yhbN |
| P23847_DPPA_ECOLI | 1 | -1 | 325 | 4 | Periplasmic dipeptide transport protein |
| P0ADA1_TESA_ECOLI | 1 | -1.1 | 9 | 2 | Acyl-CoA thioesterase I |
| P0AFK9_POTD_ECOLI | 1 | -1.1 | 88 | 4 | Spermidine/putrescine-binding periplasmic protein |
| P61316_LOLA_ECOLI | 1 | -1.1 | 41 | 4 | Outer-membrane lipoprotein carrier protein |
| P31697_FIMC_ECOLI | 1 | -1.2 | 6 | 2 | Chaperone protein fimC |
| P37329_MODA_ECOLI | 1 | -1.2 | 219 | 4 | Molybdate-binding periplasmic protein |
| P08331_CN16_ECOLI | 1 | -1.3 | 46 | 4 | 2',3'-cyclic-nucleotide 2'-phosphodiesterase |
| P0AEL6_FEPB_ECOLI | 1 | -1.4 | 9 | 2 | Ferrienterobactin-binding periplasmic protein |
| P30859_ARTI_ECOLI | 1 | -1.4 | 91 | 4 | Arginine-binding periplasmic protein 1 |
| P0AEX9_MALE_ECOLI | 1 | -1.5 | 68 | 4 | Maltose-binding periplasmic protein |
| P21338_RNI_ECOLI | 1 | -1.5 | 8 | 2 | Ribonuclease I |
| P0AG82_PSTS_ECOLI | 1 | -1.6 | 21 | 3 | Phosphate-binding protein pstS |

| | | | | | |
|---|---|---|---|---|---|
| P0AAX8_YBIS_ECOLI | 1 | -1.7 | 48 | 3 | Uncharacterized protein ybiS |
| P75797_GSIB_ECOLI | 1 | -1.7 | 43 | 3 | Glutathione-binding protein gsiB |
| P39265_ALSB_ECOLI | 1 | -1.8 | 6 | 1 | D-allose-binding periplasmic protein |
| P23843_OPPA_ECOLI | 1 | -1.9 | 730 | 4 | Periplasmic oligopeptide-binding protein |
| P36649_CUEO_ECOLI | 1 | -2.1 | 21 | 2 | Blue copper oxidase cueO |
| P37648_YHJJ_ECOLI | 1 | -2.1 | 20 | 3 | Protein yhjJ |
| P40120_OPGD_ECOLI | 1 | -2.1 | 28 | 2 | Glucans biosynthesis protein D |
| P07024_USHA_ECOLI | 1 | -2.3 | 52 | 4 | Protein ushA |
| P0AE22_APHA_ECOLI | 1 | -2.3 | 9 | 1 | Class B acid phosphatase |
| P02925_RBSB_ECOLI | 1 | -2.4 | 81 | 4 | D-ribose-binding periplasmic protein |
| P0AFX9_RSEB_ECOLI | 1 | -2.4 | 9 | 1 | Sigma-E factor regulatory protein rseB |
| P0AFH8_OSMY_ECOLI | 1 | -2.6 | 199 | 4 | Osmotically-inducible protein Y |
| P00634_PPB_ECOLI | 1 | -2.7 | 5 | 1 | Alkaline phosphatase |
| P0AEG6_DSBC_ECOLI | 1 | -2.8 | 25 | 3 | Thiol:disulfide interchange protein dsbC |
| P0AEQ3_GLNH_ECOLI | 1 | -2.9 | 107 | 4 | Glutamine-binding periplasmic protein |
| P0AEE5_DGAL_ECOLI | 1 | -3 | 120 | 4 | D-galactose-binding periplasmic protein |
| P0AEU7_SKP_ECOLI | 1 | -3 | 69 | 3 | Chaperone protein skp |
| P0AG80_UGPB_ECOLI | 1 | -3 | 45 | 4 | sn-glycerol-3-phosphate-binding periplasmic protein ugpB |
| P0AGD1_SODC_ECOLI | 1 | -3 | 40 | 4 | Superoxide dismutase [Cu-Zn] |
| P0AE85_CPXP_ECOLI | 1 | -3.1 | 12 | 2 | Periplasmic protein cpxP |
| P39172_ZNUA_ECOLI | 1 | -3.1 | 262 | 4 | High-affinity zinc uptake system protein znuA |
| P05458_PTRA_ECOLI | 1 | -3.4 | 14 | 2 | Protease 3 |
| P02924_ARAF_ECOLI | 1 | -3.7 | 17 | 1 | L-arabinose-binding periplasmic protein |
| P33363_BGLX_ECOLI | 1 | -3.7 | 31 | 2 | Periplasmic beta-glucosidase |
| P37902_GLTI_ECOLI | 1 | -3.7 | 106 | 4 | Glutamate/aspartate periplasmic-binding protein |
| P0A8X2_YCEI_ECOLI | 1 | -3.9 | 12 | 1 | Protein yceI |
| P76193_YNHG_ECOLI | 1 | -3.9 | 93 | 4 | Uncharacterized protein ynhG |
| P37387_XYLF_ECOLI | 1 | -4 | 3 | 1 | D-xylose-binding periplasmic protein |
| P39176_ERFK_ECOLI | 1 | -4 | 6 | 1 | Protein erfK/srfK |
| P0A9L3_FKBB_ECOLI | 1 | -4.1 | 19 | 3 | FKBP-type 22 kDa peptidyl-prolyl cis-trans isomerase |
| P09394_GLPQ_ECOLI | 1 | -4.2 | 44 | 4 | Glycerophosphoryl diester phosphodiesterase |
| P0AGC3_SLT_ECOLI | 1 | -4.3 | 15 | 2 | Soluble lytic murein transglycosylase |
| P23827_ECOT_ECOLI | 1 | -4.3 | 7 | 1 | Ecotin |

| | | | | | |
|---|---|---|---|---|---|
| P0AD59_IVY_ECOLI | 1 | -4.5 | 32 | 3 | Inhibitor of vertebrate lysozyme |
| P45523_FKBA_ECOLI | 1 | -4.5 | 52 | 4 | FKBP-type peptidyl-prolyl cis-trans isomerase fkpA |
| P13482_TREA_ECOLI | 1 | -4.7 | 50 | 3 | Periplasmic trehalase |
| P76108_YDCS_ECOLI | 1 | -4.7 | 34 | 1 | Putative ABC transporter periplasmic-binding protein ydcS |
| P0A862_TPX_ECOLI | 1 | -4.9 | 89 | 4 | Thiol peroxidase |
| P0C0V0_DEGP_ECOLI | 1 | -4.9 | 36 | 2 | Protease do |
| P06610_BTUE_ECOLI | 1 | -5 | 26 | 3 | Vitamin B12 transport periplasmic protein btuE. |
| P19926_AGP_ECOLI | 1 | -5.1 | 50 | 2 | Glucose-1-phosphatase |
| P0AES9_HDEA_ECOLI | 1 | -5.6 | 65 | 4 | Protein hdeA |
| P39325_YTFQ_ECOLI | 1 | -5.8 | 117 | 3 | ABC transporter periplasmic-binding protein ytfQ |
| P33136_OPGG_ECOLI | 1 | -5.9 | 32 | 2 | Glucans biosynthesis protein G |
| P77754_SPY_ECOLI | 1 | -6.2 | 36 | 2 | Spheroplast protein Y |
| P0ADU5_YGIW_ECOLI | 1 | -6.4 | 10 | 1 | Protein ygiW |
| P07102_PPA_ECOLI | 1 | -6.8 | 9 | 1 | Periplasmic appA protein |
| P77348_MPPA_ECOLI | 1 | -6.9 | 20 | 2 | Periplasmic murein peptide-binding protein |
| P0A9M0_LON_ECOLI | 0 | 6.7 | 43 | 3 | ATP-dependent protease La |
| P04949_FLIC_ECOLI | 0 | 3.1 | 99 | 4 | Flagellin. |
| P75780_FIU_ECOLI | 0 | 3.1 | 9 | 1 | Catecholate siderophore receptor fiu |
| P76002_YCGK_ECOLI | 0 | 3 | 27 | 3 | Uncharacterized protein ycgK |
| P76014_DHAL_ECOLI | 0 | 2 | 4 | 1 | PTS-dependent dihydroxyacetone kinase, ADP-binding subunit d |
| P0AEY5_MDAB_ECOLI | 0 | 1.8 | 16 | 1 | Modulator of drug activity B. |
| P23837_PHOQ_ECOLI | 0 | 1.4 | 4 | 1 | Sensor protein phoQ |
| P0A910_OMPA_ECOLI | 0 | 1 | 54 | 4 | Outer membrane protein A |
| P0ADB1_OSME_ECOLI | 0 | 0.4 | 23 | 4 | Osmotically-inducible lipoprotein E |
| P76177_YDGH_ECOLI | 0 | 0.1 | 73 | 4 | Protein ydgH |
| P64519_YODD_ECOLI | 0 | -0.1 | 4 | 1 | Uncharacterized protein yodD. |
| P22106_ASNB_ECOLI | 0 | -0.2 | 6 | 1 | Asparagine synthetase B [glutamine-hydrolyzing] |
| P0AB46_YMGD_ECOLI | 0 | -0.6 | 5 | 1 | Uncharacterized protein ymgD |
| P12281_MOEA_ECOLI | 0 | -0.8 | 17 | 2 | Molybdopterin biosynthesis protein moeA. |
| P31133_POTF_ECOLI | 0 | -0.8 | 29 | 3 | Putrescine-binding periplasmic protein |
| P76116_YNCE_ECOLI | 0 | -0.9 | 189 | 4 | Uncharacterized protein yncE |
| P04994_EX7L_ECOLI | 0 | -1.3 | 8 | 1 | Exodeoxyribonuclease 7 large subunit |
| P0A6C8_ARGB_ECOLI | 0 | -1.3 | 14 | 2 | Acetylglutamate kinase |

| | | | | | |
|---|---|---|---|---|---|
| P0ADX5_YHFG_ECOLI | 0 | -1.3 | 4 | 1 | Uncharacterized protein yhfG. |
| P0A7U3_RS19_ECOLI | 0 | -1.4 | 17 | 1 | 30S ribosomal protein S19. |
| P0AB89_PUR8_ECOLI | 0 | -1.4 | 18 | 1 | Adenylosuccinate lyase |
| P0A9K7_PHOU_ECOLI | 0 | -1.7 | 4 | 1 | Phosphate transport system protein phoU. |
| P0AFK0_PMBA_ECOLI | 0 | -1.7 | 14 | 2 | Protein pmbA |
| P23857_PSPE_ECOLI | 0 | -1.7 | 21 | 4 | Phage shock protein E |
| P24203_YJIA_ECOLI | 0 | -1.7 | 3 | 1 | Uncharacterized GTP-binding protein yjiA. |
| P0ADV7_YRBC_ECOLI | 0 | -1.8 | 21 | 2 | Protein yrbC |
| P17993_UBIG_ECOLI | 0 | -1.9 | 4 | 1 | 3-demethylubiquinone-9 3-methyltransferase |
| P76055_YDAO_ECOLI | 0 | -1.9 | 4 | 1 | UPF0021 protein ydaO. |
| P77376_YDGJ_ECOLI | 0 | -1.9 | 3 | 1 | Uncharacterized oxidoreductase ydgJ |
| P0AAQ6_YBAA_ECOLI | 0 | -2 | 13 | 1 | Uncharacterized protein ybaA. |
| P63417_YHBS_ECOLI | 0 | -2 | 6 | 1 | Uncharacterized acetyltransferase yhbS |
| P77212_YKGC_ECOLI | 0 | -2 | 3 | 1 | Probable pyridine nucleotide-disulfide oxidoreductase ykgC. |
| P0A8I3_YAAA_ECOLI | 0 | -2.1 | 6 | 1 | UPF0246 protein yaaA. |
| P22939_ISPA_ECOLI | 0 | -2.1 | 4 | 1 | Geranyltranstransferase |
| P30958_MFD_ECOLI | 0 | -2.1 | 9 | 2 | Transcription-repair-coupling factor |
| P0A9S1_FUCO_ECOLI | 0 | -2.2 | 5 | 1 | Lactaldehyde reductase |
| P27830_RFFG_ECOLI | 0 | -2.2 | 4 | 1 | dTDP-glucose 4,6-dehydratase |
| P62617_ISPF_ECOLI | 0 | -2.2 | 4 | 1 | 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase |
| P64581_YQJD_ECOLI | 0 | -2.2 | 8 | 1 | Uncharacterized protein yqjD. |
| P69908_DCEA_ECOLI | 0 | -2.2 | 3 | 1 | Glutamate decarboxylase alpha |
| P02930_TOLC_ECOLI | 0 | -2.3 | 4 | 1 | Outer membrane protein tolC |
| P05042_FUMC_ECOLI | 0 | -2.3 | 18 | 2 | Fumarate hydratase class II |
| P0ABJ1_CYOA_ECOLI | 0 | -2.3 | 3 | 1 | Ubiquinol oxidase subunit 2 |
| P0AFI7_PDXH_ECOLI | 0 | -2.3 | 4 | 1 | Pyridoxamine 5'-phosphate oxidase |
| P31455_YIDR_ECOLI | 0 | -2.3 | 5 | 1 | Uncharacterized protein yidR. |
| P76172_YNFD_ECOLI | 0 | -2.3 | 13 | 2 | Uncharacterized protein ynfD |
| Q47147_YAFJ_ECOLI | 0 | -2.3 | 8 | 1 | Putative glutamine amidotransferase yafJ |
| P00861_DCDA_ECOLI | 0 | -2.4 | 10 | 1 | Diaminopimelate decarboxylase |
| P00888_AROF_ECOLI | 0 | -2.4 | 21 | 2 | Phospho-2-dehydro-3-deoxyheptonate aldolase, Tyr-sensitive |
| P0A7B8_HSLV_ECOLI | 0 | -2.4 | 4 | 1 | ATP-dependent protease hslV |
| P0A873_TRMD_ECOLI | 0 | -2.4 | 8 | 1 | tRNA |

| | | | | | |
|---|---|---|---|---|---|
| P0ADS9_YGGN_ECOLI | 0 | -2.4 | 21 | 2 | Uncharacterized protein yggN. |
| P07762_GLGB_ECOLI | 0 | -2.5 | 5 | 1 | 1,4-alpha-glucan-branching enzyme |
| P0A6J8_DDLA_ECOLI | 0 | -2.5 | 4 | 1 | D-alanine--D-alanine ligase A |
| P0A6L9_HSCB_ECOLI | 0 | -2.5 | 10 | 1 | Co-chaperone protein hscB |
| P0A8R4_SLYX_ECOLI | 0 | -2.5 | 9 | 1 | Protein slyX. |
| P0AGG8_TLDD_ECOLI | 0 | -2.5 | 6 | 1 | Protein tldD. |
| P21367_YCAC_ECOLI | 0 | -2.5 | 6 | 1 | Uncharacterized protein ycaC. |
| P27431_YCFD_ECOLI | 0 | -2.5 | 3 | 1 | Uncharacterized protein ycfD. |
| P0A9H1_MUG_ECOLI | 0 | -2.6 | 9 | 1 | G/U mismatch-specific DNA glycosylase |
| P0AFJ1_PHNA_ECOLI | 0 | -2.6 | 17 | 2 | Protein phnA. |
| P33219_YEBF_ECOLI | 0 | -2.6 | 15 | 3 | Protein yebF |
| P33368_YOHF_ECOLI | 0 | -2.6 | 11 | 1 | Uncharacterized oxidoreductase yohF |
| P37647_KDGK_ECOLI | 0 | -2.6 | 3 | 1 | 2-dehydro-3-deoxygluconokinase |
| P69776_LPP_ECOLI | 0 | -2.6 | 3 | 1 | Major outer membrane lipoprotein |
| P0A749_MURA_ECOLI | 0 | -2.7 | 5 | 1 | UDP-N-acetylglucosamine 1-carboxyvinyltransferase |
| P0A7C2_LEXA_ECOLI | 0 | -2.7 | 4 | 1 | LexA repressor |
| P0A9R4_FER_ECOLI | 0 | -2.7 | 9 | 1 | 2Fe-2S ferredoxin. |
| P0ACC7_GLMU_ECOLI | 0 | -2.7 | 16 | 3 | Bifunctional protein glmU |
| P0ACP7_PURR_ECOLI | 0 | -2.7 | 10 | 1 | HTH-type transcriptional repressor purR |
| P0AD24_YEJL_ECOLI | 0 | -2.7 | 6 | 1 | UPF0352 protein yejL. |
| P0AED7_DAPE_ECOLI | 0 | -2.7 | 7 | 1 | Succinyl-diaminopimelate desuccinylase |
| P60664_HIS6_ECOLI | 0 | -2.7 | 9 | 1 | Imidazole glycerol phosphate synthase subunit hisF |
| P61517_CAN_ECOLI | 0 | -2.7 | 9 | 2 | Carbonic anhydrase 2 |
| P76113_YNCB_ECOLI | 0 | -2.7 | 12 | 1 | Putative NADP-dependent oxidoreductase yncB |
| P08312_SYFA_ECOLI | 0 | -2.8 | 19 | 1 | Phenylalanyl-tRNA synthetase alpha chain |
| P0A6P7_ENGB_ECOLI | 0 | -2.8 | 4 | 1 | Probable GTP-binding protein engB. |
| P33362_YEHZ_ECOLI | 0 | -2.8 | 25 | 3 | Uncharacterized protein yehZ |
| P39173_YEAD_ECOLI | 0 | -2.8 | 33 | 2 | UPF0010 protein yeaD |
| P40191_PDXK_ECOLI | 0 | -2.8 | 6 | 1 | Pyridoxine kinase |
| Q47679_YAFV_ECOLI | 0 | -2.8 | 3 | 1 | UPF0012 hydrolase yafV |
| P00893_ILVI_ECOLI | 0 | -2.9 | 9 | 1 | Acetolactate synthase isozyme 3 large subunit |
| P08660_AK3_ECOLI | 0 | -2.9 | 20 | 2 | Lysine-sensitive aspartokinase 3 |
| P09158_SPEE_ECOLI | 0 | -2.9 | 12 | 2 | Spermidine synthase |

| | | | | | |
|---|---|---|---|---|---|
| P0A972_CSPE_ECOLI | 0 | -2.9 | 13 | 2 | Cold shock-like protein cspE |
| P0ABE2_BOLA_ECOLI | 0 | -2.9 | 6 | 1 | Protein bolA. |
| P0ABU0_MENB_ECOLI | 0 | -2.9 | 8 | 1 | Naphthoate synthase |
| P0AE63_CHAB_ECOLI | 0 | -2.9 | 4 | 1 | Cation transport regulator chaB. |
| P0AGG0_THIL_ECOLI | 0 | -2.9 | 5 | 1 | Thiamine-monophosphate kinase |
| P13000_BIOD_ECOLI | 0 | -2.9 | 9 | 1 | Dethiobiotin synthetase |
| P0ACX5_YDHZ_ECOLI | 0 | -3 | 4 | 1 | Uncharacterized protein ydhZ. |
| P32662_GPH_ECOLI | 0 | -3 | 12 | 1 | Phosphoglycolate phosphatase |
| P63284_CLPB_ECOLI | 0 | -3 | 16 | 2 | Chaperone protein clpB |
| P00909_TRPC_ECOLI | 0 | -3.1 | 28 | 2 | Tryptophan biosynthesis protein trpCF |
| P0ACA3_SSPA_ECOLI | 0 | -3.1 | 10 | 2 | Stringent starvation protein A. |
| P0AGJ5_YFIF_ECOLI | 0 | -3.1 | 7 | 1 | Uncharacterized tRNA/rRNA methyltransferase yfiF |
| P11880_MURF_ECOLI | 0 | -3.1 | 3 | 1 | UDP-N-acetylmuramoyl-tripeptide--D-alanyl-D-alanine ligase |
| P0A6K3_DEF_ECOLI | 0 | -3.2 | 7 | 1 | Peptide deformylase |
| P0AAI5_FABF_ECOLI | 0 | -3.2 | 13 | 1 | 3-oxoacyl-[acyl-carrier-protein] synthase 2 |
| P0AES6_GYRB_ECOLI | 0 | -3.2 | 30 | 3 | DNA gyrase subunit B |
| P17169_GLMS_ECOLI | 0 | -3.2 | 3 | 1 | Glucosamine--fructose-6-phosphate aminotransferase |
| P67095_YFCE_ECOLI | 0 | -3.2 | 6 | 1 | Phosphodiesterase yfcE |
| Q57261_TRUD_ECOLI | 0 | -3.2 | 14 | 1 | tRNA pseudouridine synthase D |
| P0ACR4_YEIE_ECOLI | 0 | -3.3 | 11 | 1 | Uncharacterized HTH-type transcriptional regulator yeiE. |
| P0ADS2_YGFE_ECOLI | 0 | -3.3 | 13 | 1 | Uncharacterized protein ygfE. |
| P37744_RMLA1_ECOLI | 0 | -3.3 | 24 | 2 | Glucose-1-phosphate thymidylyltransferase 1 |
| P60340_TRUB_ECOLI | 0 | -3.3 | 12 | 1 | tRNA pseudouridine synthase B |
| P67826_CUTC_ECOLI | 0 | -3.3 | 6 | 1 | Copper homeostasis protein cutC. |
| P69222_IF1_ECOLI | 0 | -3.3 | 9 | 2 | Translation initiation factor IF-1. |
| P69407_RCSB_ECOLI | 0 | -3.3 | 5 | 1 | Capsular synthesis regulator component B. |
| P69503_APT_ECOLI | 0 | -3.3 | 15 | 1 | Adenine phosphoribosyltransferase |
| P76344_YODA_ECOLI | 0 | -3.3 | 208 | 4 | Metal-binding protein yodA |
| P00934_THRC_ECOLI | 0 | -3.4 | 88 | 4 | Threonine synthase |
| P09147_GALE_ECOLI | 0 | -3.4 | 18 | 2 | UDP-glucose 4-epimerase |
| P0A763_NDK_ECOLI | 0 | -3.4 | 67 | 4 | Nucleoside diphosphate kinase |
| P0A955_ALKH_ECOLI | 0 | -3.4 | 39 | 3 | KHG/KDPG aldolase |
| P0AEH5_ELAB_ECOLI | 0 | -3.4 | 11 | 1 | Protein elaB. |

| | | | | | |
|---|---|---|---|---|---|
| P20099_BISC_ECOLI | 0 | -3.4 | 8 | 2 | Biotin sulfoxide reductase |
| P24215_UXUA_ECOLI | 0 | -3.4 | 10 | 1 | Mannonate dehydratase |
| P24232_HMP_ECOLI | 0 | -3.4 | 4 | 1 | Flavohemoprotein |
| P37051_PURU_ECOLI | 0 | -3.4 | 8 | 1 | Formyltetrahydrofolate deformylase |
| P45955_YBGF_ECOLI | 0 | -3.4 | 16 | 2 | Uncharacterized protein ybgF |
| P64455_YDCY_ECOLI | 0 | -3.4 | 10 | 2 | Uncharacterized protein ydcY. |
| P69829_PTSN_ECOLI | 0 | -3.4 | 9 | 1 | Nitrogen regulatory protein |
| P76223_YNJB_ECOLI | 0 | -3.4 | 13 | 2 | Protein ynjB. |
| P04951_KDSB_ECOLI | 0 | -3.5 | 7 | 2 | 3-deoxy-manno-octulosonate cytidylyltransferase |
| P04995_EX1_ECOLI | 0 | -3.5 | 8 | 2 | Exodeoxyribonuclease I |
| P0A9C0_GLPA_ECOLI | 0 | -3.5 | 17 | 2 | Anaerobic glycerol-3-phosphate dehydrogenase subunit A |
| P0AFR4_YCIO_ECOLI | 0 | -3.5 | 19 | 2 | Uncharacterized protein yciO. |
| P24178_YFFB_ECOLI | 0 | -3.5 | 10 | 1 | Protein yffB. |
| P25524_CODA_ECOLI | 0 | -3.5 | 15 | 2 | Cytosine deaminase |
| P30136_THIC_ECOLI | 0 | -3.5 | 12 | 2 | Thiamine biosynthesis protein thiC. |
| P37685_ALDB_ECOLI | 0 | -3.5 | 4 | 1 | Aldehyde dehydrogenase B |
| P69797_PTNAB_ECOLI | 0 | -3.5 | 38 | 4 | PTS system mannose-specific EIIAB component |
| P00363_FRDA_ECOLI | 0 | -3.6 | 4 | 1 | Fumarate reductase flavoprotein subunit |
| P07004_PROA_ECOLI | 0 | -3.6 | 30 | 2 | Gamma-glutamyl phosphate reductase |
| P0A6N4_EFP_ECOLI | 0 | -3.6 | 7 | 1 | Elongation factor P |
| P0A8N7_SYK3_ECOLI | 0 | -3.6 | 4 | 1 | Putative lysyl-tRNA synthetase |
| P0ACB2_HEM2_ECOLI | 0 | -3.6 | 18 | 1 | Delta-aminolevulinic acid dehydratase |
| P0ACE7_HINT_ECOLI | 0 | -3.6 | 4 | 1 | HIT-like protein hinT. |
| P0ADY1_PPID_ECOLI | 0 | -3.6 | 4 | 1 | Peptidyl-prolyl cis-trans isomerase D |
| P0AFM6_PSPA_ECOLI | 0 | -3.6 | 28 | 2 | Phage shock protein A. |
| P11454_ENTF_ECOLI | 0 | -3.6 | 15 | 1 | Enterobactin synthetase component F |
| P30126_LEUD_ECOLI | 0 | -3.6 | 28 | 3 | 3-isopropylmalate dehydratase small subunit |
| P33602_NUOG_ECOLI | 0 | -3.6 | 3 | 1 | NADH-quinone oxidoreductase subunit G |
| P46837_YHGF_ECOLI | 0 | -3.6 | 41 | 3 | Protein yhgF. |
| P77695_GNSB_ECOLI | 0 | -3.6 | 15 | 3 | Protein gnsB. |
| P78067_YNJE_ECOLI | 0 | -3.6 | 22 | 2 | Putative thiosulfate sulfurtransferase ynjE |
| Q2EEQ2_RL362_ECOLI | 0 | -3.6 | 11 | 1 | 50S ribosomal protein L36 2. |
| Q46851_YGHZ_ECOLI | 0 | -3.6 | 20 | 3 | Uncharacterized protein yghZ. |

| | | | | | |
|---|---|---|---|---|---|
| P00579_RPOD_ECOLI | 0 | -3.7 | 21 | 1 | RNA polymerase sigma factor rpoD |
| P0A9W3_YJJK_ECOLI | 0 | -3.7 | 24 | 1 | Uncharacterized ABC transporter ATP-binding protein yjjK. |
| P0ADP9_YIHD_ECOLI | 0 | -3.7 | 26 | 3 | Protein yihD. |
| P10408_SECA_ECOLI | 0 | -3.7 | 46 | 4 | Protein translocase subunit secA. |
| P37191_GATZ_ECOLI | 0 | -3.7 | 39 | 4 | Putative tagatose 6-phosphate kinase gatZ |
| P39180_AG43_ECOLI | 0 | -3.7 | 28 | 2 | Antigen 43 |
| P63224_GMHA_ECOLI | 0 | -3.7 | 27 | 1 | Phosphoheptose isomerase |
| P75874_YCCU_ECOLI | 0 | -3.7 | 21 | 2 | Uncharacterized protein yccU. |
| P00448_SODM_ECOLI | 0 | -3.8 | 72 | 2 | Superoxide dismutase [Mn] |
| P00960_SYGA_ECOLI | 0 | -3.8 | 10 | 1 | Glycyl-tRNA synthetase alpha subunit |
| P05194_AROD_ECOLI | 0 | -3.8 | 13 | 1 | 3-dehydroquinate dehydratase |
| P06986_HIS8_ECOLI | 0 | -3.8 | 15 | 2 | Histidinol-phosphate aminotransferase |
| P0A6A6_LEU2_ECOLI | 0 | -3.8 | 23 | 2 | 3-isopropylmalate dehydratase large subunit |
| P0A6N1_EFTU_ECOLI | 0 | -3.8 | 255 | 4 | Elongation factor Tu |
| P0A6N8_EFPL_ECOLI | 0 | -3.8 | 8 | 2 | Elongation factor P-like protein. |
| P0A9T4_TAS_ECOLI | 0 | -3.8 | 17 | 1 | Protein tas. |
| P17846_CYSI_ECOLI | 0 | -3.8 | 26 | 1 | Sulfite reductase [NADPH] hemoprotein beta-component |
| P31120_GLMM_ECOLI | 0 | -3.8 | 27 | 3 | Phosphoglucosamine mutase |
| P33018_YEIG_ECOLI | 0 | -3.8 | 10 | 2 | Esterase yeiG |
| P33221_PURT_ECOLI | 0 | -3.8 | 15 | 2 | Phosphoribosylglycinamide formyltransferase 2 |
| P37759_RFBB_ECOLI | 0 | -3.8 | 27 | 2 | dTDP-glucose 4,6-dehydratase |
| P40874_MTOX_ECOLI | 0 | -3.8 | 10 | 1 | N-methyl-L-tryptophan oxidase |
| P67080_YGGS_ECOLI | 0 | -3.8 | 6 | 1 | UPF0001 protein yggS. |
| P69441_KAD_ECOLI | 0 | -3.8 | 104 | 3 | Adenylate kinase |
| P75949_NAGZ_ECOLI | 0 | -3.8 | 7 | 2 | Beta-hexosaminidase |
| P77454_GLSA1_ECOLI | 0 | -3.8 | 9 | 1 | Glutaminase 1 |
| P77499_SUFC_ECOLI | 0 | -3.8 | 11 | 1 | Probable ATP-dependent transporter sufC. |
| P05791_ILVD_ECOLI | 0 | -3.9 | 14 | 2 | Dihydroxy-acid dehydratase |
| P0A744_MSRA_ECOLI | 0 | -3.9 | 12 | 1 | Peptide methionine sulfoxide reductase msrA |
| P0A8A0_YEBC_ECOLI | 0 | -3.9 | 13 | 1 | UPF0082 protein yebC. |
| P0A9K3_PHOL_ECOLI | 0 | -3.9 | 23 | 3 | PhoH-like protein. |
| P0AAB4_UBID_ECOLI | 0 | -3.9 | 16 | 1 | 3-octaprenyl-4-hydroxybenzoate carboxy-lyase |
| P0AES0_GSP_ECOLI | 0 | -3.9 | 15 | 2 | Bifunctional glutathionylspermidine synthetase/amidase |

| | | | | | |
|---|---|---|---|---|---|
| P23908_ARGE_ECOLI | 0 | -3.9 | 4 | 1 | Acetylornithine deacetylase |
| P67910_HLDD_ECOLI | 0 | -3.9 | 26 | 2 | ADP-L-glycero-D-manno-heptose-6-epimerase |
| P00509_AAT_ECOLI | 0 | -4 | 104 | 2 | Aspartate aminotransferase |
| P08244_PYRF_ECOLI | 0 | -4 | 12 | 1 | Orotidine 5'-phosphate decarboxylase |
| P08373_MURB_ECOLI | 0 | -4 | 6 | 2 | UDP-N-acetylenolpyruvoylglucosamine reductase |
| P0A7S9_RS13_ECOLI | 0 | -4 | 50 | 4 | 30S ribosomal protein S13. |
| P0A867_TALA_ECOLI | 0 | -4 | 72 | 2 | Transaldolase A |
| P0ABU5_ELBB_ECOLI | 0 | -4 | 30 | 2 | Enhancing lycopene biosynthesis protein 2 |
| P0ADU2_YGIN_ECOLI | 0 | -4 | 9 | 1 | Protein ygiN. |
| P0AGD3_SODF_ECOLI | 0 | -4 | 20 | 1 | Superoxide dismutase [Fe] |
| P28861_FENR_ECOLI | 0 | -4 | 9 | 1 | Ferredoxin--NADP reductase |
| P37192_GATY_ECOLI | 0 | -4 | 23 | 2 | Tagatose-1,6-bisphosphate aldolase gatY |
| P77541_PRPB_ECOLI | 0 | -4 | 26 | 4 | Methylisocitrate lyase |
| P00452_RIR1_ECOLI | 0 | -4.1 | 14 | 2 | Ribonucleoside-diphosphate reductase 1 subunit alpha |
| P0AB80_ILVE_ECOLI | 0 | -4.1 | 16 | 3 | Branched-chain-amino-acid aminotransferase |
| P0AG18_PUR6_ECOLI | 0 | -4.1 | 10 | 2 | Phosphoribosylaminoimidazole carboxylase catalytic subunit |
| P14900_MURD_ECOLI | 0 | -4.1 | 9 | 2 | UDP-N-acetylmuramoylalanine--D-glutamate ligase |
| P30125_LEU3_ECOLI | 0 | -4.1 | 51 | 2 | 3-isopropylmalate dehydrogenase |
| P45578_LUXS_ECOLI | 0 | -4.1 | 55 | 4 | S-ribosylhomocysteine lyase |
| P64463_YDFZ_ECOLI | 0 | -4.1 | 46 | 4 | Putative selenoprotein ydfZ. |
| P75694_YAHO_ECOLI | 0 | -4.1 | 82 | 4 | UPF0379 protein yahO |
| P05020_PYRC_ECOLI | 0 | -4.2 | 32 | 2 | Dihydroorotase |
| P07012_RF2_ECOLI | 0 | -4.2 | 46 | 4 | Peptide chain release factor 2 |
| P0A7U7_RS20_ECOLI | 0 | -4.2 | 32 | 3 | 30S ribosomal protein S20. |
| P0A7Z4_RPOA_ECOLI | 0 | -4.2 | 91 | 2 | DNA-directed RNA polymerase subunit alpha |
| P0AAI9_FABD_ECOLI | 0 | -4.2 | 47 | 3 | Malonyl CoA-acyl carrier protein transacylase |
| P0AD10_YECJ_ECOLI | 0 | -4.2 | 8 | 1 | Uncharacterized protein yecJ. |
| P15640_PUR2_ECOLI | 0 | -4.2 | 55 | 4 | Phosphoribosylamine--glycine ligase |
| P24171_DCP_ECOLI | 0 | -4.2 | 41 | 2 | Peptidyl-dipeptidase dcp |
| P37760_RFBD_ECOLI | 0 | -4.2 | 4 | 1 | dTDP-4-dehydrorhamnose reductase |
| P52697_6PGL_ECOLI | 0 | -4.2 | 21 | 3 | 6-phosphogluconolactonase |
| P07003_POXB_ECOLI | 0 | -4.3 | 23 | 1 | Pyruvate dehydrogenase [cytochrome] |
| P08957_T1MK_ECOLI | 0 | -4.3 | 7 | 1 | Type I restriction enzyme EcoKI M protein |

| | | | | | |
|---|---|---|---|---|---|
| P0A800_RPOZ_ECOLI | 0 | -4.3 | 19 | 3 | DNA-directed RNA polymerase subunit omega |
| P0A8G3_UXAC_ECOLI | 0 | -4.3 | 31 | 2 | Uronate isomerase |
| P0A8S1_ICIA_ECOLI | 0 | -4.3 | 7 | 1 | Chromosome initiation inhibitor |
| P0A8T7_RPOC_ECOLI | 0 | -4.3 | 152 | 3 | DNA-directed RNA polymerase subunit beta' |
| P0A993_F16P_ECOLI | 0 | -4.3 | 15 | 2 | Fructose-1,6-bisphosphatase |
| P0A9C3_GALM_ECOLI | 0 | -4.3 | 24 | 2 | Aldose 1-epimerase |
| P0ABB4_ATPB_ECOLI | 0 | -4.3 | 45 | 3 | ATP synthase subunit beta |
| P0ACX3_YDHR_ECOLI | 0 | -4.3 | 34 | 3 | Protein ydhR |
| P0ADE8_YGFZ_ECOLI | 0 | -4.3 | 24 | 1 | tRNA-modifying protein ygfZ. |
| P0AE18_AMPM_ECOLI | 0 | -4.3 | 13 | 1 | Methionine aminopeptidase |
| P0AG59_RS14_ECOLI | 0 | -4.3 | 25 | 2 | 30S ribosomal protein S14. |
| P0AG84_YGHA_ECOLI | 0 | -4.3 | 31 | 2 | Uncharacterized oxidoreductase yghA |
| P10378_ENTE_ECOLI | 0 | -4.3 | 5 | 1 | Enterobactin synthetase component E |
| P30177_YBIB_ECOLI | 0 | -4.3 | 16 | 2 | Uncharacterized protein ybiB. |
| P30178_YBIC_ECOLI | 0 | -4.3 | 17 | 1 | Uncharacterized oxidoreductase ybiC |
| P36938_PGM_ECOLI | 0 | -4.3 | 41 | 2 | Phosphoglucomutase |
| P00547_KHSE_ECOLI | 0 | -4.4 | 25 | 3 | Homoserine kinase |
| P08142_ILVB_ECOLI | 0 | -4.4 | 15 | 2 | Acetolactate synthase isozyme 1 large subunit |
| P08178_PUR5_ECOLI | 0 | -4.4 | 25 | 1 | Phosphoribosylformylglycinamidine cyclo-ligase |
| P0A6D7_AROK_ECOLI | 0 | -4.4 | 12 | 1 | Shikimate kinase 1 |
| P0A6R0_FABH_ECOLI | 0 | -4.4 | 31 | 3 | 3-oxoacyl-[acyl-carrier-protein] synthase 3 |
| P0A7Z0_RPIA_ECOLI | 0 | -4.4 | 39 | 4 | Ribose-5-phosphate isomerase A |
| P0AAN9_IRAP_ECOLI | 0 | -4.4 | 8 | 1 | Protein iraP. |
| P0ACG1_STPA_ECOLI | 0 | -4.4 | 56 | 2 | DNA-binding protein stpA |
| P0ADA3_NLPD_ECOLI | 0 | -4.4 | 8 | 2 | Lipoprotein nlpD |
| P13033_GLPB_ECOLI | 0 | -4.4 | 7 | 1 | Anaerobic glycerol-3-phosphate dehydrogenase subunit B |
| P21362_YCIF_ECOLI | 0 | -4.4 | 21 | 2 | Protein yciF. |
| P60757_HIS1_ECOLI | 0 | -4.4 | 26 | 2 | ATP phosphoribosyltransferase |
| P76569_YFGD_ECOLI | 0 | -4.4 | 6 | 1 | Uncharacterized protein yfgD. |
| P0A6W9_GSH1_ECOLI | 0 | -4.5 | 9 | 1 | Glutamate--cysteine ligase |
| P0A8M3_SYT_ECOLI | 0 | -4.5 | 24 | 1 | Threonyl-tRNA synthetase |
| P0A8M6_YEEX_ECOLI | 0 | -4.5 | 37 | 3 | UPF0265 protein yeeX. |
| P0A8W8_YFBU_ECOLI | 0 | -4.5 | 13 | 2 | UPF0304 protein yfbU. |

| | | | | | |
|---|---|---|---|---|---|
| P0AAB6_GALF_ECOLI | 0 | -4.5 | 94 | 4 | UTP--glucose-1-phosphate uridylyltransferase |
| P0AB14_YCCJ_ECOLI | 0 | -4.5 | 30 | 2 | Uncharacterized protein yccJ. |
| P0ABT2_DPS_ECOLI | 0 | -4.5 | 18 | 2 | DNA protection during starvation protein |
| P0AC13_DHPS_ECOLI | 0 | -4.5 | 11 | 1 | Dihydropteroate synthase |
| P0AC81_LGUL_ECOLI | 0 | -4.5 | 6 | 1 | Lactoylglutathione lyase |
| P0ADY3_RL14_ECOLI | 0 | -4.5 | 35 | 4 | 50S ribosomal protein L14. |
| P0AE01_TRMJ_ECOLI | 0 | -4.5 | 11 | 2 | tRNA |
| P0AGE0_SSB_ECOLI | 0 | -4.5 | 26 | 4 | Single-stranded DNA-binding protein |
| P0C018_RL18_ECOLI | 0 | -4.5 | 60 | 4 | 50S ribosomal protein L18. |
| P17115_GUTQ_ECOLI | 0 | -4.5 | 9 | 1 | Protein gutQ. |
| P21888_SYC_ECOLI | 0 | -4.5 | 47 | 2 | Cysteinyl-tRNA synthetase |
| P25665_METE_ECOLI | 0 | -4.5 | 120 | 3 | 5-methyltetrahydropteroyltriglutamate--homocysteine methyltr |
| P27434_YFGA_ECOLI | 0 | -4.5 | 12 | 1 | Uncharacterized HTH-type transcriptional regulator yfgA. |
| P30749_MOAE_ECOLI | 0 | -4.5 | 9 | 1 | Molybdopterin-converting factor subunit 2 |
| P56262_DLHH_ECOLI | 0 | -4.5 | 16 | 2 | Putative carboxymethylenebutenolidase |
| Q46856_YQHD_ECOLI | 0 | -4.5 | 30 | 2 | Alcohol dehydrogenase yqhD |
| P06960_OTC2_ECOLI | 0 | -4.6 | 7 | 1 | Ornithine carbamoyltransferase chain F |
| P0A6B7_ISCS_ECOLI | 0 | -4.6 | 15 | 1 | Cysteine desulfurase |
| P0A6Q3_FABA_ECOLI | 0 | -4.6 | 24 | 2 | 3-hydroxydecanoyl-[acyl-carrier-protein] dehydratase |
| P0A6T3_GAL1_ECOLI | 0 | -4.6 | 12 | 1 | Galactokinase |
| P0A729_YCEF_ECOLI | 0 | -4.6 | 6 | 1 | Maf-like protein yceF. |
| P0A7E5_PYRG_ECOLI | 0 | -4.6 | 39 | 2 | CTP synthase |
| P0A870_TALB_ECOLI | 0 | -4.6 | 120 | 4 | Transaldolase B |
| P0A8L1_SYS_ECOLI | 0 | -4.6 | 63 | 4 | Seryl-tRNA synthetase |
| P0AAV6_YBGS_ECOLI | 0 | -4.6 | 18 | 3 | Uncharacterized protein ybgS |
| P0ABB0_ATPA_ECOLI | 0 | -4.6 | 38 | 3 | ATP synthase subunit alpha |
| P0AC41_DHSA_ECOLI | 0 | -4.6 | 9 | 1 | Succinate dehydrogenase flavoprotein subunit |
| P0ACA7_YLIJ_ECOLI | 0 | -4.6 | 16 | 2 | Uncharacterized GST-like protein yliJ. |
| P0AEK4_FABI_ECOLI | 0 | -4.6 | 66 | 4 | Enoyl-[acyl-carrier-protein] reductase [NADH] |
| P10121_FTSY_ECOLI | 0 | -4.6 | 11 | 2 | Cell division protein ftsY. |
| P11447_ARLY_ECOLI | 0 | -4.6 | 57 | 1 | Argininosuccinate lyase |
| P18843_NADE_ECOLI | 0 | -4.6 | 43 | 3 | NH(3)-dependent NAD(+) synthetase |
| P31130_YDEI_ECOLI | 0 | -4.6 | 11 | 1 | Uncharacterized protein ydeI |

| | | | | | |
|---|---|---|---|---|---|
| P60438_RL3_ECOLI | 0 | -4.6 | 58 | 3 | 50S ribosomal protein L3. |
| P68191_SRA_ECOLI | 0 | -4.6 | 28 | 4 | Stationary-phase-induced ribosome-associated protein |
| P75849_YCBL_ECOLI | 0 | -4.6 | 8 | 1 | Uncharacterized protein ycbL |
| P76143_YNEB_ECOLI | 0 | -4.6 | 46 | 4 | Uncharacterized aldolase yneB |
| P76536_YFEX_ECOLI | 0 | -4.6 | 9 | 2 | Uncharacterized protein yfeX. |
| Q46829_BGLA_ECOLI | 0 | -4.6 | 10 | 1 | 6-phospho-beta-glucosidase bglA |
| P00946_MANA_ECOLI | 0 | -4.7 | 35 | 3 | Mannose-6-phosphate isomerase |
| P00957_SYA_ECOLI | 0 | -4.7 | 69 | 2 | Alanyl-tRNA synthetase |
| P0A6D3_AROA_ECOLI | 0 | -4.7 | 25 | 2 | 3-phosphoshikimate 1-carboxyvinyltransferase |
| P0A6G7_CLPP_ECOLI | 0 | -4.7 | 16 | 3 | ATP-dependent Clp protease proteolytic subunit |
| P0A6H1_CLPX_ECOLI | 0 | -4.7 | 11 | 1 | ATP-dependent Clp protease ATP-binding subunit clpX. |
| P0A7R1_RL9_ECOLI | 0 | -4.7 | 92 | 4 | 50S ribosomal protein L9. |
| P0A877_TRPA_ECOLI | 0 | -4.7 | 88 | 4 | Tryptophan synthase alpha chain |
| P0A879_TRPB_ECOLI | 0 | -4.7 | 48 | 3 | Tryptophan synthase beta chain |
| P0A881_TRPR_ECOLI | 0 | -4.7 | 6 | 1 | Trp operon repressor. |
| P0A8B5_YBAB_ECOLI | 0 | -4.7 | 32 | 2 | UPF0133 protein ybaB. |
| P0A998_FTNA_ECOLI | 0 | -4.7 | 19 | 1 | Ferritin-1 |
| P0A9X4_MREB_ECOLI | 0 | -4.7 | 49 | 4 | Rod shape-determining protein mreB. |
| P0ABK5_CYSK_ECOLI | 0 | -4.7 | 368 | 4 | Cysteine synthase A |
| P0AE52_BCP_ECOLI | 0 | -4.7 | 40 | 3 | Putative peroxiredoxin bcp |
| P0AED0_USPA_ECOLI | 0 | -4.7 | 45 | 4 | Universal stress protein A. |
| P14407_FUMB_ECOLI | 0 | -4.7 | 10 | 1 | Fumarate hydratase class I, anaerobic |
| P16456_SELD_ECOLI | 0 | -4.7 | 20 | 1 | Selenide, water dikinase |
| P37689_GPMI_ECOLI | 0 | -4.7 | 47 | 3 | 2,3-bisphosphoglycerate-independent phosphoglycerate mutase |
| P60595_HIS5_ECOLI | 0 | -4.7 | 12 | 1 | Imidazole glycerol phosphate synthase subunit hisH |
| P77444_SUFS_ECOLI | 0 | -4.7 | 6 | 1 | Cysteine desulfurase |
| P06999_K6PF2_ECOLI | 0 | -4.8 | 35 | 3 | 6-phosphofructokinase isozyme 2 |
| P0A8E7_YAJQ_ECOLI | 0 | -4.8 | 89 | 4 | UPF0234 protein yajQ. |
| P0A9Q5_ACCD_ECOLI | 0 | -4.8 | 15 | 2 | Acetyl-coenzyme A carboxylase carboxyl transferase subunit b |
| P0AA04_PTHP_ECOLI | 0 | -4.8 | 63 | 4 | Phosphocarrier protein HPr |
| P0ABP8_DEOD_ECOLI | 0 | -4.8 | 39 | 1 | Purine nucleoside phosphorylase deoD-type |
| P0AC69_GLRX4_ECOLI | 0 | -4.8 | 16 | 1 | Glutaredoxin-4 |
| P0ACR9_MPRA_ECOLI | 0 | -4.8 | 26 | 2 | Transcriptional repressor mprA |

| | | | | | |
|---|---|---|---|---|---|
| P0AET2_HDEB_ECOLI | 0 | -4.8 | 163 | 4 | Protein hdeB |
| P0AFF6_NUSA_ECOLI | 0 | -4.8 | 39 | 3 | Transcription elongation protein nusA |
| P0AG07_RPE_ECOLI | 0 | -4.8 | 20 | 1 | Ribulose-phosphate 3-epimerase |
| P15639_PUR9_ECOLI | 0 | -4.8 | 57 | 4 | Bifunctional purine biosynthesis protein purH |
| P33195_GCSP_ECOLI | 0 | -4.8 | 13 | 1 | Glycine dehydrogenase [decarboxylating] |
| P37349_DHAM_ECOLI | 0 | -4.8 | 4 | 1 | PTS-dependent dihydroxyacetone kinase, phosphotransferase su |
| P37773_MPL_ECOLI | 0 | -4.8 | 8 | 1 | UDP-N-acetylmuramate:L-alanyl-gamma-D-glutamyl-meso-diaminop |
| P38489_NFNB_ECOLI | 0 | -4.8 | 70 | 4 | Oxygen-insensitive NAD(P)H nitroreductase |
| P60422_RL2_ECOLI | 0 | -4.8 | 77 | 4 | 50S ribosomal protein L2. |
| P69783_PTGA_ECOLI | 0 | -4.8 | 108 | 4 | Glucose-specific phosphotransferase enzyme IIA component |
| P76316_DCYD_ECOLI | 0 | -4.8 | 17 | 1 | D-cysteine desulfhydrase |
| P04805_SYE_ECOLI | 0 | -4.9 | 48 | 2 | Glutamyl-tRNA synthetase |
| P09030_EX3_ECOLI | 0 | -4.9 | 22 | 2 | Exodeoxyribonuclease III |
| P09151_LEU1_ECOLI | 0 | -4.9 | 13 | 1 | 2-isopropylmalate synthase |
| P0A7A2_GPMB_ECOLI | 0 | -4.9 | 3 | 1 | Probable phosphoglycerate mutase gpmB |
| P0A9D8_DAPD_ECOLI | 0 | -4.9 | 122 | 2 | 2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-succinyltrans |
| P0A9Y6_CSPC_ECOLI | 0 | -4.9 | 46 | 4 | Cold shock-like protein cspC |
| P0AC62_GLRX3_ECOLI | 0 | -4.9 | 24 | 2 | Glutaredoxin-3 |
| P0AEP3_GALU_ECOLI | 0 | -4.9 | 65 | 3 | UTP--glucose-1-phosphate uridylyltransferase |
| P0AEZ3_MIND_ECOLI | 0 | -4.9 | 27 | 4 | Septum site-determining protein minD |
| P0AF96_YJGK_ECOLI | 0 | -4.9 | 16 | 2 | Uncharacterized protein yjgK. |
| P0AG55_RL6_ECOLI | 0 | -4.9 | 98 | 3 | 50S ribosomal protein L6. |
| P15288_PEPD_ECOLI | 0 | -4.9 | 60 | 2 | Aminoacyl-histidine dipeptidase |
| P16703_CYSM_ECOLI | 0 | -4.9 | 18 | 3 | Cysteine synthase B |
| P21165_PEPQ_ECOLI | 0 | -4.9 | 18 | 2 | Xaa-Pro dipeptidase |
| P23869_PPIB_ECOLI | 0 | -4.9 | 80 | 4 | Peptidyl-prolyl cis-trans isomerase B |
| P28635_METQ_ECOLI | 0 | -4.9 | 12 | 1 | D-methionine-binding lipoprotein metQ |
| P32132_TYPA_ECOLI | 0 | -4.9 | 35 | 3 | GTP-binding protein typA/BipA |
| P52061_RDGB_ECOLI | 0 | -4.9 | 9 | 1 | Nucleoside-triphosphatase rdgB |
| P60560_GUAC_ECOLI | 0 | -4.9 | 28 | 4 | GMP reductase |
| P61714_RISB_ECOLI | 0 | -4.9 | 24 | 3 | 6,7-dimethyl-8-ribityllumazine synthase |
| P76402_YEGP_ECOLI | 0 | -4.9 | 57 | 3 | UPF0339 protein yegP. |
| P06149_DLD_ECOLI | 0 | -5 | 3 | 1 | D-lactate dehydrogenase |

| | | | | | |
|---|---|---|---|---|---|
| P0A734_MINE_ECOLI | 0 | -5 | 3 | 1 | Cell division topological specificity factor. |
| P0A794_PDXJ_ECOLI | 0 | -5 | 24 | 1 | Pyridoxine 5'-phosphate synthase |
| P0A7D7_PUR7_ECOLI | 0 | -5 | 78 | 4 | Phosphoribosylaminoimidazole-succinocarboxamide synthase |
| P0A7R5_RS10_ECOLI | 0 | -5 | 42 | 4 | 30S ribosomal protein S10. |
| P0AEZ1_METF_ECOLI | 0 | -5 | 49 | 2 | 5,10-methylenetetrahydrofolate reductase |
| P36683_ACON2_ECOLI | 0 | -5 | 130 | 3 | Aconitate hydratase 2 |
| P75863_YCBX_ECOLI | 0 | -5 | 7 | 1 | Uncharacterized protein ycbX. |
| P06959_ODP2_ECOLI | 0 | -5.1 | 91 | 4 | Dihydrolipoyllysine-residue acetyltransferase component of p |
| P07118_SYV_ECOLI | 0 | -5.1 | 62 | 2 | Valyl-tRNA synthetase |
| P0A6V1_GLGC_ECOLI | 0 | -5.1 | 33 | 2 | Glucose-1-phosphate adenylyltransferase |
| P0A7J0_RIBB_ECOLI | 0 | -5.1 | 17 | 1 | 3,4-dihydroxy-2-butanone 4-phosphate synthase |
| P0A805_RRF_ECOLI | 0 | -5.1 | 91 | 4 | Ribosome recycling factor |
| P0A8V0_RNZ_ECOLI | 0 | -5.1 | 3 | 1 | Ribonuclease Z |
| P0AB77_KBL_ECOLI | 0 | -5.1 | 7 | 1 | 2-amino-3-ketobutyrate coenzyme A ligase |
| P39353_YJHC_ECOLI | 0 | -5.1 | 4 | 1 | Putative oxidoreductase yjhC |
| P75823_LTAE_ECOLI | 0 | -5.1 | 24 | 1 | Low specificity L-threonine aldolase |
| P77735_YAJO_ECOLI | 0 | -5.1 | 33 | 4 | Uncharacterized oxidoreductase yajO |
| P0A8N3_SYK1_ECOLI | 0 | -5.2 | 24 | 2 | Lysyl-tRNA synthetase |
| P0AB71_ALF_ECOLI | 0 | -5.2 | 127 | 4 | Fructose-bisphosphate aldolase class 2 |
| P0ACF8_HNS_ECOLI | 0 | -5.2 | 158 | 4 | DNA-binding protein H-NS |
| P0AE88_CPXR_ECOLI | 0 | -5.2 | 18 | 2 | Transcriptional regulatory protein cpxR. |
| P0AGJ9_SYY_ECOLI | 0 | -5.2 | 31 | 2 | Tyrosyl-tRNA synthetase |
| P22256_GABT_ECOLI | 0 | -5.2 | 75 | 4 | 4-aminobutyrate aminotransferase |
| P27848_YIGL_ECOLI | 0 | -5.2 | 15 | 1 | Uncharacterized protein yigL. |
| P0A6M8_EFG_ECOLI | 0 | -5.3 | 394 | 4 | Elongation factor G |
| P0AB28_YCED_ECOLI | 0 | -5.3 | 8 | 1 | Uncharacterized protein yceD |
| P0AET8_HDHA_ECOLI | 0 | -5.3 | 15 | 2 | 7-alpha-hydroxysteroid dehydrogenase |
| P22255_CYSQ_ECOLI | 0 | -5.3 | 19 | 2 | Protein cysQ. |
| Q46871_YQJH_ECOLI | 0 | -5.3 | 11 | 1 | Uncharacterized protein yqjH. |
| P04079_GUAA_ECOLI | 0 | -5.4 | 42 | 2 | GMP synthase [glutamine-hydrolyzing] |
| P04693_TYRB_ECOLI | 0 | -5.4 | 12 | 1 | Aromatic-amino-acid aminotransferase |
| P06983_HEM3_ECOLI | 0 | -5.4 | 15 | 1 | Porphobilinogen deaminase |
| P0A7M6_RL29_ECOLI | 0 | -5.4 | 32 | 3 | 50S ribosomal protein L29. |

| | | | | | |
|---|---|---|---|---|---|
| P0A7N4_RL32_ECOLI | 0 | -5.4 | 32 | 1 | 50S ribosomal protein L32. |
| P0A7P5_RL34_ECOLI | 0 | -5.4 | 7 | 1 | 50S ribosomal protein L34. |
| P0A7W7_RS8_ECOLI | 0 | -5.4 | 20 | 2 | 30S ribosomal protein S8. |
| P0A825_GLYA_ECOLI | 0 | -5.4 | 309 | 4 | Serine hydroxymethyltransferase |
| P0ABD3_BFR_ECOLI | 0 | -5.4 | 4 | 1 | Bacterioferritin |
| P0AE78_CORC_ECOLI | 0 | -5.4 | 8 | 1 | Magnesium and cobalt efflux protein corC. |
| P0AF93_YJGF_ECOLI | 0 | -5.4 | 33 | 1 | UPF0076 protein yjgF. |
| P23882_FMT_ECOLI | 0 | -5.4 | 10 | 1 | Methionyl-tRNA formyltransferase |
| P39315_YTFG_ECOLI | 0 | -5.4 | 5 | 1 | Uncharacterized oxidoreductase ytfG |
| P75691_YAHK_ECOLI | 0 | -5.4 | 55 | 3 | Zinc-type alcohol dehydrogenase-like protein yahK |
| P00561_AK1H_ECOLI | 0 | -5.5 | 6 | 1 | Bifunctional aspartokinase/homoserine dehydrogenase I |
| P0A717_KPRS_ECOLI | 0 | -5.5 | 18 | 2 | Ribose-phosphate pyrophosphokinase |
| P0A759_NAGB_ECOLI | 0 | -5.5 | 13 | 2 | Glucosamine-6-phosphate deaminase |
| P0A7G6_RECA_ECOLI | 0 | -5.5 | 17 | 2 | Protein recA |
| P0AE08_AHPC_ECOLI | 0 | -5.5 | 155 | 4 | Alkyl hydroperoxide reductase subunit C |
| P0AFG0_NUSG_ECOLI | 0 | -5.5 | 20 | 2 | Transcription antitermination protein nusG. |
| P68919_RL25_ECOLI | 0 | -5.5 | 23 | 2 | 50S ribosomal protein L25. |
| P76015_DHAK_ECOLI | 0 | -5.5 | 8 | 1 | PTS-dependent dihydroxyacetone kinase, dihydroxyacetone-bind |
| P77739_YNIA_ECOLI | 0 | -5.5 | 7 | 1 | Uncharacterized protein yniA. |
| P0A6A3_ACKA_ECOLI | 0 | -5.6 | 50 | 4 | Acetate kinase |
| P0A6P9_ENO_ECOLI | 0 | -5.6 | 545 | 4 | Enolase |
| P0A7K2_RL7_ECOLI | 0 | -5.6 | 37 | 2 | 50S ribosomal protein L7/L12 |
| P0A9A6_FTSZ_ECOLI | 0 | -5.6 | 53 | 2 | Cell division protein ftsZ. |
| P0AA16_OMPR_ECOLI | 0 | -5.6 | 27 | 3 | Transcriptional regulatory protein ompR. |
| P0AB43_YCGL_ECOLI | 0 | -5.6 | 6 | 1 | Uncharacterized protein ycgL. |
| P22259_PPCK_ECOLI | 0 | -5.6 | 60 | 4 | Phosphoenolpyruvate carboxykinase [ATP] |
| P61889_MDH_ECOLI | 0 | -5.6 | 321 | 4 | Malate dehydrogenase |
| P76004_YCGM_ECOLI | 0 | -5.6 | 10 | 2 | Uncharacterized protein ycgM. |
| P0A6F1_CARA_ECOLI | 0 | -5.7 | 63 | 4 | Carbamoyl-phosphate synthase small chain |
| P0A6T5_GCH1_ECOLI | 0 | -5.7 | 14 | 1 | GTP cyclohydrolase 1 |
| P0A7F6_SPED_ECOLI | 0 | -5.7 | 17 | 2 | S-adenosylmethionine decarboxylase proenzyme |
| P0A8F0_UPP_ECOLI | 0 | -5.7 | 52 | 3 | Uracil phosphoribosyltransferase |
| P0A8F8_UVRB_ECOLI | 0 | -5.7 | 15 | 1 | UvrABC system protein B |

| | | | | | |
|---|---|---|---|---|---|
| P0A8W2_SLYA_ECOLI | 0 | -5.7 | 10 | 1 | Transcriptional regulator slyA. |
| P0A9L8_P5CR_ECOLI | 0 | -5.7 | 12 | 3 | Pyrroline-5-carboxylate reductase |
| P23836_PHOP_ECOLI | 0 | -5.7 | 11 | 1 | Transcriptional regulatory protein phoP. |
| P24182_ACCC_ECOLI | 0 | -5.7 | 33 | 3 | Biotin carboxylase |
| P00954_SYW_ECOLI | 0 | -5.8 | 32 | 2 | Tryptophanyl-tRNA synthetase |
| P08997_MASY_ECOLI | 0 | -5.8 | 147 | 4 | Malate synthase A |
| P0A6K6_DEOB_ECOLI | 0 | -5.8 | 51 | 2 | Phosphopentomutase |
| P0A799_PGK_ECOLI | 0 | -5.8 | 299 | 3 | Phosphoglycerate kinase |
| P0A9C5_GLNA_ECOLI | 0 | -5.8 | 57 | 2 | Glutamine synthetase |
| P0AA25_THIO_ECOLI | 0 | -5.8 | 40 | 3 | Thioredoxin-1 |
| P18335_ARGD_ECOLI | 0 | -5.8 | 22 | 2 | Acetylornithine/succinyldiaminopimelate aminotransferase |
| P21889_SYD_ECOLI | 0 | -5.8 | 70 | 1 | Aspartyl-tRNA synthetase |
| P25906_YDBC_ECOLI | 0 | -5.8 | 18 | 1 | Putative oxidoreductase ydbC |
| P27298_OPDA_ECOLI | 0 | -5.8 | 46 | 2 | Oligopeptidase A |
| P37666_TKRA_ECOLI | 0 | -5.8 | 17 | 2 | 2-ketogluconate reductase |
| P08839_PT1_ECOLI | 0 | -5.9 | 241 | 4 | Phosphoenolpyruvate-protein phosphotransferase |
| P0A6L2_DAPA_ECOLI | 0 | -5.9 | 45 | 3 | Dihydrodipicolinate synthase |
| P0A6T1_G6PI_ECOLI | 0 | -5.9 | 66 | 2 | Glucose-6-phosphate isomerase |
| P0A7R9_RS11_ECOLI | 0 | -5.9 | 34 | 2 | 30S ribosomal protein S11. |
| P0A836_SUCC_ECOLI | 0 | -5.9 | 137 | 3 | Succinyl-CoA synthetase beta chain |
| P0A9Q9_DHAS_ECOLI | 0 | -5.9 | 137 | 4 | Aspartate-semialdehyde dehydrogenase |
| P0ABD5_ACCA_ECOLI | 0 | -5.9 | 24 | 2 | Acetyl-coenzyme A carboxylase carboxyl transferase subunit a |
| P0ADG7_IMDH_ECOLI | 0 | -5.9 | 145 | 4 | Inosine-5'-monophosphate dehydrogenase |
| P0ADX7_YHHA_ECOLI | 0 | -5.9 | 47 | 3 | Uncharacterized protein yhhA |
| P13035_GLPD_ECOLI | 0 | -5.9 | 33 | 2 | Aerobic glycerol-3-phosphate dehydrogenase |
| P16659_SYP_ECOLI | 0 | -5.9 | 33 | 1 | Prolyl-tRNA synthetase |
| P29217_YCEH_ECOLI | 0 | -5.9 | 38 | 4 | UPF0502 protein yceH |
| P32157_YIIM_ECOLI | 0 | -5.9 | 12 | 2 | Protein yiiM. |
| P69428_TATA_ECOLI | 0 | -5.9 | 7 | 2 | Sec-independent protein translocase protein tatA. |
| Q46857_DKGA_ECOLI | 0 | -5.9 | 37 | 2 | 2,5-diketo-D-gluconic acid reductase A |
| P07395_SYFB_ECOLI | 0 | -6 | 52 | 2 | Phenylalanyl-tRNA synthetase beta chain |
| P0A705_IF2_ECOLI | 0 | -6 | 61 | 3 | Translation initiation factor IF-2. |
| P0A715_KDSA_ECOLI | 0 | -6 | 61 | 3 | 2-dehydro-3-deoxyphosphooctonate aldolase |

| | | | | | |
|---|---|---|---|---|---|
| P0A796_K6PF1_ECOLI | 0 | -6 | 30 | 1 | 6-phosphofructokinase isozyme 1 |
| P0A7D4_PURA_ECOLI | 0 | -6 | 111 | 4 | Adenylosuccinate synthetase |
| P0A7V3_RS3_ECOLI | 0 | -6 | 63 | 4 | 30S ribosomal protein S3. |
| P0A850_TIG_ECOLI | 0 | -6 | 175 | 4 | Trigger factor |
| P0A858_TPIS_ECOLI | 0 | -6 | 105 | 3 | Triosephosphate isomerase |
| P0A991_ALF1_ECOLI | 0 | -6 | 47 | 4 | Fructose-bisphosphate aldolase class 1 |
| P0A9D2_GST_ECOLI | 0 | -6 | 17 | 2 | Glutathione S-transferase |
| P0A9P4_TRXB_ECOLI | 0 | -6 | 38 | 3 | Thioredoxin reductase |
| P0AB91_AROG_ECOLI | 0 | -6 | 39 | 3 | Phospho-2-dehydro-3-deoxyheptonate aldolase, Phe-sensitive |
| P0AG63_RS17_ECOLI | 0 | -6 | 27 | 2 | 30S ribosomal protein S17. |
| P23721_SERC_ECOLI | 0 | -6 | 68 | 2 | Phosphoserine aminotransferase |
| P32685_YJBD_ECOLI | 0 | -6 | 3 | 1 | Uncharacterized protein yjbD. |
| P37747_GLF_ECOLI | 0 | -6 | 16 | 1 | UDP-galactopyranose mutase |
| P37903_USPF_ECOLI | 0 | -6 | 19 | 2 | Universal stress protein F. |
| P42616_YQJC_ECOLI | 0 | -6 | 29 | 4 | Protein yqjC |
| P68066_GRCA_ECOLI | 0 | -6 | 180 | 4 | Autonomous glycyl radical cofactor. |
| P02413_RL15_ECOLI | 0 | -6.1 | 65 | 4 | 50S ribosomal protein L15. |
| P06612_TOP1_ECOLI | 0 | -6.1 | 41 | 3 | DNA topoisomerase 1 |
| P06721_METC_ECOLI | 0 | -6.1 | 9 | 1 | Cystathionine beta-lyase |
| P0A7S3_RS12_ECOLI | 0 | -6.1 | 28 | 3 | 30S ribosomal protein S12. |
| P0A7T7_RS18_ECOLI | 0 | -6.1 | 31 | 3 | 30S ribosomal protein S18. |
| P0AAC0_USPE_ECOLI | 0 | -6.1 | 28 | 4 | Universal stress protein E. |
| P0ABU2_ENGD_ECOLI | 0 | -6.1 | 30 | 3 | GTP-dependent nucleic acid-binding protein engD. |
| P0AC38_ASPA_ECOLI | 0 | -6.1 | 14 | 1 | Aspartate ammonia-lyase |
| P0ACW6_YDCH_ECOLI | 0 | -6.1 | 8 | 1 | Uncharacterized protein ydcH. |
| P0AG16_PUR1_ECOLI | 0 | -6.1 | 20 | 1 | Amidophosphoribosyltransferase |
| P0AG86_SECB_ECOLI | 0 | -6.1 | 15 | 1 | Protein-export protein secB. |
| P13029_CATA_ECOLI | 0 | -6.1 | 53 | 2 | Peroxidase/catalase HPI |
| P21513_RNE_ECOLI | 0 | -6.1 | 31 | 3 | Ribonuclease E |
| P21599_KPYK2_ECOLI | 0 | -6.1 | 86 | 4 | Pyruvate kinase II |
| P31663_PANC_ECOLI | 0 | -6.1 | 43 | 3 | Pantothenate synthetase |
| P33012_GYRI_ECOLI | 0 | -6.1 | 3 | 1 | DNA gyrase inhibitory protein |
| P06988_HISX_ECOLI | 0 | -6.2 | 100 | 4 | Histidinol dehydrogenase |

| | | | | | |
|---|---|---|---|---|---|
| P07650_TYPH_ECOLI | 0 | -6.2 | 37 | 4 | Thymidine phosphorylase |
| P07813_SYL_ECOLI | 0 | -6.2 | 61 | 2 | Leucyl-tRNA synthetase |
| P09373_PFLB_ECOLI | 0 | -6.2 | 287 | 4 | Formate acetyltransferase 1 |
| P09424_MTLD_ECOLI | 0 | -6.2 | 57 | 3 | Mannitol-1-phosphate 5-dehydrogenase |
| P0A7J7_RL11_ECOLI | 0 | -6.2 | 69 | 3 | 50S ribosomal protein L11. |
| P0A7V8_RS4_ECOLI | 0 | -6.2 | 96 | 4 | 30S ribosomal protein S4. |
| P0A988_DPO3B_ECOLI | 0 | -6.2 | 15 | 1 | DNA polymerase III subunit beta |
| P0A9K9_SLYD_ECOLI | 0 | -6.2 | 4 | 1 | FKBP-type peptidyl-prolyl cis-trans isomerase slyD |
| P0ACF0_DBHA_ECOLI | 0 | -6.2 | 136 | 3 | DNA-binding protein HU-alpha |
| P0ADI7_YECD_ECOLI | 0 | -6.2 | 5 | 1 | Isochorismatase family protein yecD. |
| P0AG67_RS1_ECOLI | 0 | -6.2 | 148 | 4 | 30S ribosomal protein S1. |
| P11875_SYR_ECOLI | 0 | -6.2 | 18 | 1 | Arginyl-tRNA synthetase |
| P42620_YQJG_ECOLI | 0 | -6.2 | 4 | 1 | Uncharacterized protein yqjG. |
| P61175_RL22_ECOLI | 0 | -6.2 | 80 | 4 | 50S ribosomal protein L22. |
| P76142_YNEA_ECOLI | 0 | -6.2 | 64 | 2 | Uncharacterized protein yneA |
| P76243_YEAO_ECOLI | 0 | -6.2 | 19 | 2 | Uncharacterized protein yeaO. |
| P00350_6PGD_ECOLI | 0 | -6.3 | 168 | 4 | 6-phosphogluconate dehydrogenase, decarboxylating |
| P00968_CARB_ECOLI | 0 | -6.3 | 124 | 4 | Carbamoyl-phosphate synthase large chain |
| P02359_RS7_ECOLI | 0 | -6.3 | 131 | 4 | 30S ribosomal protein S7. |
| P05459_PDXB_ECOLI | 0 | -6.3 | 20 | 3 | Erythronate-4-phosphate dehydrogenase |
| P08200_IDH_ECOLI | 0 | -6.3 | 267 | 4 | Isocitrate dehydrogenase [NADP] |
| P0A6F3_GLPK_ECOLI | 0 | -6.3 | 160 | 3 | Glycerol kinase |
| P0A6H5_HSLU_ECOLI | 0 | -6.3 | 75 | 4 | ATP-dependent hsl protease ATP-binding subunit hslU |
| P0A6L0_DEOC_ECOLI | 0 | -6.3 | 92 | 4 | Deoxyribose-phosphate aldolase |
| P0A6P1_EFTS_ECOLI | 0 | -6.3 | 121 | 2 | Elongation factor Ts |
| P0A6S7_GPDA_ECOLI | 0 | -6.3 | 8 | 1 | Glycerol-3-phosphate dehydrogenase [NAD(P)+] |
| P0A6Y8_DNAK_ECOLI | 0 | -6.3 | 535 | 4 | Chaperone protein dnaK |
| P0A786_PYRB_ECOLI | 0 | -6.3 | 62 | 4 | Aspartate carbamoyltransferase catalytic chain |
| P0A7N1_RL31B_ECOLI | 0 | -6.3 | 20 | 3 | 50S ribosomal protein L31 type B. |
| P0A8V2_RPOB_ECOLI | 0 | -6.3 | 138 | 3 | DNA-directed RNA polymerase subunit beta |
| P0A9B2_G3P1_ECOLI | 0 | -6.3 | 535 | 4 | Glyceraldehyde-3-phosphate dehydrogenase A |
| P0A9D4_CYSE_ECOLI | 0 | -6.3 | 7 | 1 | Serine acetyltransferase |
| P0AD61_KPYK1_ECOLI | 0 | -6.3 | 227 | 4 | Pyruvate kinase I |

| | | | | | |
|---|---|---|---|---|---|
| P0ADS6_YGGE_ECOLI | 0 | -6.3 | 29 | 3 | Uncharacterized protein yggE. |
| P0AES4_GYRA_ECOLI | 0 | -6.3 | 24 | 2 | DNA gyrase subunit A |
| P0AFG8_ODP1_ECOLI | 0 | -6.3 | 99 | 2 | Pyruvate dehydrogenase E1 component |
| P0C0L2_OSMC_ECOLI | 0 | -6.3 | 88 | 2 | Peroxiredoxin osmC |
| P15254_PUR4_ECOLI | 0 | -6.3 | 89 | 2 | Phosphoribosylformylglycinamidine synthase |
| P16681_PHNB_ECOLI | 0 | -6.3 | 21 | 3 | Protein phnB. |
| P23538_PPSA_ECOLI | 0 | -6.3 | 38 | 2 | Phosphoenolpyruvate synthase |
| P26646_YHDH_ECOLI | 0 | -6.3 | 31 | 2 | Protein yhdH. |
| P28304_QOR_ECOLI | 0 | -6.3 | 26 | 2 | Quinone oxidoreductase |
| P46853_YHHX_ECOLI | 0 | -6.3 | 13 | 2 | Putative oxidoreductase yhhX |
| P60624_RL24_ECOLI | 0 | -6.3 | 19 | 2 | 50S ribosomal protein L24. |
| P62707_GPMA_ECOLI | 0 | -6.3 | 345 | 4 | 2,3-bisphosphoglycerate-dependent phosphoglycerate mutase |
| P75913_GHRA_ECOLI | 0 | -6.3 | 5 | 1 | Putative 2-hydroxyacid dehydrogenase ghrA |
| P00370_DHE4_ECOLI | 0 | -6.4 | 89 | 3 | NADP-specific glutamate dehydrogenase |
| P00961_SYGB_ECOLI | 0 | -6.4 | 63 | 3 | Glycyl-tRNA synthetase beta subunit |
| P00962_SYQ_ECOLI | 0 | -6.4 | 34 | 2 | Glutaminyl-tRNA synthetase |
| P05055_PNP_ECOLI | 0 | -6.4 | 91 | 4 | Polyribonucleotide nucleotidyltransferase |
| P0A6E4_ASSY_ECOLI | 0 | -6.4 | 80 | 3 | Argininosuccinate synthase |
| P0A6F5_CH60_ECOLI | 0 | -6.4 | 617 | 4 | 60 kDa chaperonin |
| P0A6W5_GREA_ECOLI | 0 | -6.4 | 12 | 1 | Transcription elongation factor greA |
| P0A7A9_IPYR_ECOLI | 0 | -6.4 | 57 | 4 | Inorganic pyrophosphatase |
| P0A7J3_RL10_ECOLI | 0 | -6.4 | 86 | 4 | 50S ribosomal protein L10 |
| P0A7L0_RL1_ECOLI | 0 | -6.4 | 101 | 4 | 50S ribosomal protein L1. |
| P0A817_METK_ECOLI | 0 | -6.4 | 49 | 3 | S-adenosylmethionine synthetase |
| P0A9Q7_ADHE_ECOLI | 0 | -6.4 | 149 | 3 | Aldehyde-alcohol dehydrogenase |
| P0AB55_YCII_ECOLI | 0 | -6.4 | 27 | 3 | Protein yciI. |
| P0ACJ0_LRP_ECOLI | 0 | -6.4 | 26 | 3 | Leucine-responsive regulatory protein. |
| P0ACY1_YDJA_ECOLI | 0 | -6.4 | 18 | 2 | Protein ydjA. |
| P0ACY3_YEAG_ECOLI | 0 | -6.4 | 51 | 2 | Uncharacterized protein yeaG. |
| P0ADY7_RL16_ECOLI | 0 | -6.4 | 33 | 2 | 50S ribosomal protein L16. |
| P0AEK2_FABG_ECOLI | 0 | -6.4 | 92 | 4 | 3-oxoacyl-[acyl-carrier-protein] reductase |
| P0AEQ1_GLCG_ECOLI | 0 | -6.4 | 16 | 1 | Protein glcG. |
| P0AF36_YIIU_ECOLI | 0 | -6.4 | 36 | 2 | Uncharacterized protein yiiU. |

| | | | | | |
|---|---|---|---|---|---|
| P0AFM4_PSIF_ECOLI | 0 | -6.4 | 21 | 1 | Phosphate starvation-inducible protein psiF |
| P0AG44_RL17_ECOLI | 0 | -6.4 | 64 | 3 | 50S ribosomal protein L17. |
| P25516_ACON1_ECOLI | 0 | -6.4 | 124 | 3 | Aconitate hydratase 1 |
| P25553_ALDA_ECOLI | 0 | -6.4 | 70 | 2 | Aldehyde dehydrogenase A |
| P37330_MASZ_ECOLI | 0 | -6.4 | 90 | 2 | Malate synthase G |
| P60723_RL4_ECOLI | 0 | -6.4 | 71 | 4 | 50S ribosomal protein L4. |
| P69913_CSRA_ECOLI | 0 | -6.4 | 24 | 2 | Carbon storage regulator. |
| P00864_CAPP_ECOLI | 0 | -6.5 | 112 | 4 | Phosphoenolpyruvate carboxylase |
| P09832_GLTD_ECOLI | 0 | -6.5 | 88 | 4 | Glutamate synthase [NADPH] small chain |
| P0A7K6_RL19_ECOLI | 0 | -6.5 | 24 | 3 | 50S ribosomal protein L19. |
| P0A7M2_RL28_ECOLI | 0 | -6.5 | 23 | 3 | 50S ribosomal protein L28. |
| P0A7T3_RS16_ECOLI | 0 | -6.5 | 36 | 4 | 30S ribosomal protein S16. |
| P0A7W1_RS5_ECOLI | 0 | -6.5 | 88 | 3 | 30S ribosomal protein S5. |
| P0A953_FABB_ECOLI | 0 | -6.5 | 71 | 3 | 3-oxoacyl-[acyl-carrier-protein] synthase 1 |
| P0A9M8_PTA_ECOLI | 0 | -6.5 | 76 | 4 | Phosphate acetyltransferase |
| P0ABH7_CISY_ECOLI | 0 | -6.5 | 9 | 1 | Citrate synthase |
| P0ABS1_DKSA_ECOLI | 0 | -6.5 | 14 | 1 | DnaK suppressor protein. |
| P0AC59_GLRX2_ECOLI | 0 | -6.5 | 69 | 4 | Glutaredoxin-2 |
| P0ADE6_YGAU_ECOLI | 0 | -6.5 | 45 | 4 | Uncharacterized protein ygaU. |
| P0ADZ0_RL23_ECOLI | 0 | -6.5 | 18 | 1 | 50S ribosomal protein L23. |
| P0AG30_RHO_ECOLI | 0 | -6.5 | 19 | 1 | Transcription termination factor rho |
| P0AGE9_SUCD_ECOLI | 0 | -6.5 | 64 | 3 | Succinyl-CoA ligase [ADP-forming] subunit alpha |
| P21177_FADB_ECOLI | 0 | -6.5 | 38 | 1 | Fatty acid oxidation complex subunit alpha |
| P27550_ACSA_ECOLI | 0 | -6.5 | 127 | 3 | Acetyl-coenzyme A synthetase |
| P31142_THTM_ECOLI | 0 | -6.5 | 15 | 2 | 3-mercaptopyruvate sulfurtransferase |
| P37095_PEPB_ECOLI | 0 | -6.5 | 23 | 3 | Peptidase B |
| P38038_CYSJ_ECOLI | 0 | -6.5 | 42 | 2 | Sulfite reductase [NADPH] flavoprotein alpha-component |
| P60906_SYH_ECOLI | 0 | -6.5 | 33 | 1 | Histidyl-tRNA synthetase |
| P62399_RL5_ECOLI | 0 | -6.5 | 91 | 4 | 50S ribosomal protein L5. |
| P62768_YAEH_ECOLI | 0 | -6.5 | 24 | 1 | UPF0325 protein yaeH. |
| P76621_CSID_ECOLI | 0 | -6.5 | 16 | 1 | Protein csiD. |
| P02358_RS6_ECOLI | 0 | -6.6 | 54 | 4 | 30S ribosomal protein S6 |
| P04036_DAPB_ECOLI | 0 | -6.6 | 70 | 4 | Dihydrodipicolinate reductase |

| | | | | | |
|---|---|---|---|---|---|
| P05793_ILVC_ECOLI | 0 | -6.6 | 184 | 4 | Ketol-acid reductoisomerase |
| P09831_GLTB_ECOLI | 0 | -6.6 | 115 | 1 | Glutamate synthase [NADPH] large chain |
| P0A6F9_CH10_ECOLI | 0 | -6.6 | 74 | 3 | 10 kDa chaperonin |
| P0A6X7_IHFA_ECOLI | 0 | -6.6 | 13 | 2 | Integration host factor subunit alpha |
| P0A6Y1_IHFB_ECOLI | 0 | -6.6 | 42 | 3 | Integration host factor subunit beta |
| P0A707_IF3_ECOLI | 0 | -6.6 | 44 | 3 | Translation initiation factor IF-3. |
| P0A7F3_PYRI_ECOLI | 0 | -6.6 | 53 | 4 | Aspartate carbamoyltransferase regulatory chain. |
| P0A7N9_RL33_ECOLI | 0 | -6.6 | 26 | 3 | 50S ribosomal protein L33. |
| P0A7X3_RS9_ECOLI | 0 | -6.6 | 25 | 3 | 30S ribosomal protein S9. |
| P0A8M0_SYN_ECOLI | 0 | -6.6 | 22 | 2 | Asparaginyl-tRNA synthetase |
| P0A9Q1_ARCA_ECOLI | 0 | -6.6 | 75 | 3 | Aerobic respiration control protein arcA |
| P0ABD8_BCCP_ECOLI | 0 | -6.6 | 25 | 3 | Biotin carboxyl carrier protein of acetyl-CoA carboxylase |
| P0ADN2_YIFE_ECOLI | 0 | -6.6 | 35 | 2 | UPF0438 protein yifE. |
| P0AFG3_ODO1_ECOLI | 0 | -6.6 | 70 | 1 | 2-oxoglutarate dehydrogenase E1 component |
| P0AFG6_ODO2_ECOLI | 0 | -6.6 | 67 | 3 | Dihydrolipoyllysine-residue succinyltransferase component of |
| P31658_HCHA_ECOLI | 0 | -6.6 | 36 | 1 | Chaperone protein hchA |
| P33570_TKT2_ECOLI | 0 | -6.6 | 51 | 2 | Transketolase 2 |
| P36659_CBPA_ECOLI | 0 | -6.6 | 15 | 2 | Curved DNA-binding protein. |
| P68679_RS21_ECOLI | 0 | -6.6 | 34 | 2 | 30S ribosomal protein S21. |
| P76558_MAO2_ECOLI | 0 | -6.6 | 70 | 4 | NADP-dependent malic enzyme |
| P77581_ASTC_ECOLI | 0 | -6.6 | 7 | 2 | Succinylornithine transaminase |
| P00956_SYI_ECOLI | 0 | -6.7 | 37 | 2 | Isoleucyl-tRNA synthetase |
| P00959_SYM_ECOLI | 0 | -6.7 | 26 | 2 | Methionyl-tRNA synthetase |
| P06715_GSHR_ECOLI | 0 | -6.7 | 17 | 2 | Glutathione reductase |
| P0A7V0_RS2_ECOLI | 0 | -6.7 | 98 | 3 | 30S ribosomal protein S2. |
| P0A9G6_ACEA_ECOLI | 0 | -6.7 | 365 | 4 | Isocitrate lyase |
| P0A9P0_DLDH_ECOLI | 0 | -6.7 | 73 | 3 | Dihydrolipoyl dehydrogenase |
| P0AD33_YFCZ_ECOLI | 0 | -6.7 | 10 | 2 | UPF0381 protein yfcZ. |
| P0AFW2_RMF_ECOLI | 0 | -6.7 | 29 | 4 | Ribosome modulation factor |
| P0AGE6_YIEF_ECOLI | 0 | -6.7 | 14 | 1 | Uncharacterized protein yieF. |
| P21179_CATE_ECOLI | 0 | -6.7 | 89 | 3 | Catalase HPII |
| P25526_GABD_ECOLI | 0 | -6.7 | 60 | 2 | Succinate-semialdehyde dehydrogenase [NADP+] |
| P26616_MAO1_ECOLI | 0 | -6.7 | 26 | 1 | NAD-dependent malic enzyme |

| | | | | | |
|---|---|---|---|---|---|
| P27302_TKT1_ECOLI | 0 | -6.7 | 81 | 3 | Transketolase 1 |
| P35340_AHPF_ECOLI | 0 | -6.7 | 51 | 3 | Alkyl hydroperoxide reductase subunit F |
| P39177_USPG_ECOLI | 0 | -6.7 | 34 | 3 | Universal stress protein G. |
| P68206_YJBJ_ECOLI | 0 | -6.7 | 78 | 4 | UPF0337 protein yjbJ. |
| Q46845_YGHU_ECOLI | 0 | -6.7 | 22 | 2 | Uncharacterized GST-like protein yghU. |
| P09372_GRPE_ECOLI | 0 | -6.8 | 62 | 3 | Protein grpE |
| P0A6Z3_HTPG_ECOLI | 0 | -6.8 | 100 | 4 | Chaperone protein htpG |
| P0A9T0_SERA_ECOLI | 0 | -6.8 | 85 | 4 | D-3-phosphoglycerate dehydrogenase |
| P0ACF4_DBHB_ECOLI | 0 | -6.8 | 121 | 3 | DNA-binding protein HU-beta |
| P0AD49_RAIA_ECOLI | 0 | -6.8 | 13 | 2 | Ribosome-associated inhibitor A |
| P0ADZ4_RS15_ECOLI | 0 | -6.8 | 7 | 1 | 30S ribosomal protein S15. |
| P45395_KDSD_ECOLI | 0 | -6.8 | 9 | 2 | Arabinose 5-phosphate isomerase |
| P69786_PTGCB_ECOLI | 0 | -6.8 | 9 | 1 | PTS system glucose-specific EIICB component |
| Q59385_ATCU_ECOLI | 0 | -6.8 | 23 | 2 | Copper-transporting P-type ATPase |
| P00582_DPO1_ECOLI | 0 | -6.9 | 35 | 1 | DNA polymerase I |
| P0AA10_RL13_ECOLI | 0 | -6.9 | 27 | 2 | 50S ribosomal protein L13. |
| P0AG51_RL30_ECOLI | 0 | -6.9 | 11 | 1 | 50S ribosomal protein L30. |
| P12758_UDP_ECOLI | 0 | -6.9 | 88 | 2 | Uridine phosphorylase |
| P15034_AMPP_ECOLI | 0 | -6.9 | 27 | 1 | Xaa-Pro aminopeptidase |
| P00550_PTM3C_ECOLI | 0 | -7 | 3 | 1 | PTS system mannitol-specific EIICBA component |
| P0A9M2_HPRT_ECOLI | 0 | -7 | 11 | 1 | Hypoxanthine phosphoribosyltransferase |
| P0A9Z1_GLNB_ECOLI | 0 | -7 | 15 | 1 | Nitrogen regulatory protein P-II 1. |
| P0ABA6_ATPG_ECOLI | 0 | -7 | 17 | 1 | ATP synthase gamma chain |
| P0AC53_G6PD_ECOLI | 0 | -7 | 5 | 1 | Glucose-6-phosphate 1-dehydrogenase |
| P0AD12_YEEZ_ECOLI | 0 | -7 | 12 | 1 | Protein yeeZ |
| P0AFU8_RISA_ECOLI | 0 | -7 | 11 | 1 | Riboflavin synthase alpha chain |
| P29745_PEPT_ECOLI | 0 | -7 | 16 | 1 | Peptidase T |
| P39371_YJHT_ECOLI | 0 | -7 | 3 | 1 | Kelch domain-containing protein yjhT |
| P52647_NIFJ_ECOLI | 0 | -7 | 14 | 1 | Probable pyruvate-flavodoxin oxidoreductase |
| P60651_SPEB_ECOLI | 0 | -7 | 9 | 1 | Agmatinase |
| P76550_YFFS_ECOLI | 0 | -7 | 10 | 1 | Uncharacterized protein yffS. |
| P77522_SUFB_ECOLI | 0 | -7 | 13 | 1 | Protein sufB. |

# Appendix F
# H2O Isolation Protocol, All PP:PP Protein Isotope Ratios

| Protein | PP | log2ratio | Chr. Features | Reps | Protein Description |
|---|---|---|---|---|---|
| P0AA57_YOBA_ECOLI | 1 | 2.7 | 12 | 2 | Protein yobA |
| P0AEL6_FEPB_ECOLI | 1 | 2.4 | 7 | 1 | Ferrienterobactin-binding periplasmic protein |
| P0AB24_YCDO_ECOLI | 1 | 1.9 | 52 | 2 | UPF0409 protein ycdO |
| P77754_SPY_ECOLI | 1 | 1.8 | 14 | 2 | Spheroplast protein Y |
| P0ADV1_YHBN_ECOLI | 1 | 1.6 | 15 | 2 | Protein yhbN |
| P77804_YDGA_ECOLI | 1 | 1.2 | 10 | 2 | Protein ydgA |
| P37329_MODA_ECOLI | 1 | 0.5 | 248 | 2 | Molybdate-binding periplasmic protein |
| P0A8X2_YCEI_ECOLI | 1 | 0.4 | 16 | 2 | Protein yceI |
| P0ADU5_YGIW_ECOLI | 1 | 0.4 | 26 | 2 | Protein ygiW |
| P76193_YNHG_ECOLI | 1 | 0.4 | 44 | 2 | Uncharacterized protein ynhG |
| P18956_GGT_ECOLI | 1 | 0.3 | 6 | 1 | Gamma-glutamyltranspeptidase |
| P0AFL3_PPIA_ECOLI | 1 | 0.2 | 29 | 2 | Peptidyl-prolyl cis-trans isomerase A |
| P77348_MPPA_ECOLI | 1 | 0.1 | 32 | 2 | Periplasmic murein peptide-binding protein |
| P13482_TREA_ECOLI | 1 | 0.1 | 44 | 2 | Periplasmic trehalase |
| P0C0V0_DEGP_ECOLI | 1 | 0 | 89 | 2 | Protease do |
| P31697_FIMC_ECOLI | 1 | 0 | 18 | 2 | Chaperone protein fimC |
| P0ADA1_TESA_ECOLI | 1 | 0 | 30 | 2 | Acyl-CoA thioesterase I |
| P76128_DDPA_ECOLI | 1 | 0 | 3 | 1 | Probable D,D-dipeptide-binding periplasmic protein ddpA |
| P00805_ASPG2_ECOLI | 1 | -0.1 | 17 | 2 | L-asparaginase 2 |
| P0AEM9_FLIY_ECOLI | 1 | -0.1 | 194 | 2 | Cystine-binding periplasmic protein |
| P61316_LOLA_ECOLI | 1 | -0.1 | 34 | 2 | Outer-membrane lipoprotein carrier protein |
| P0ABZ6_SURA_ECOLI | 1 | -0.1 | 124 | 2 | Chaperone surA |
| P31550_THIB_ECOLI | 1 | -0.1 | 69 | 2 | Thiamine-binding periplasmic protein |
| P0AD59_IVY_ECOLI | 1 | -0.2 | 44 | 2 | Inhibitor of vertebrate lysozyme |
| P0AFM2_PROX_ECOLI | 1 | -0.2 | 113 | 2 | Glycine betaine-binding periplasmic protein |
| P0AEU7_SKP_ECOLI | 1 | -0.2 | 70 | 2 | Chaperone protein skp |
| P0AD96_LIVJ_ECOLI | 1 | -0.3 | 389 | 2 | Leu/Ile/Val-binding protein |
| P37028_BTUF_ECOLI | 1 | -0.3 | 7 | 1 | Vitamin B12-binding protein |
| P00811_AMPC_ECOLI | 1 | -0.4 | 23 | 2 | Beta-lactamase |
| P0AEG6_DSBC_ECOLI | 1 | -0.4 | 40 | 2 | Thiol:disulfide interchange protein dsbC |

| | | | | | |
|---|---|---|---|---|---|
| P45523_FKBA_ECOLI | 1 | -0.4 | 93 | 2 | FKBP-type peptidyl-prolyl cis-trans isomerase fkpA |
| P40120_OPGD_ECOLI | 1 | -0.4 | 23 | 2 | Glucans biosynthesis protein D |
| P05458_PTRA_ECOLI | 1 | -0.4 | 26 | 2 | Protease 3 |
| P39172_ZNUA_ECOLI | 1 | -0.4 | 205 | 2 | High-affinity zinc uptake system protein znuA |
| P07024_USHA_ECOLI | 1 | -0.4 | 33 | 1 | Protein ushA |
| P0AEG4_DSBA_ECOLI | 1 | -0.5 | 33 | 2 | Thiol:disulfide interchange protein dsbA |
| P75797_GSIB_ECOLI | 1 | -0.5 | 126 | 2 | Glutathione-binding protein gsiB |
| P0AFK9_POTD_ECOLI | 1 | -0.5 | 150 | 2 | Spermidine/putrescine-binding periplasmic protein |
| P0AG78_SUBI_ECOLI | 1 | -0.5 | 15 | 2 | Sulfate-binding protein |
| P26648_SUFI_ECOLI | 1 | -0.5 | 13 | 2 | Protein sufI |
| P0A855_TOLB_ECOLI | 1 | -0.5 | 97 | 2 | Protein tolB |
| P64596_YRAP_ECOLI | 1 | -0.5 | 10 | 2 | Uncharacterized protein yraP |
| P39099_DEGQ_ECOLI | 1 | -0.6 | 33 | 2 | Protease degQ |
| P0AFH8_OSMY_ECOLI | 1 | -0.6 | 206 | 2 | Osmotically-inducible protein Y |
| P23865_PRC_ECOLI | 1 | -0.6 | 50 | 2 | Tail-specific protease |
| P37648_YHJJ_ECOLI | 1 | -0.6 | 45 | 2 | Protein yhjJ |
| P30859_ARTI_ECOLI | 1 | -0.7 | 128 | 2 | Arginine-binding periplasmic protein 1 |
| P33363_BGLX_ECOLI | 1 | -0.7 | 55 | 2 | Periplasmic beta-glucosidase |
| P0AES9_HDEA_ECOLI | 1 | -0.7 | 28 | 2 | Protein hdeA |
| P0AG82_PSTS_ECOLI | 1 | -0.7 | 20 | 2 | Phosphate-binding protein pstS |
| P0A862_TPX_ECOLI | 1 | -0.7 | 49 | 2 | Thiol peroxidase |
| P0AFX9_RSEB_ECOLI | 1 | -0.7 | 7 | 1 | Sigma-E factor regulatory protein rseB |
| P16700_CYSP_ECOLI | 1 | -0.8 | 181 | 2 | Thiosulfate-binding protein |
| P37902_GLTI_ECOLI | 1 | -0.8 | 160 | 2 | Glutamate/aspartate periplasmic-binding protein |
| P33136_OPGG_ECOLI | 1 | -0.8 | 20 | 2 | Glucans biosynthesis protein G |
| P0AGD1_SODC_ECOLI | 1 | -0.8 | 18 | 2 | Superoxide dismutase [Cu-Zn] |
| P0A9L3_FKBB_ECOLI | 1 | -0.9 | 17 | 2 | FKBP-type 22 kDa peptidyl-prolyl cis-trans isomerase |
| P0AEQ3_GLNH_ECOLI | 1 | -0.9 | 195 | 2 | Glutamine-binding periplasmic protein |
| P04816_LIVK_ECOLI | 1 | -0.9 | 190 | 2 | Leucine-specific-binding protein |
| P0AAX8_YBIS_ECOLI | 1 | -0.9 | 71 | 2 | Uncharacterized protein ybiS |
| P09551_ARGT_ECOLI | 1 | -1 | 112 | 2 | Lysine-arginine-ornithine-binding periplasmic protein |
| P23847_DPPA_ECOLI | 1 | -1 | 405 | 2 | Periplasmic dipeptide transport protein |
| P23827_ECOT_ECOLI | 1 | -1 | 25 | 2 | Ecotin |

| | | | | | |
|---|---|---|---|---|---|
| P0AEU0_HISJ_ECOLI | 1 | -1 | 187 | 2 | Histidine-binding periplasmic protein |
| P23843_OPPA_ECOLI | 1 | -1.1 | 650 | 2 | Periplasmic oligopeptide-binding protein |
| P0AGC3_SLT_ECOLI | 1 | -1.1 | 30 | 2 | Soluble lytic murein transglycosylase |
| P06610_BTUE_ECOLI | 1 | -1.1 | 6 | 1 | Vitamin B12 transport periplasmic protein btuE. |
| P08331_CN16_ECOLI | 1 | -1.2 | 59 | 2 | 2',3'-cyclic-nucleotide 2'-phosphodiesterase |
| P0AEX9_MALE_ECOLI | 1 | -1.3 | 85 | 2 | Maltose-binding periplasmic protein |
| P76108_YDCS_ECOLI | 1 | -1.4 | 36 | 2 | Putative ABC transporter periplasmic-binding protein ydcS |
| P39325_YTFQ_ECOLI | 1 | -1.5 | 120 | 2 | ABC transporter periplasmic-binding protein ytfQ |
| P30860_ARTJ_ECOLI | 1 | -1.6 | 183 | 2 | Arginine-binding periplasmic protein 2 |
| P07102_PPA_ECOLI | 1 | -1.7 | 16 | 2 | Periplasmic appA protein |
| P0AG80_UGPB_ECOLI | 1 | -1.7 | 89 | 2 | sn-glycerol-3-phosphate-binding periplasmic protein ugpB |
| P19926_AGP_ECOLI | 1 | -1.8 | 126 | 2 | Glucose-1-phosphatase |
| P0ABK9_NRFA_ECOLI | 1 | -1.9 | 6 | 1 | Cytochrome c-552 |
| P0AEE5_DGAL_ECOLI | 1 | -2.1 | 73 | 2 | D-galactose-binding periplasmic protein |
| P09394_GLPQ_ECOLI | 1 | -2.2 | 52 | 2 | Glycerophosphoryl diester phosphodiesterase |
| P02925_RBSB_ECOLI | 1 | -2.4 | 96 | 2 | D-ribose-binding periplasmic protein |
| P77214_CUSF_ECOLI | 1 | -2.5 | 35 | 2 | Cation efflux system protein cusF |
| P00634_PPB_ECOLI | 1 | -2.6 | 6 | 1 | Alkaline phosphatase |
| P02924_ARAF_ECOLI | 1 | -3.4 | 50 | 2 | L-arabinose-binding periplasmic protein |
| P36649_CUEO_ECOLI | 1 | -3.5 | 60 | 2 | Blue copper oxidase cueO |
| P0AEJ2_ENTC_ECOLI | 0 | 7 | 15 | 2 | Isochorismate synthase entC |
| P69222_IF1_ECOLI | 0 | 7 | 12 | 2 | Translation initiation factor IF-1. |
| P0ADU2_YGIN_ECOLI | 0 | 7 | 4 | 1 | Protein ygiN. |
| P17315_CIRA_ECOLI | 0 | 6.2 | 225 | 2 | Colicin I receptor |
| P75780_FIU_ECOLI | 0 | 5.7 | 72 | 2 | Catecholate siderophore receptor fiu |
| P05825_FEPA_ECOLI | 0 | 4.9 | 202 | 2 | Ferrienterobactin receptor |
| P0AG84_YGHA_ECOLI | 0 | 4.8 | 5 | 1 | Uncharacterized oxidoreductase yghA |
| P76116_YNCE_ECOLI | 0 | 3.9 | 254 | 2 | Uncharacterized protein yncE |
| P75818_YBJP_ECOLI | 0 | 3.5 | 4 | 1 | Uncharacterized lipoprotein ybjP |
| P10378_ENTE_ECOLI | 0 | 3.4 | 3 | 1 | Enterobactin synthetase component E |
| P37636_MDTE_ECOLI | 0 | 3.1 | 7 | 2 | Multidrug resistance protein mdtE |
| P0AB40_YCFR_ECOLI | 0 | 3 | 12 | 2 | UPF0379 protein ycfR |
| P64614_YHCN_ECOLI | 0 | 2.7 | 6 | 1 | UPF0379 protein yhcN |

| | | | | | |
|---|---|---|---|---|---|
| P68066_GRCA_ECOLI | 0 | 2.2 | 116 | 2 | Autonomous glycyl radical cofactor. |
| P68191_SRA_ECOLI | 0 | 2.2 | 10 | 2 | Stationary-phase-induced ribosome-associated protein |
| P0A910_OMPA_ECOLI | 0 | 2.1 | 266 | 2 | Outer membrane protein A |
| P64463_YDFZ_ECOLI | 0 | 2.1 | 21 | 2 | Putative selenoprotein ydfZ. |
| P0A917_OMPX_ECOLI | 0 | 1.8 | 112 | 2 | Outer membrane protein X |
| P33219_YEBF_ECOLI | 0 | 1.8 | 24 | 2 | Protein yebF |
| P0AB38_YCFM_ECOLI | 0 | 1.7 | 11 | 2 | Uncharacterized protein ycfM. |
| P76002_YCGK_ECOLI | 0 | 1.7 | 69 | 2 | Uncharacterized protein ycgK |
| P08506_DACC_ECOLI | 0 | 1.6 | 3 | 1 | D-alanyl-D-alanine carboxypeptidase dacC |
| P36659_CBPA_ECOLI | 0 | 1.5 | 9 | 2 | Curved DNA-binding protein. |
| P76009_EMTA_ECOLI | 0 | 1.5 | 34 | 2 | Endotype membrane-bound lytic murein transglycosylase A |
| P06971_FHUA_ECOLI | 0 | 1.5 | 3 | 1 | Ferrichrome-iron receptor |
| P31554_OSTA_ECOLI | 0 | 1.5 | 3 | 1 | LPS-assembly protein |
| P62707_GPMA_ECOLI | 0 | 1.4 | 129 | 2 | 2,3-bisphosphoglycerate-dependent phosphoglycerate mutase |
| P0AE06_ACRA_ECOLI | 0 | 1.4 | 4 | 1 | Acriflavine resistance protein A |
| P0ADB1_OSME_ECOLI | 0 | 1.3 | 42 | 2 | Osmotically-inducible lipoprotein E |
| P0A7N9_RL33_ECOLI | 0 | 1.3 | 7 | 2 | 50S ribosomal protein L33. |
| P0A905_SLYB_ECOLI | 0 | 1.3 | 27 | 2 | Outer membrane lipoprotein slyB |
| P16869_FHUE_ECOLI | 0 | 1.3 | 17 | 1 | FhuE receptor |
| P0AFW2_RMF_ECOLI | 0 | 1.3 | 3 | 1 | Ribosome modulation factor |
| P0ADT2_YGIB_ECOLI | 0 | 1.3 | 3 | 1 | UPF0441 protein ygiB. |
| P0AG55_RL6_ECOLI | 0 | 1.2 | 44 | 2 | 50S ribosomal protein L6. |
| P37194_SLP_ECOLI | 0 | 1.2 | 14 | 2 | Outer membrane protein slp |
| P0A7Q1_RL35_ECOLI | 0 | 1.2 | 3 | 1 | 50S ribosomal protein L35 |
| P06996_OMPC_ECOLI | 0 | 1.1 | 57 | 2 | Outer membrane protein C |
| P34210_OMPP_ECOLI | 0 | 1.1 | 8 | 2 | Outer membrane protease ompP |
| P0AA04_PTHP_ECOLI | 0 | 1.1 | 24 | 2 | Phosphocarrier protein HPr |
| P0ADX7_YHHA_ECOLI | 0 | 1.1 | 82 | 2 | Uncharacterized protein yhhA |
| P0AF70_YJEI_ECOLI | 0 | 1.1 | 7 | 2 | Uncharacterized protein yjeI |
| P0A9P6_DEAD_ECOLI | 0 | 1.1 | 3 | 1 | Cold-shock DEAD box protein A |
| P30139_THIG_ECOLI | 0 | 1.1 | 9 | 1 | Thiazole biosynthesis protein thiG. |
| P0AEH5_ELAB_ECOLI | 0 | 1 | 17 | 2 | Protein elaB. |
| P09169_OMPT_ECOLI | 0 | 1 | 216 | 2 | Protease 7 |

| | | | | | |
|---|---|---|---|---|---|
| P68919_RL25_ECOLI | 0 | 1 | 24 | 2 | 50S ribosomal protein L25. |
| P0AG51_RL30_ECOLI | 0 | 1 | 19 | 2 | 50S ribosomal protein L30. |
| P0A7G2_RBFA_ECOLI | 0 | 1 | 5 | 1 | Ribosome-binding factor A |
| P0A8B5_YBAB_ECOLI | 0 | 1 | 4 | 1 | UPF0133 protein ybaB. |
| P42616_YQJC_ECOLI | 0 | 1 | 7 | 1 | Protein yqjC |
| P0A7U7_RS20_ECOLI | 0 | 0.9 | 21 | 2 | 30S ribosomal protein S20. |
| P0ACW4_YDCA_ECOLI | 0 | 0.9 | 17 | 2 | Uncharacterized protein ydcA |
| P31130_YDEI_ECOLI | 0 | 0.9 | 11 | 2 | Uncharacterized protein ydeI |
| P0ADE6_YGAU_ECOLI | 0 | 0.9 | 25 | 2 | Uncharacterized protein ygaU. |
| P75820_AMID_ECOLI | 0 | 0.9 | 3 | 1 | N-acetylmuramoyl-L-alanine amidase amiD |
| P0A6X7_IHFA_ECOLI | 0 | 0.9 | 3 | 1 | Integration host factor subunit alpha |
| P0AB43_YCGL_ECOLI | 0 | 0.9 | 4 | 1 | Uncharacterized protein ycgL. |
| P64503_YEBV_ECOLI | 0 | 0.9 | 7 | 1 | Uncharacterized protein yebV. |
| P0ACF4_DBHB_ECOLI | 0 | 0.8 | 10 | 2 | DNA-binding protein HU-beta |
| P45464_YRAM_ECOLI | 0 | 0.8 | 14 | 2 | Uncharacterized protein yraM. |
| P0AEY5_MDAB_ECOLI | 0 | 0.8 | 6 | 1 | Modulator of drug activity B. |
| P0ACF0_DBHA_ECOLI | 0 | 0.7 | 31 | 2 | DNA-binding protein HU-alpha |
| P05459_PDXB_ECOLI | 0 | 0.7 | 11 | 2 | Erythronate-4-phosphate dehydrogenase |
| P60624_RL24_ECOLI | 0 | 0.7 | 15 | 2 | 50S ribosomal protein L24. |
| P0A7M6_RL29_ECOLI | 0 | 0.7 | 18 | 2 | 50S ribosomal protein L29. |
| P0A7U3_RS19_ECOLI | 0 | 0.7 | 7 | 2 | 30S ribosomal protein S19. |
| P0ADA5_YAJG_ECOLI | 0 | 0.7 | 13 | 2 | Uncharacterized lipoprotein yajG |
| P0A986_CSPI_ECOLI | 0 | 0.6 | 15 | 2 | Cold shock-like protein cspI |
| P02931_OMPF_ECOLI | 0 | 0.6 | 21 | 2 | Outer membrane protein F |
| P28304_QOR_ECOLI | 0 | 0.6 | 12 | 2 | Quinone oxidoreductase |
| P0C018_RL18_ECOLI | 0 | 0.6 | 17 | 2 | 50S ribosomal protein L18. |
| P00448_SODM_ECOLI | 0 | 0.6 | 37 | 2 | Superoxide dismutase [Mn] |
| P02930_TOLC_ECOLI | 0 | 0.6 | 20 | 2 | Outer membrane protein tolC |
| P0A877_TRPA_ECOLI | 0 | 0.6 | 77 | 2 | Tryptophan synthase alpha chain |
| P0ADS6_YGGE_ECOLI | 0 | 0.6 | 12 | 2 | Uncharacterized protein yggE. |
| P0ADA3_NLPD_ECOLI | 0 | 0.6 | 3 | 1 | Lipoprotein nlpD |
| P0AB55_YCII_ECOLI | 0 | 0.6 | 5 | 1 | Protein yciI. |
| P0A972_CSPE_ECOLI | 0 | 0.5 | 10 | 2 | Cold shock-like protein cspE |

| | | | | | |
|---|---|---|---|---|---|
| P04128_FIMA1_ECOLI | 0 | 0.5 | 9 | 2 | Type-1 fimbrial protein, A chain |
| P60723_RL4_ECOLI | 0 | 0.5 | 37 | 2 | 50S ribosomal protein L4. |
| P0A7T7_RS18_ECOLI | 0 | 0.5 | 15 | 2 | 30S ribosomal protein S18. |
| P68679_RS21_ECOLI | 0 | 0.5 | 10 | 2 | 30S ribosomal protein S21. |
| P76177_YDGH_ECOLI | 0 | 0.5 | 130 | 2 | Protein ydgH |
| P0A6Y1_IHFB_ECOLI | 0 | 0.4 | 6 | 2 | Integration host factor subunit beta |
| P0C0L2_OSMC_ECOLI | 0 | 0.4 | 47 | 2 | Peroxiredoxin osmC |
| P0A7J7_RL11_ECOLI | 0 | 0.4 | 40 | 2 | 50S ribosomal protein L11. |
| P0ADY7_RL16_ECOLI | 0 | 0.4 | 26 | 2 | 50S ribosomal protein L16. |
| P0A7K2_RL7_ECOLI | 0 | 0.4 | 16 | 2 | 50S ribosomal protein L7/L12 |
| P0A805_RRF_ECOLI | 0 | 0.4 | 48 | 2 | Ribosome recycling factor |
| P0A7T3_RS16_ECOLI | 0 | 0.4 | 31 | 2 | 30S ribosomal protein S16. |
| P0A7W7_RS8_ECOLI | 0 | 0.4 | 8 | 2 | 30S ribosomal protein S8. |
| P77717_YBAY_ECOLI | 0 | 0.4 | 22 | 2 | Uncharacterized lipoprotein ybaY |
| P0A7N1_RL31B_ECOLI | 0 | 0.4 | 7 | 1 | 50S ribosomal protein L31 type B. |
| P0ABJ1_CYOA_ECOLI | 0 | 0.3 | 8 | 2 | Ubiquinol oxidase subunit 2 |
| P0A6N4_EFP_ECOLI | 0 | 0.3 | 15 | 2 | Elongation factor P |
| P61320_LOLB_ECOLI | 0 | 0.3 | 8 | 2 | Outer-membrane lipoprotein lolB |
| P61175_RL22_ECOLI | 0 | 0.3 | 20 | 2 | 50S ribosomal protein L22. |
| P0A7W1_RS5_ECOLI | 0 | 0.3 | 36 | 2 | 30S ribosomal protein S5. |
| P22256_GABT_ECOLI | 0 | 0.3 | 4 | 1 | 4-aminobutyrate aminotransferase |
| P0A903_NLPB_ECOLI | 0 | 0.3 | 9 | 1 | Lipoprotein 34 |
| P33570_TKT2_ECOLI | 0 | 0.3 | 14 | 1 | Transketolase 2 |
| P77625_YFBT_ECOLI | 0 | 0.3 | 4 | 1 | Phosphatase yfbT |
| P76569_YFGD_ECOLI | 0 | 0.3 | 5 | 1 | Uncharacterized protein yfgD. |
| P77774_YFGL_ECOLI | 0 | 0.3 | 4 | 1 | Lipoprotein yfgL |
| P0A9W6_YRBA_ECOLI | 0 | 0.3 | 3 | 1 | Uncharacterized protein yrbA. |
| P00350_6PGD_ECOLI | 0 | 0.2 | 56 | 2 | 6-phosphogluconate dehydrogenase, decarboxylating |
| P21179_CATE_ECOLI | 0 | 0.2 | 23 | 2 | Catalase HPII |
| P0A9Y6_CSPC_ECOLI | 0 | 0.2 | 47 | 2 | Cold shock-like protein cspC |
| P0A9B2_G3P1_ECOLI | 0 | 0.2 | 192 | 2 | Glyceraldehyde-3-phosphate dehydrogenase A |
| P69776_LPP_ECOLI | 0 | 0.2 | 14 | 2 | Major outer membrane lipoprotein |
| P0A908_MIPA_ECOLI | 0 | 0.2 | 25 | 2 | MltA-interacting protein |

| | | | | | |
|---|---|---|---|---|---|
| P0A7K6_RL19_ECOLI | 0 | 0.2 | 20 | 2 | 50S ribosomal protein L19. |
| P0A7R1_RL9_ECOLI | 0 | 0.2 | 42 | 2 | 50S ribosomal protein L9. |
| P0A8M6_YEEX_ECOLI | 0 | 0.2 | 8 | 2 | UPF0265 protein yeeX. |
| P24186_FOLD_ECOLI | 0 | 0.2 | 3 | 1 | Bifunctional protein folD |
| P0AAI3_FTSH_ECOLI | 0 | 0.2 | 7 | 1 | Cell division protease ftsH |
| P10100_RLPA_ECOLI | 0 | 0.2 | 4 | 1 | Rare lipoprotein A |
| P46130_YBHC_ECOLI | 0 | 0.2 | 8 | 1 | Acyl-CoA thioester hydrolase ybgC |
| P31063_YEDD_ECOLI | 0 | 0.2 | 4 | 1 | Uncharacterized lipoprotein yedD |
| P06129_BTUB_ECOLI | 0 | 0.1 | 8 | 2 | Vitamin B12 transporter btuB |
| P0A8P3_FETP_ECOLI | 0 | 0.1 | 19 | 2 | Probable Fe(2+)-trafficking protein. |
| P69829_PTSN_ECOLI | 0 | 0.1 | 10 | 2 | Nitrogen regulatory protein |
| P61714_RISB_ECOLI | 0 | 0.1 | 47 | 2 | 6,7-dimethyl-8-ribityllumazine synthase |
| P02413_RL15_ECOLI | 0 | 0.1 | 36 | 2 | 50S ribosomal protein L15. |
| P0A867_TALA_ECOLI | 0 | 0.1 | 17 | 2 | Transaldolase A |
| P0A8E7_YAJQ_ECOLI | 0 | 0.1 | 48 | 2 | UPF0234 protein yajQ. |
| P0AAV6_YBGS_ECOLI | 0 | 0.1 | 6 | 2 | Uncharacterized protein ybgS |
| P39173_YEAD_ECOLI | 0 | 0.1 | 23 | 2 | UPF0010 protein yeaD |
| P0AC02_YFIO_ECOLI | 0 | 0.1 | 9 | 2 | UPF0169 lipoprotein yfiO |
| P52697_6PGL_ECOLI | 0 | 0.1 | 9 | 1 | 6-phosphogluconolactonase |
| P0AFF6_NUSA_ECOLI | 0 | 0.1 | 4 | 1 | Transcription elongation protein nusA |
| P21513_RNE_ECOLI | 0 | 0.1 | 8 | 1 | Ribonuclease E |
| P0A7S9_RS13_ECOLI | 0 | 0.1 | 9 | 1 | 30S ribosomal protein S13. |
| P68206_YJBJ_ECOLI | 0 | 0.1 | 10 | 1 | UPF0337 protein yjbJ. |
| P24182_ACCC_ECOLI | 0 | 0 | 11 | 2 | Biotin carboxylase |
| P33235_FLGK_ECOLI | 0 | 0 | 23 | 2 | Flagellar hook-associated protein 1 |
| P04949_FLIC_ECOLI | 0 | 0 | 1343 | 2 | Flagellin. |
| P0ACF8_HNS_ECOLI | 0 | 0 | 37 | 2 | DNA-binding protein H-NS |
| P00935_METB_ECOLI | 0 | 0 | 15 | 2 | Cystathionine gamma-synthase |
| P38489_NFNB_ECOLI | 0 | 0 | 26 | 2 | Oxygen-insensitive NAD(P)H nitroreductase |
| P0A7E5_PYRG_ECOLI | 0 | 0 | 13 | 2 | CTP synthase |
| P0AG44_RL17_ECOLI | 0 | 0 | 18 | 2 | 50S ribosomal protein L17. |
| P0A7L0_RL1_ECOLI | 0 | 0 | 39 | 2 | 50S ribosomal protein L1. |
| P0A7M2_RL28_ECOLI | 0 | 0 | 11 | 2 | 50S ribosomal protein L28. |

| | | | | | |
|---|---|---|---|---|---|
| P60422_RL2_ECOLI | 0 | 0 | 44 | 2 | 50S ribosomal protein L2. |
| P60438_RL3_ECOLI | 0 | 0 | 31 | 2 | 50S ribosomal protein L3. |
| P76344_YODA_ECOLI | 0 | 0 | 306 | 2 | Metal-binding protein yodA |
| P0ADV7_YRBC_ECOLI | 0 | 0 | 37 | 2 | Protein yrbC |
| P0AE91_CREA_ECOLI | 0 | 0 | 7 | 1 | Protein creA. |
| P0ACJ0_LRP_ECOLI | 0 | 0 | 5 | 1 | Leucine-responsive regulatory protein. |
| Q47702_YFEK_ECOLI | 0 | 0 | 3 | 1 | Uncharacterized protein yfeK |
| P0ADE8_YGFZ_ECOLI | 0 | 0 | 26 | 1 | tRNA-modifying protein ygfZ. |
| P04036_DAPB_ECOLI | 0 | -0.1 | 10 | 2 | Dihydrodipicolinate reductase |
| P0AEK2_FABG_ECOLI | 0 | -0.1 | 24 | 2 | 3-oxoacyl-[acyl-carrier-protein] reductase |
| P0A707_IF3_ECOLI | 0 | -0.1 | 8 | 2 | Translation initiation factor IF-3. |
| P61889_MDH_ECOLI | 0 | -0.1 | 135 | 2 | Malate dehydrogenase |
| P0AG30_RHO_ECOLI | 0 | -0.1 | 24 | 2 | Transcription termination factor rho |
| P0A7J3_RL10_ECOLI | 0 | -0.1 | 16 | 2 | 50S ribosomal protein L10 |
| P0AA10_RL13_ECOLI | 0 | -0.1 | 12 | 2 | 50S ribosomal protein L13. |
| P0ADY3_RL14_ECOLI | 0 | -0.1 | 8 | 2 | 50S ribosomal protein L14. |
| P0A7R5_RS10_ECOLI | 0 | -0.1 | 17 | 2 | 30S ribosomal protein S10. |
| P02358_RS6_ECOLI | 0 | -0.1 | 21 | 2 | 30S ribosomal protein S6 |
| P02359_RS7_ECOLI | 0 | -0.1 | 53 | 2 | 30S ribosomal protein S7. |
| P0ACG1_STPA_ECOLI | 0 | -0.1 | 26 | 2 | DNA-binding protein stpA |
| P0AA25_THIO_ECOLI | 0 | -0.1 | 20 | 2 | Thioredoxin-1 |
| P39177_USPG_ECOLI | 0 | -0.1 | 32 | 2 | Universal stress protein G. |
| P0A8J4_YBED_ECOLI | 0 | -0.1 | 8 | 2 | UPF0250 protein ybeD. |
| P66948_YFGC_ECOLI | 0 | -0.1 | 15 | 2 | TPR repeat-containing protein yfgC |
| P04825_AMPN_ECOLI | 0 | -0.1 | 3 | 1 | Aminopeptidase N |
| P0A6E1_AROL_ECOLI | 0 | -0.1 | 3 | 1 | Shikimate kinase 2 |
| P36995_CSPB_ECOLI | 0 | -0.1 | 3 | 1 | Cold shock-like protein cspB |
| Q46857_DKGA_ECOLI | 0 | -0.1 | 7 | 1 | 2,5-diketo-D-gluconic acid reductase A |
| P0AC62_GLRX3_ECOLI | 0 | -0.1 | 4 | 1 | Glutaredoxin-3 |
| P76558_MAO2_ECOLI | 0 | -0.1 | 9 | 1 | NADP-dependent malic enzyme |
| P08244_PYRF_ECOLI | 0 | -0.1 | 4 | 1 | Orotidine 5'-phosphate decarboxylase |
| P37744_RMLA1_ECOLI | 0 | -0.1 | 9 | 1 | Glucose-1-phosphate thymidylyltransferase 1 |
| P76172_YNFD_ECOLI | 0 | -0.1 | 7 | 1 | Uncharacterized protein ynfD |

| | | | | | |
|---|---|---|---|---|---|
| Q46856_YQHD_ECOLI | 0 | -0.1 | 3 | 1 | Alcohol dehydrogenase yqhD |
| P0ABB0_ATPA_ECOLI | 0 | -0.2 | 29 | 2 | ATP synthase subunit alpha |
| P29744_FLGL_ECOLI | 0 | -0.2 | 54 | 2 | Flagellar hook-associated protein 3 |
| P0A6W5_GREA_ECOLI | 0 | -0.2 | 11 | 2 | Transcription elongation factor greA |
| P0ADG7_IMDH_ECOLI | 0 | -0.2 | 61 | 2 | Inosine-5'-monophosphate dehydrogenase |
| P69441_KAD_ECOLI | 0 | -0.2 | 91 | 2 | Adenylate kinase |
| P26616_MAO1_ECOLI | 0 | -0.2 | 9 | 2 | NAD-dependent malic enzyme |
| P28635_METQ_ECOLI | 0 | -0.2 | 25 | 2 | D-methionine-binding lipoprotein metQ |
| P0A912_PAL_ECOLI | 0 | -0.2 | 31 | 2 | Peptidoglycan-associated lipoprotein |
| P0A799_PGK_ECOLI | 0 | -0.2 | 187 | 2 | Phosphoglycerate kinase |
| P0A7Z0_RPIA_ECOLI | 0 | -0.2 | 9 | 2 | Ribose-5-phosphate isomerase A |
| P0A7Z4_RPOA_ECOLI | 0 | -0.2 | 25 | 2 | DNA-directed RNA polymerase subunit alpha |
| P08312_SYFA_ECOLI | 0 | -0.2 | 3 | 1 | Phenylalanyl-tRNA synthetase alpha chain |
| P0ACY1_YDJA_ECOLI | 0 | -0.2 | 6 | 1 | Protein ydjA. |
| P0ABD8_BCCP_ECOLI | 0 | -0.3 | 7 | 2 | Biotin carboxyl carrier protein of acetyl-CoA carboxylase |
| P0A9P0_DLDH_ECOLI | 0 | -0.3 | 25 | 2 | Dihydrolipoyl dehydrogenase |
| P0AAI5_FABF_ECOLI | 0 | -0.3 | 11 | 2 | 3-oxoacyl-[acyl-carrier-protein] synthase 2 |
| P0AEK4_FABI_ECOLI | 0 | -0.3 | 40 | 2 | Enoyl-[acyl-carrier-protein] reductase [NADH] |
| P24216_FLID_ECOLI | 0 | -0.3 | 37 | 2 | Flagellar hook-associated protein 2 |
| P06715_GSHR_ECOLI | 0 | -0.3 | 9 | 2 | Glutathione reductase |
| P08200_IDH_ECOLI | 0 | -0.3 | 152 | 2 | Isocitrate dehydrogenase [NADP] |
| P25665_METE_ECOLI | 0 | -0.3 | 90 | 2 | 5-methyltetrahydropteroyltriglutamate--homocysteine |
| P0AFG6_ODO2_ECOLI | 0 | -0.3 | 26 | 2 | Dihydrolipoyllysine-residue succinyltransferase component |
| P31663_PANC_ECOLI | 0 | -0.3 | 14 | 2 | Pantothenate synthetase |
| P09373_PFLB_ECOLI | 0 | -0.3 | 297 | 2 | Formate acetyltransferase 1 |
| P23869_PPIB_ECOLI | 0 | -0.3 | 38 | 2 | Peptidyl-prolyl cis-trans isomerase B |
| P0AG18_PUR6_ECOLI | 0 | -0.3 | 13 | 2 | Phosphoribosylaminoimidazole carboxylase catalytic subunit |
| P0A7J0_RIBB_ECOLI | 0 | -0.3 | 8 | 2 | 3,4-dihydroxy-2-butanone 4-phosphate synthase |
| P0A7R9_RS11_ECOLI | 0 | -0.3 | 18 | 2 | 30S ribosomal protein S11. |
| P0A7X3_RS9_ECOLI | 0 | -0.3 | 9 | 2 | 30S ribosomal protein S9. |
| P21889_SYD_ECOLI | 0 | -0.3 | 22 | 2 | Aspartyl-tRNA synthetase |
| P0AGE6_YIEF_ECOLI | 0 | -0.3 | 13 | 2 | Uncharacterized protein yieF. |
| P0AE67_CHEY_ECOLI | 0 | -0.3 | 4 | 1 | Chemotaxis protein cheY. |

| | | | | | |
|---|---|---|---|---|---|
| P0ABP8_DEOD_ECOLI | 0 | -0.3 | 7 | 1 | Purine nucleoside phosphorylase deoD-type |
| P0AG48_RL21_ECOLI | 0 | -0.3 | 8 | 1 | 50S ribosomal protein L21. |
| P0ADS9_YGGN_ECOLI | 0 | -0.3 | 3 | 1 | Uncharacterized protein yggN. |
| P00509_AAT_ECOLI | 0 | -0.4 | 67 | 2 | Aspartate aminotransferase |
| P0A9G6_ACEA_ECOLI | 0 | -0.4 | 89 | 2 | Isocitrate lyase |
| P0A9Q7_ADHE_ECOLI | 0 | -0.4 | 76 | 2 | Aldehyde-alcohol dehydrogenase |
| P0AE08_AHPC_ECOLI | 0 | -0.4 | 120 | 2 | Alkyl hydroperoxide reductase subunit C |
| P0AE52_BCP_ECOLI | 0 | -0.4 | 13 | 2 | Putative peroxiredoxin bcp |
| P0A6F5_CH60_ECOLI | 0 | -0.4 | 124 | 2 | 60 kDa chaperonin |
| P0AAI9_FABD_ECOLI | 0 | -0.4 | 53 | 2 | Malonyl CoA-acyl carrier protein transacylase |
| P0A9C5_GLNA_ECOLI | 0 | -0.4 | 54 | 2 | Glutamine synthetase |
| P0A825_GLYA_ECOLI | 0 | -0.4 | 139 | 2 | Serine hydroxymethyltransferase |
| P05793_ILVC_ECOLI | 0 | -0.4 | 123 | 2 | Ketol-acid reductoisomerase |
| P0A7A9_IPYR_ECOLI | 0 | -0.4 | 12 | 2 | Inorganic pyrophosphatase |
| P0A796_K6PF1_ECOLI | 0 | -0.4 | 39 | 2 | 6-phosphofructokinase isozyme 1 |
| P0A715_KDSA_ECOLI | 0 | -0.4 | 27 | 2 | 2-dehydro-3-deoxyphosphooctonate aldolase |
| P09151_LEU1_ECOLI | 0 | -0.4 | 14 | 2 | 2-isopropylmalate synthase |
| P69797_PTNAB_ECOLI | 0 | -0.4 | 21 | 2 | PTS system mannose-specific EIIAB component |
| P0A7G6_RECA_ECOLI | 0 | -0.4 | 20 | 2 | Protein recA |
| P0AG67_RS1_ECOLI | 0 | -0.4 | 23 | 2 | 30S ribosomal protein S1. |
| P0A7V8_RS4_ECOLI | 0 | -0.4 | 20 | 2 | 30S ribosomal protein S4. |
| P23721_SERC_ECOLI | 0 | -0.4 | 61 | 2 | Phosphoserine aminotransferase |
| P27302_TKT1_ECOLI | 0 | -0.4 | 30 | 2 | Transketolase 1 |
| P0A879_TRPB_ECOLI | 0 | -0.4 | 18 | 2 | Tryptophan synthase beta chain |
| P0AF93_YJGF_ECOLI | 0 | -0.4 | 35 | 2 | UPF0076 protein yjgF. |
| P39371_YJHT_ECOLI | 0 | -0.4 | 13 | 2 | Kelch domain-containing protein yjhT |
| P0ABA6_ATPG_ECOLI | 0 | -0.4 | 4 | 1 | ATP synthase gamma chain |
| P0AFM4_PSIF_ECOLI | 0 | -0.4 | 7 | 1 | Phosphate starvation-inducible protein psiF |
| P0AG86_SECB_ECOLI | 0 | -0.4 | 3 | 1 | Protein-export protein secB. |
| P0ACA3_SSPA_ECOLI | 0 | -0.4 | 6 | 1 | Stringent starvation protein A. |
| P60906_SYH_ECOLI | 0 | -0.4 | 8 | 1 | Histidyl-tRNA synthetase |
| P00962_SYQ_ECOLI | 0 | -0.4 | 13 | 1 | Glutaminyl-tRNA synthetase |
| P76227_YNJH_ECOLI | 0 | -0.4 | 10 | 1 | Uncharacterized protein ynjH |

| | | | | | |
|---|---|---|---|---|---|
| P35340_AHPF_ECOLI | 0 | -0.5 | 10 | 2 | Alkyl hydroperoxide reductase subunit F |
| P0A991_ALF1_ECOLI | 0 | -0.5 | 13 | 2 | Fructose-bisphosphate aldolase class 1 |
| P0A955_ALKH_ECOLI | 0 | -0.5 | 27 | 2 | KHG/KDPG aldolase |
| P0A6D3_AROA_ECOLI | 0 | -0.5 | 7 | 2 | 3-phosphoshikimate 1-carboxyvinyltransferase |
| P0ABH7_CISY_ECOLI | 0 | -0.5 | 13 | 2 | Citrate synthase |
| P0A6Q3_FABA_ECOLI | 0 | -0.5 | 20 | 2 | 3-hydroxydecanoyl-[acyl-carrier-protein] dehydratase |
| P75937_FLGE_ECOLI | 0 | -0.5 | 106 | 2 | Flagellar hook protein flgE. |
| P04425_GSHB_ECOLI | 0 | -0.5 | 11 | 2 | Glutathione synthetase |
| P0AD61_KPYK1_ECOLI | 0 | -0.5 | 77 | 2 | Pyruvate kinase I |
| P00946_MANA_ECOLI | 0 | -0.5 | 23 | 2 | Mannose-6-phosphate isomerase |
| P09424_MTLD_ECOLI | 0 | -0.5 | 30 | 2 | Mannitol-1-phosphate 5-dehydrogenase |
| P0AF24_NAGD_ECOLI | 0 | -0.5 | 16 | 2 | Protein nagD. |
| P33221_PURT_ECOLI | 0 | -0.5 | 14 | 2 | Phosphoribosylglycinamide formyltransferase 2 |
| P0A7V0_RS2_ECOLI | 0 | -0.5 | 19 | 2 | 30S ribosomal protein S2. |
| P0A9T0_SERA_ECOLI | 0 | -0.5 | 56 | 2 | D-3-phosphoglycerate dehydrogenase |
| P0AGE9_SUCD_ECOLI | 0 | -0.5 | 12 | 2 | Succinyl-CoA ligase [ADP-forming] subunit alpha |
| P04805_SYE_ECOLI | 0 | -0.5 | 12 | 2 | Glutamyl-tRNA synthetase |
| P0A8M0_SYN_ECOLI | 0 | -0.5 | 14 | 2 | Asparaginyl-tRNA synthetase |
| P07118_SYV_ECOLI | 0 | -0.5 | 23 | 2 | Valyl-tRNA synthetase |
| P0AGL2_TDCF_ECOLI | 0 | -0.5 | 7 | 2 | Protein tdcF. |
| P0A858_TPIS_ECOLI | 0 | -0.5 | 34 | 2 | Triosephosphate isomerase |
| P0AAB8_USPD_ECOLI | 0 | -0.5 | 19 | 2 | Universal stress protein D. |
| P64506_YEBY_ECOLI | 0 | -0.5 | 7 | 2 | Uncharacterized protein yebY |
| P26646_YHDH_ECOLI | 0 | -0.5 | 13 | 2 | Protein yhdH. |
| P0A6D7_AROK_ECOLI | 0 | -0.5 | 4 | 1 | Shikimate kinase 1 |
| P76015_DHAK_ECOLI | 0 | -0.5 | 11 | 1 | PTS-dependent dihydroxyacetone kinase |
| P06959_ODP2_ECOLI | 0 | -0.5 | 12 | 1 | Dihydrolipoyllysine-residue acetyltransferase component |
| P34209_YDCF_ECOLI | 0 | -0.5 | 6 | 1 | Protein ydcF. |
| P32157_YIIM_ECOLI | 0 | -0.5 | 3 | 1 | Protein yiiM. |
| P0A9W9_YRDA_ECOLI | 0 | -0.5 | 8 | 1 | Protein yrdA. |
| P0AB71_ALF_ECOLI | 0 | -0.6 | 64 | 2 | Fructose-bisphosphate aldolase class 2 |
| P0A6F9_CH10_ECOLI | 0 | -0.6 | 27 | 2 | 10 kDa chaperonin |
| P0A6P1_EFTS_ECOLI | 0 | -0.6 | 84 | 2 | Elongation factor Ts |

| | | | | | |
|---|---|---|---|---|---|
| P0A953_FABB_ECOLI | 0 | -0.6 | 42 | 2 | 3-oxoacyl-[acyl-carrier-protein] synthase 1 |
| P13035_GLPD_ECOLI | 0 | -0.6 | 37 | 2 | Aerobic glycerol-3-phosphate dehydrogenase |
| P0AC59_GLRX2_ECOLI | 0 | -0.6 | 31 | 2 | Glutaredoxin-2 |
| P0AET2_HDEB_ECOLI | 0 | -0.6 | 52 | 2 | Protein hdeB |
| P60757_HIS1_ECOLI | 0 | -0.6 | 8 | 2 | ATP phosphoribosyltransferase |
| P06988_HISX_ECOLI | 0 | -0.6 | 19 | 2 | Histidinol dehydrogenase |
| P0A705_IF2_ECOLI | 0 | -0.6 | 20 | 2 | Translation initiation factor IF-2. |
| P08997_MASY_ECOLI | 0 | -0.6 | 77 | 2 | Malate synthase A |
| P0AFG0_NUSG_ECOLI | 0 | -0.6 | 13 | 2 | Transcription antitermination protein nusG. |
| P0AFG8_ODP1_ECOLI | 0 | -0.6 | 55 | 2 | Pyruvate dehydrogenase E1 component |
| P15288_PEPD_ECOLI | 0 | -0.6 | 15 | 2 | Aminoacyl-histidine dipeptidase |
| P15254_PUR4_ECOLI | 0 | -0.6 | 63 | 2 | Phosphoribosylformylglycinamidine synthase |
| P15639_PUR9_ECOLI | 0 | -0.6 | 12 | 2 | Bifunctional purine biosynthesis protein purH |
| P0A7F3_PYRI_ECOLI | 0 | -0.6 | 28 | 2 | Aspartate carbamoyltransferase regulatory chain. |
| P62399_RL5_ECOLI | 0 | -0.6 | 40 | 2 | 50S ribosomal protein L5. |
| P0A836_SUCC_ECOLI | 0 | -0.6 | 55 | 2 | Succinyl-CoA synthetase beta chain |
| P0A8L1_SYS_ECOLI | 0 | -0.6 | 31 | 2 | Seryl-tRNA synthetase |
| P0A870_TALB_ECOLI | 0 | -0.6 | 60 | 2 | Transaldolase B |
| P0A8F0_UPP_ECOLI | 0 | -0.6 | 14 | 2 | Uracil phosphoribosyltransferase |
| P33362_YEHZ_ECOLI | 0 | -0.6 | 21 | 2 | Uncharacterized protein yehZ |
| P22106_ASNB_ECOLI | 0 | -0.6 | 4 | 1 | Asparagine synthetase B [glutamine-hydrolyzing] |
| P04079_GUAA_ECOLI | 0 | -0.6 | 3 | 1 | GMP synthase [glutamine-hydrolyzing] |
| P0A6A3_ACKA_ECOLI | 0 | -0.7 | 14 | 2 | Acetate kinase |
| P0A6A8_ACP_ECOLI | 0 | -0.7 | 13 | 2 | Acyl carrier protein |
| P0A6G7_CLPP_ECOLI | 0 | -0.7 | 12 | 2 | ATP-dependent Clp protease proteolytic subunit |
| P0A6K6_DEOB_ECOLI | 0 | -0.7 | 25 | 2 | Phosphopentomutase |
| P0A6L0_DEOC_ECOLI | 0 | -0.7 | 29 | 2 | Deoxyribose-phosphate aldolase |
| P0A9Q9_DHAS_ECOLI | 0 | -0.7 | 110 | 2 | Aspartate-semialdehyde dehydrogenase |
| P00370_DHE4_ECOLI | 0 | -0.7 | 48 | 2 | NADP-specific glutamate dehydrogenase |
| P0A6T1_G6PI_ECOLI | 0 | -0.7 | 43 | 2 | Glucose-6-phosphate isomerase |
| P0A817_METK_ECOLI | 0 | -0.7 | 25 | 2 | S-adenosylmethionine synthetase |
| P0A786_PYRB_ECOLI | 0 | -0.7 | 56 | 2 | Aspartate carbamoyltransferase catalytic chain |
| P69407_RCSB_ECOLI | 0 | -0.7 | 17 | 2 | Capsular synthesis regulator component B. |

| | | | | | |
|---|---|---|---|---|---|
| P0A7I4_RF3_ECOLI | 0 | -0.7 | 7 | 2 | Peptide chain release factor 3 |
| P0A9K9_SLYD_ECOLI | 0 | -0.7 | 23 | 2 | FKBP-type peptidyl-prolyl cis-trans isomerase slyD |
| P00954_SYW_ECOLI | 0 | -0.7 | 20 | 2 | Tryptophanyl-tRNA synthetase |
| P0A850_TIG_ECOLI | 0 | -0.7 | 88 | 2 | Trigger factor |
| P75694_YAHO_ECOLI | 0 | -0.7 | 52 | 2 | UPF0379 protein yahO |
| P0ACX3_YDHR_ECOLI | 0 | -0.7 | 10 | 2 | Protein ydhR |
| P0A8W8_YFBU_ECOLI | 0 | -0.7 | 10 | 2 | UPF0304 protein yfbU. |
| P76223_YNJB_ECOLI | 0 | -0.7 | 21 | 2 | Protein ynjB. |
| P0A6T9_GCSH_ECOLI | 0 | -0.7 | 4 | 1 | Glycine cleavage system H protein. |
| P75823_LTAE_ECOLI | 0 | -0.7 | 4 | 1 | Low specificity L-threonine aldolase |
| P0A7L3_RL20_ECOLI | 0 | -0.7 | 5 | 1 | 50S ribosomal protein L20. |
| P00864_CAPP_ECOLI | 0 | -0.8 | 45 | 2 | Phosphoenolpyruvate carboxylase |
| P0A9D8_DAPD_ECOLI | 0 | -0.8 | 69 | 2 | 2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-succinyltrans |
| P24171_DCP_ECOLI | 0 | -0.8 | 33 | 2 | Peptidyl-dipeptidase dcp |
| P0A6Y8_DNAK_ECOLI | 0 | -0.8 | 78 | 2 | Chaperone protein dnaK |
| P0ABT2_DPS_ECOLI | 0 | -0.8 | 72 | 2 | DNA protection during starvation protein |
| P0A6P9_ENO_ECOLI | 0 | -0.8 | 305 | 2 | Enolase |
| P25526_GABD_ECOLI | 0 | -0.8 | 6 | 2 | Succinate-semialdehyde dehydrogenase [NADP+] |
| P0AAB6_GALF_ECOLI | 0 | -0.8 | 13 | 2 | UTP--glucose-1-phosphate uridylyltransferase |
| P09372_GRPE_ECOLI | 0 | -0.8 | 16 | 2 | Protein grpE |
| P0ACE7_HINT_ECOLI | 0 | -0.8 | 16 | 2 | HIT-like protein hinT. |
| P69783_PTGA_ECOLI | 0 | -0.8 | 123 | 2 | Glucose-specific phosphotransferase enzyme IIA component |
| P07012_RF2_ECOLI | 0 | -0.8 | 16 | 2 | Peptide chain release factor 2 |
| P0AED0_USPA_ECOLI | 0 | -0.8 | 27 | 2 | Universal stress protein A. |
| P0ACA7_YLIJ_ECOLI | 0 | -0.8 | 15 | 2 | Uncharacterized GST-like protein yliJ. |
| P18843_NADE_ECOLI | 0 | -0.8 | 12 | 1 | NH(3)-dependent NAD(+) synthetase |
| P05055_PNP_ECOLI | 0 | -0.8 | 4 | 1 | Polyribonucleotide nucleotidyltransferase |
| P0AFU8_RISA_ECOLI | 0 | -0.8 | 3 | 1 | Riboflavin synthase alpha chain |
| P0AGJ9_SYY_ECOLI | 0 | -0.8 | 7 | 1 | Tyrosyl-tRNA synthetase |
| P0ABB4_ATPB_ECOLI | 0 | -0.9 | 55 | 2 | ATP synthase subunit beta |
| P63284_CLPB_ECOLI | 0 | -0.9 | 28 | 2 | Chaperone protein clpB |
| P0ABK5_CYSK_ECOLI | 0 | -0.9 | 216 | 2 | Cysteine synthase A |
| P0A6L2_DAPA_ECOLI | 0 | -0.9 | 15 | 2 | Dihydrodipicolinate synthase |

| | | | | | |
|---|---|---|---|---|---|
| P21599_KPYK2_ECOLI | 0 | -0.9 | 59 | 2 | Pyruvate kinase II |
| P30125_LEU3_ECOLI | 0 | -0.9 | 23 | 2 | 3-isopropylmalate dehydrogenase |
| P45578_LUXS_ECOLI | 0 | -0.9 | 25 | 2 | S-ribosylhomocysteine lyase |
| P0A744_MSRA_ECOLI | 0 | -0.9 | 6 | 2 | Peptide methionine sulfoxide reductase msrA |
| P37095_PEPB_ECOLI | 0 | -0.9 | 11 | 2 | Peptidase B |
| P21165_PEPQ_ECOLI | 0 | -0.9 | 21 | 2 | Xaa-Pro dipeptidase |
| P31133_POTF_ECOLI | 0 | -0.9 | 63 | 2 | Putrescine-binding periplasmic protein |
| P08839_PT1_ECOLI | 0 | -0.9 | 102 | 2 | Phosphoenolpyruvate-protein phosphotransferase |
| P0A7D4_PURA_ECOLI | 0 | -0.9 | 53 | 2 | Adenylosuccinate synthetase |
| P0A7V3_RS3_ECOLI | 0 | -0.9 | 29 | 2 | 30S ribosomal protein S3. |
| P17169_GLMS_ECOLI | 0 | -0.9 | 4 | 1 | Glucosamine--fructose-6-phosphate aminotransferase [isomeriz |
| P06960_OTC2_ECOLI | 0 | -0.9 | 3 | 1 | Ornithine carbamoyltransferase chain F |
| P0AB89_PUR8_ECOLI | 0 | -0.9 | 7 | 1 | Adenylosuccinate lyase |
| P0A8N3_SYK1_ECOLI | 0 | -0.9 | 5 | 1 | Lysyl-tRNA synthetase |
| P0A8M3_SYT_ECOLI | 0 | -0.9 | 3 | 1 | Threonyl-tRNA synthetase |
| P0AB91_AROG_ECOLI | 0 | -1 | 33 | 2 | Phospho-2-dehydro-3-deoxyheptonate aldolase, Phe-sensitive |
| P0A6V8_GLK_ECOLI | 0 | -1 | 10 | 2 | Glucokinase |
| P09832_GLTD_ECOLI | 0 | -1 | 25 | 2 | Glutamate synthase [NADPH] small chain |
| P0AB80_ILVE_ECOLI | 0 | -1 | 10 | 2 | Branched-chain-amino-acid aminotransferase |
| P0AF12_MTNN_ECOLI | 0 | -1 | 14 | 2 | MTA/SAH nucleosidase |
| P0A749_MURA_ECOLI | 0 | -1 | 24 | 2 | UDP-N-acetylglucosamine 1-carboxyvinyltransferase |
| P04391_OTC1_ECOLI | 0 | -1 | 26 | 2 | Ornithine carbamoyltransferase chain I |
| P29745_PEPT_ECOLI | 0 | -1 | 11 | 2 | Peptidase T |
| P09158_SPEE_ECOLI | 0 | -1 | 10 | 2 | Spermidine synthase |
| P32132_TYPA_ECOLI | 0 | -1 | 27 | 2 | GTP-binding protein typA/BipA |
| P77735_YAJO_ECOLI | 0 | -1 | 6 | 2 | Uncharacterized oxidoreductase yajO |
| P0A6Z3_HTPG_ECOLI | 0 | -1 | 9 | 1 | Chaperone protein htpG |
| P0A6B7_ISCS_ECOLI | 0 | -1 | 5 | 1 | Cysteine desulfurase |
| P39180_AG43_ECOLI | 0 | -1.1 | 691 | 2 | Antigen 43 |
| P08660_AK3_ECOLI | 0 | -1.1 | 40 | 2 | Lysine-sensitive aspartokinase 3 |
| P25553_ALDA_ECOLI | 0 | -1.1 | 9 | 2 | Aldehyde dehydrogenase A |
| P0A9A6_FTSZ_ECOLI | 0 | -1.1 | 9 | 2 | Cell division protein ftsZ. |
| P31658_HCHA_ECOLI | 0 | -1.1 | 20 | 2 | Chaperone protein hchA |

| | | | | | |
|---|---|---|---|---|---|
| P0A717_KPRS_ECOLI | 0 | -1.1 | 13 | 2 | Ribose-phosphate pyrophosphokinase |
| P16659_SYP_ECOLI | 0 | -1.1 | 17 | 2 | Prolyl-tRNA synthetase |
| P12758_UDP_ECOLI | 0 | -1.1 | 36 | 2 | Uridine phosphorylase |
| P0A8G6_WRBA_ECOLI | 0 | -1.1 | 82 | 2 | Flavoprotein wrbA |
| P0A6F1_CARA_ECOLI | 0 | -1.1 | 7 | 1 | Carbamoyl-phosphate synthase small chain |
| P52643_LDHD_ECOLI | 0 | -1.1 | 3 | 1 | D-lactate dehydrogenase |
| P31057_PANB_ECOLI | 0 | -1.1 | 7 | 1 | 3-methyl-2-oxobutanoate hydroxymethyltransferase |
| P00561_AK1H_ECOLI | 0 | -1.2 | 24 | 2 | Bifunctional aspartokinase/homoserine dehydrogenase I |
| P69503_APT_ECOLI | 0 | -1.2 | 14 | 2 | Adenine phosphoribosyltransferase |
| P11447_ARLY_ECOLI | 0 | -1.2 | 14 | 2 | Argininosuccinate lyase |
| P0A6E4_ASSY_ECOLI | 0 | -1.2 | 32 | 2 | Argininosuccinate synthase |
| P00968_CARB_ECOLI | 0 | -1.2 | 48 | 2 | Carbamoyl-phosphate synthase large chain |
| P0AC41_DHSA_ECOLI | 0 | -1.2 | 14 | 2 | Succinate dehydrogenase flavoprotein subunit |
| P0A6M8_EFG_ECOLI | 0 | -1.2 | 262 | 2 | Elongation factor G |
| P0A993_F16P_ECOLI | 0 | -1.2 | 12 | 2 | Fructose-1,6-bisphosphatase |
| P76658_HLDE_ECOLI | 0 | -1.2 | 7 | 2 | Bifunctional protein hldE |
| P0AEZ1_METF_ECOLI | 0 | -1.2 | 16 | 2 | 5,10-methylenetetrahydrofolate reductase |
| P15640_PUR2_ECOLI | 0 | -1.2 | 19 | 2 | Phosphoribosylamine--glycine ligase |
| P0A8V2_RPOB_ECOLI | 0 | -1.2 | 39 | 2 | DNA-directed RNA polymerase subunit beta |
| P00961_SYGB_ECOLI | 0 | -1.2 | 13 | 2 | Glycyl-tRNA synthetase beta subunit |
| P00956_SYI_ECOLI | 0 | -1.2 | 10 | 2 | Isoleucyl-tRNA synthetase |
| P0A998_FTNA_ECOLI | 0 | -1.2 | 4 | 1 | Ferritin-1 |
| P08142_ILVB_ECOLI | 0 | -1.2 | 4 | 1 | Acetolactate synthase isozyme 1 large subunit |
| P0A7B1_PPK_ECOLI | 0 | -1.2 | 3 | 1 | Polyphosphate kinase |
| P0AEP3_GALU_ECOLI | 0 | -1.3 | 35 | 2 | UTP--glucose-1-phosphate uridylyltransferase |
| P37689_GPMI_ECOLI | 0 | -1.3 | 59 | 2 | 2,3-bisphosphoglycerate-independent phosphoglycerate mutase |
| P23893_GSA_ECOLI | 0 | -1.3 | 13 | 2 | Glutamate-1-semialdehyde 2,1-aminomutase |
| P0A9X4_MREB_ECOLI | 0 | -1.3 | 18 | 2 | Rod shape-determining protein mreB. |
| P0A763_NDK_ECOLI | 0 | -1.3 | 20 | 2 | Nucleoside diphosphate kinase |
| P27298_OPDA_ECOLI | 0 | -1.3 | 24 | 2 | Oligopeptidase A |
| P78067_YNJE_ECOLI | 0 | -1.3 | 80 | 2 | Putative thiosulfate sulfurtransferase ynjE |
| P0A6H5_HSLU_ECOLI | 0 | -1.3 | 3 | 1 | ATP-dependent hsl protease ATP-binding subunit hslU |
| P62620_ISPG_ECOLI | 0 | -1.3 | 3 | 1 | 4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase |

| | | | | | |
|---|---|---|---|---|---|
| P0A6C8_ARGB_ECOLI | 0 | -1.4 | 12 | 2 | Acetylglutamate kinase |
| P18335_ARGD_ECOLI | 0 | -1.4 | 16 | 2 | Acetylornithine/succinyldiaminopimelate aminotransferase |
| P0A6N1_EFTU_ECOLI | 0 | -1.4 | 298 | 2 | Elongation factor Tu |
| P0AES4_GYRA_ECOLI | 0 | -1.4 | 12 | 2 | DNA gyrase subunit A |
| P0A7D7_PUR7_ECOLI | 0 | -1.4 | 32 | 2 | Phosphoribosylaminoimidazole-succinocarboxamide synthase |
| P05020_PYRC_ECOLI | 0 | -1.4 | 36 | 2 | Dihydroorotase |
| P00909_TRPC_ECOLI | 0 | -1.4 | 15 | 2 | Tryptophan biosynthesis protein trpCF |
| P76142_YNEA_ECOLI | 0 | -1.4 | 44 | 2 | Uncharacterized protein yneA |
| Q59385_ATCU_ECOLI | 0 | -1.4 | 9 | 1 | Copper-transporting P-type ATPase |
| P0AES0_GSP_ECOLI | 0 | -1.4 | 3 | 1 | Bifunctional glutathionylspermidine synthetase/amidase |
| P0AEZ3_MIND_ECOLI | 0 | -1.5 | 15 | 2 | Septum site-determining protein minD |
| P23538_PPSA_ECOLI | 0 | -1.5 | 7 | 2 | Phosphoenolpyruvate synthase |
| P23857_PSPE_ECOLI | 0 | -1.5 | 59 | 2 | Phage shock protein E |
| P0A8T7_RPOC_ECOLI | 0 | -1.5 | 29 | 2 | DNA-directed RNA polymerase subunit beta' |
| P21888_SYC_ECOLI | 0 | -1.5 | 12 | 2 | Cysteinyl-tRNA synthetase |
| P0A8A2_YEEN_ECOLI | 0 | -1.5 | 9 | 2 | UPF0082 protein yeeN. |
| P0ADP9_YIHD_ECOLI | 0 | -1.5 | 25 | 2 | Protein yihD. |
| P07395_SYFB_ECOLI | 0 | -1.5 | 9 | 1 | Phenylalanyl-tRNA synthetase beta chain |
| P29217_YCEH_ECOLI | 0 | -1.5 | 5 | 1 | UPF0502 protein yceH |
| P0ABU2_ENGD_ECOLI | 0 | -1.6 | 16 | 2 | GTP-dependent nucleic acid-binding protein engD. |
| P0A9S5_GLDA_ECOLI | 0 | -1.6 | 10 | 2 | Glycerol dehydrogenase |
| P09831_GLTB_ECOLI | 0 | -1.6 | 64 | 2 | Glutamate synthase [NADPH] large chain |
| P22259_PPCK_ECOLI | 0 | -1.6 | 32 | 2 | Phosphoenolpyruvate carboxykinase [ATP] |
| P76536_YFEX_ECOLI | 0 | -1.6 | 8 | 2 | Uncharacterized protein yfeX. |
| P24203_YJIA_ECOLI | 0 | -1.6 | 6 | 2 | Uncharacterized GTP-binding protein yjiA. |
| P31677_OTSA_ECOLI | 0 | -1.6 | 3 | 1 | Alpha,alpha-trehalose-phosphate synthase [UDP-forming] |
| P77318_YDEN_ECOLI | 0 | -1.6 | 3 | 1 | Uncharacterized sulfatase ydeN |
| P36683_ACON2_ECOLI | 0 | -1.7 | 54 | 2 | Aconitate hydratase 2 |
| P37330_MASZ_ECOLI | 0 | -1.7 | 16 | 2 | Malate synthase G |
| P0AB14_YCCJ_ECOLI | 0 | -1.7 | 10 | 2 | Uncharacterized protein yccJ. |
| P09147_GALE_ECOLI | 0 | -1.7 | 3 | 1 | UDP-glucose 4-epimerase |
| P17846_CYSI_ECOLI | 0 | -1.8 | 22 | 2 | Sulfite reductase [NADPH] hemoprotein beta-component |
| P37192_GATY_ECOLI | 0 | -1.8 | 29 | 2 | Tagatose-1,6-bisphosphate aldolase gatY |

| | | | | | |
|---|---|---|---|---|---|
| P37759_RFBB_ECOLI | 0 | -1.8 | 14 | 2 | dTDP-glucose 4,6-dehydratase |
| P0A9W3_YJJK_ECOLI | 0 | -1.8 | 31 | 2 | Uncharacterized ABC transporter ATP-binding protein yjjK. |
| P0A9J6_RBSK_ECOLI | 0 | -1.8 | 3 | 1 | Ribokinase |
| P0A6F3_GLPK_ECOLI | 0 | -1.9 | 68 | 2 | Glycerol kinase |
| P05791_ILVD_ECOLI | 0 | -1.9 | 15 | 2 | Dihydroxy-acid dehydratase |
| P30126_LEUD_ECOLI | 0 | -1.9 | 47 | 2 | 3-isopropylmalate dehydratase small subunit |
| P00934_THRC_ECOLI | 0 | -1.9 | 71 | 2 | Threonine synthase |
| P0A915_OMPW_ECOLI | 0 | -1.9 | 6 | 1 | Outer membrane protein W |
| P07623_META_ECOLI | 0 | -2 | 12 | 2 | Homoserine O-succinyltransferase |
| P0AC33_FUMA_ECOLI | 0 | -2 | 14 | 1 | Fumarate hydratase class I, aerobic |
| P0A6T3_GAL1_ECOLI | 0 | -2 | 6 | 1 | Galactokinase |
| P0A6W9_GSH1_ECOLI | 0 | -2 | 8 | 1 | Glutamate--cysteine ligase |
| P0AG16_PUR1_ECOLI | 0 | -2.1 | 7 | 2 | Amidophosphoribosyltransferase |
| P0AAC0_USPE_ECOLI | 0 | -2.1 | 33 | 2 | Universal stress protein E. |
| P0AFG3_ODO1_ECOLI | 0 | -2.1 | 7 | 1 | 2-oxoglutarate dehydrogenase E1 component |
| P69828_PTKA_ECOLI | 0 | -2.2 | 13 | 2 | Galactitol-specific phosphotransferase enzyme IIA component |
| P07639_AROB_ECOLI | 0 | -2.3 | 10 | 2 | 3-dehydroquinate synthase |
| P0AGD3_SODF_ECOLI | 0 | -2.5 | 14 | 2 | Superoxide dismutase [Fe] |
| P0A9P4_TRXB_ECOLI | 0 | -2.5 | 4 | 1 | Thioredoxin reductase |
| P07650_TYPH_ECOLI | 0 | -2.5 | 10 | 1 | Thymidine phosphorylase |
| P13033_GLPB_ECOLI | 0 | -2.6 | 7 | 1 | Anaerobic glycerol-3-phosphate dehydrogenase subunit B |
| P10408_SECA_ECOLI | 0 | -2.6 | 4 | 1 | Protein translocase subunit secA. |
| P25516_ACON1_ECOLI | 0 | -2.7 | 44 | 2 | Aconitate hydratase 1 |
| P13029_CATA_ECOLI | 0 | -2.7 | 40 | 2 | Peroxidase/catalase HPI |
| P0A9C3_GALM_ECOLI | 0 | -2.7 | 16 | 2 | Aldose 1-epimerase |
| P0AF36_YIIU_ECOLI | 0 | -2.7 | 22 | 2 | Uncharacterized protein yiiU. |
| P40874_MTOX_ECOLI | 0 | -2.7 | 3 | 1 | N-methyl-L-tryptophan oxidase |
| P27550_ACSA_ECOLI | 0 | -3.1 | 35 | 2 | Acetyl-coenzyme A synthetase |
| P67080_YGGS_ECOLI | 0 | -3.1 | 7 | 2 | UPF0001 protein yggS. |
| P0A6A6_LEU2_ECOLI | 0 | -3.3 | 45 | 2 | 3-isopropylmalate dehydratase large subunit |
| P0A9M8_PTA_ECOLI | 0 | -3.4 | 13 | 2 | Phosphate acetyltransferase |
| P12281_MOEA_ECOLI | 0 | -3.4 | 8 | 1 | Molybdopterin biosynthesis protein moeA. |
| P37191_GATZ_ECOLI | 0 | -3.6 | 13 | 2 | Putative tagatose 6-phosphate kinase gatZ |

| | | | | | |
|---|---|---|---|---|---|
| P39315_YTFG_ECOLI | 0 | -3.9 | 6 | 1 | Uncharacterized oxidoreductase ytfG |
| P0A9C0_GLPA_ECOLI | 0 | -4.1 | 4 | 1 | Anaerobic glycerol-3-phosphate dehydrogenase subunit A |

# Appendix G
# GdCl3 Isolation Protocol, All PP:PP Protein Isotope Ratios

| Protein | PP | log2ratio | Chr. Features | Reps | Protein Description |
|---|---|---|---|---|---|
| P0AEL6_FEPB_ECOLI | 1 | 1.3 | 12 | 1 | Ferrienterobactin-binding periplasmic protein |
| P0AA57_YOBA_ECOLI | 1 | 1 | 7 | 1 | Protein yobA |
| P77754_SPY_ECOLI | 1 | 0.8 | 20 | 2 | Spheroplast protein Y |
| P0ADV1_YHBN_ECOLI | 1 | 0.6 | 11 | 2 | Protein yhbN |
| P0C0V0_DEGP_ECOLI | 1 | 0.3 | 34 | 2 | Protease do |
| P45523_FKBA_ECOLI | 1 | 0.3 | 18 | 2 | FKBP-type peptidyl-prolyl cis-trans isomerase fkpA |
| P0AB24_YCDO_ECOLI | 1 | 0.2 | 22 | 2 | UPF0409 protein ycdO |
| P0AEM9_FLIY_ECOLI | 1 | 0.1 | 238 | 2 | Cystine-binding periplasmic protein |
| P0AD59_IVY_ECOLI | 1 | -0.2 | 19 | 2 | Inhibitor of vertebrate lysozyme |
| P0AD96_LIVJ_ECOLI | 1 | -0.4 | 520 | 2 | Leu/Ile/Val-binding protein |
| P0AFL3_PPIA_ECOLI | 1 | -0.5 | 40 | 2 | Peptidyl-prolyl cis-trans isomerase A |
| P39172_ZNUA_ECOLI | 1 | -0.5 | 106 | 2 | High-affinity zinc uptake system protein znuA |
| P64596_YRAP_ECOLI | 1 | -0.5 | 6 | 1 | Uncharacterized protein yraP |
| P37329_MODA_ECOLI | 1 | -0.6 | 226 | 2 | Molybdate-binding periplasmic protein |
| P23865_PRC_ECOLI | 1 | -0.7 | 8 | 1 | Tail-specific protease |
| P0A9L3_FKBB_ECOLI | 1 | -0.8 | 11 | 2 | FKBP-type 22 kDa peptidyl-prolyl cis-trans isomerase |
| P0ABZ6_SURA_ECOLI | 1 | -0.8 | 115 | 2 | Chaperone surA |
| P0ADA1_TESA_ECOLI | 1 | -0.8 | 21 | 2 | Acyl-CoA thioesterase I |
| P0AEG4_DSBA_ECOLI | 1 | -0.9 | 48 | 2 | Thiol:disulfide interchange protein dsbA |
| P61316_LOLA_ECOLI | 1 | -0.9 | 48 | 2 | Outer-membrane lipoprotein carrier protein |
| P0AE85_CPXP_ECOLI | 1 | -0.9 | 3 | 1 | Periplasmic protein cpxP |
| P31697_FIMC_ECOLI | 1 | -0.9 | 3 | 1 | Chaperone protein fimC |
| P30859_ARTI_ECOLI | 1 | -1 | 157 | 2 | Arginine-binding periplasmic protein 1 |
| P06610_BTUE_ECOLI | 1 | -1 | 13 | 2 | Vitamin B12 transport periplasmic protein btuE. |
| P0AES9_HDEA_ECOLI | 1 | -1 | 53 | 2 | Protein hdeA |
| P0AFM2_PROX_ECOLI | 1 | -1 | 131 | 2 | Glycine betaine-binding periplasmic protein |
| P31550_THIB_ECOLI | 1 | -1 | 49 | 2 | Thiamine-binding periplasmic protein |
| P0A855_TOLB_ECOLI | 1 | -1 | 66 | 2 | Protein tolB |
| P0A862_TPX_ECOLI | 1 | -1 | 69 | 2 | Thiol peroxidase |
| P13482_TREA_ECOLI | 1 | -1 | 29 | 2 | Periplasmic trehalase |

| | | | | | |
|---|---|---|---|---|---|
| P07024_USHA_ECOLI | 1 | -1 | 69 | 2 | Protein ushA |
| P23827_ECOT_ECOLI | 1 | -1.1 | 9 | 2 | Ecotin |
| P18956_GGT_ECOLI | 1 | -1.1 | 8 | 1 | Gamma-glutamyltranspeptidase |
| P00811_AMPC_ECOLI | 1 | -1.2 | 12 | 2 | Beta-lactamase |
| P37902_GLTI_ECOLI | 1 | -1.2 | 71 | 2 | Glutamate/aspartate periplasmic-binding protein |
| P75797_GSIB_ECOLI | 1 | -1.2 | 52 | 2 | Glutathione-binding protein gsiB |
| P0AFH8_OSMY_ECOLI | 1 | -1.2 | 34 | 2 | Osmotically-inducible protein Y |
| P16700_CYSP_ECOLI | 1 | -1.3 | 158 | 2 | Thiosulfate-binding protein |
| P0AFK9_POTD_ECOLI | 1 | -1.3 | 153 | 2 | Spermidine/putrescine-binding periplasmic protein |
| P0AGC3_SLT_ECOLI | 1 | -1.3 | 34 | 2 | Soluble lytic murein transglycosylase |
| P0AG78_SUBI_ECOLI | 1 | -1.3 | 41 | 2 | Sulfate-binding protein |
| P76193_YNHG_ECOLI | 1 | -1.3 | 70 | 2 | Uncharacterized protein ynhG |
| P09551_ARGT_ECOLI | 1 | -1.4 | 121 | 2 | Lysine-arginine-ornithine-binding periplasmic protein |
| P04816_LIVK_ECOLI | 1 | -1.4 | 320 | 2 | Leucine-specific-binding protein |
| P0AEU7_SKP_ECOLI | 1 | -1.4 | 60 | 2 | Chaperone protein skp |
| P08331_CN16_ECOLI | 1 | -1.5 | 65 | 2 | 2',3'-cyclic-nucleotide 2'-phosphodiesterase |
| P0AEQ3_GLNH_ECOLI | 1 | -1.5 | 77 | 2 | Glutamine-binding periplasmic protein |
| P0AEU0_HISJ_ECOLI | 1 | -1.5 | 195 | 2 | Histidine-binding periplasmic protein |
| P00634_PPB_ECOLI | 1 | -1.5 | 15 | 2 | Alkaline phosphatase |
| P37648_YHJJ_ECOLI | 1 | -1.5 | 26 | 2 | Protein yhjJ |
| P21338_RNI_ECOLI | 1 | -1.5 | 10 | 1 | Ribonuclease I |
| Q46863_YGIS_ECOLI | 1 | -1.5 | 5 | 1 | Putative binding protein ygiS |
| P0AFX9_RSEB_ECOLI | 1 | -1.6 | 7 | 1 | Sigma-E factor regulatory protein rseB |
| P39265_ALSB_ECOLI | 1 | -1.7 | 6 | 2 | D-allose-binding periplasmic protein |
| P0AE22_APHA_ECOLI | 1 | -1.8 | 19 | 2 | Class B acid phosphatase |
| P23843_OPPA_ECOLI | 1 | -1.8 | 349 | 2 | Periplasmic oligopeptide-binding protein |
| P05458_PTRA_ECOLI | 1 | -1.9 | 28 | 2 | Protease 3 |
| P0AGD1_SODC_ECOLI | 1 | -1.9 | 21 | 2 | Superoxide dismutase [Cu-Zn] |
| P39099_DEGQ_ECOLI | 1 | -1.9 | 6 | 1 | Protease degQ |
| P33136_OPGG_ECOLI | 1 | -1.9 | 20 | 1 | Glucans biosynthesis protein G |
| P76108_YDCS_ECOLI | 1 | -1.9 | 16 | 1 | Putative ABC transporter periplasmic-binding protein ydcS |
| P0AG82_PSTS_ECOLI | 1 | -2 | 15 | 2 | Phosphate-binding protein pstS |
| P0AEE5_DGAL_ECOLI | 1 | -2.1 | 78 | 2 | D-galactose-binding periplasmic protein |

| | | | | | |
|---|---|---|---|---|---|
| P39325_YTFQ_ECOLI | 1 | -2.1 | 148 | 2 | ABC transporter periplasmic-binding protein ytfQ |
| P23847_DPPA_ECOLI | 1 | -2.2 | 319 | 2 | Periplasmic dipeptide transport protein |
| P39176_ERFK_ECOLI | 1 | -2.2 | 11 | 2 | Protein erfK/srfK |
| P40120_OPGD_ECOLI | 1 | -2.2 | 29 | 2 | Glucans biosynthesis protein D |
| P0AAX8_YBIS_ECOLI | 1 | -2.2 | 40 | 2 | Uncharacterized protein ybiS |
| P19926_AGP_ECOLI | 1 | -2.3 | 113 | 2 | Glucose-1-phosphatase |
| P0AEG6_DSBC_ECOLI | 1 | -2.3 | 21 | 2 | Thiol:disulfide interchange protein dsbC |
| P0AEX9_MALE_ECOLI | 1 | -2.3 | 73 | 2 | Maltose-binding periplasmic protein |
| P30860_ARTJ_ECOLI | 1 | -2.6 | 187 | 2 | Arginine-binding periplasmic protein 2 |
| P09394_GLPQ_ECOLI | 1 | -2.6 | 64 | 2 | Glycerophosphoryl diester phosphodiesterase |
| P0AG80_UGPB_ECOLI | 1 | -2.7 | 90 | 2 | sn-glycerol-3-phosphate-binding periplasmic protein ugpB |
| P07102_PPA_ECOLI | 1 | -2.7 | 11 | 1 | Periplasmic appA protein |
| P00805_ASPG2_ECOLI | 1 | -3 | 16 | 2 | L-asparaginase 2 |
| P02925_RBSB_ECOLI | 1 | -3.3 | 101 | 2 | D-ribose-binding periplasmic protein |
| P33937_NAPA_ECOLI | 1 | -3.5 | 8 | 1 | Periplasmic nitrate reductase |
| P02924_ARAF_ECOLI | 1 | -3.6 | 24 | 2 | L-arabinose-binding periplasmic protein |
| P36649_CUEO_ECOLI | 1 | -4.1 | 71 | 2 | Blue copper oxidase cueO |
| P0ABK9_NRFA_ECOLI | 1 | -6.2 | 32 | 1 | Cytochrome c-552 |
| P17315_CIRA_ECOLI | 0 | 5.1 | 93 | 2 | Colicin I receptor |
| P75780_FIU_ECOLI | 0 | 4.9 | 43 | 2 | Catecholate siderophore receptor fiu |
| P0AEJ2_ENTC_ECOLI | 0 | 4 | 18 | 2 | Isochorismate synthase entC |
| P05825_FEPA_ECOLI | 0 | 3.8 | 132 | 2 | Ferrienterobactin receptor |
| P76116_YNCE_ECOLI | 0 | 2.2 | 167 | 2 | Uncharacterized protein yncE |
| P10100_RLPA_ECOLI | 0 | 2.2 | 4 | 1 | Rare lipoprotein A |
| P0A9X9_CSPA_ECOLI | 0 | 2.1 | 4 | 1 | Cold shock protein cspA |
| P0AB40_YCFR_ECOLI | 0 | 1.8 | 7 | 2 | UPF0379 protein ycfR |
| P0A910_OMPA_ECOLI | 0 | 1.7 | 131 | 2 | Outer membrane protein A |
| P68191_SRA_ECOLI | 0 | 1.6 | 21 | 2 | Stationary-phase-induced ribosome-associated protein |
| P0A917_OMPX_ECOLI | 0 | 1.4 | 40 | 2 | Outer membrane protein X |
| P0A905_SLYB_ECOLI | 0 | 1.4 | 23 | 2 | Outer membrane lipoprotein slyB |
| P68066_GRCA_ECOLI | 0 | 1.3 | 88 | 2 | Autonomous glycyl radical cofactor. |
| P0A800_RPOZ_ECOLI | 0 | 1.3 | 8 | 2 | DNA-directed RNA polymerase subunit omega |
| P0A7N4_RL32_ECOLI | 0 | 1.1 | 8 | 2 | 50S ribosomal protein L32. |

| | | | | | |
|---|---|---|---|---|---|
| P76014_DHAL_ECOLI | 0 | 1.1 | 5 | 1 | PTS-dependent dihydroxyacetone kinase, ADP-binding subunit d |
| P0A7N9_RL33_ECOLI | 0 | 1 | 11 | 2 | 50S ribosomal protein L33. |
| P0A8M6_YEEX_ECOLI | 0 | 1 | 12 | 2 | UPF0265 protein yeeX. |
| P64463_YDFZ_ECOLI | 0 | 0.9 | 9 | 2 | Putative selenoprotein ydfZ. |
| P68919_RL25_ECOLI | 0 | 0.9 | 16 | 1 | 50S ribosomal protein L25. |
| P0ADB1_OSME_ECOLI | 0 | 0.8 | 36 | 2 | Osmotically-inducible lipoprotein E |
| P60624_RL24_ECOLI | 0 | 0.8 | 18 | 2 | 50S ribosomal protein L24. |
| P0AE06_ACRA_ECOLI | 0 | 0.8 | 3 | 1 | Acriflavine resistance protein A |
| P0AEH5_ELAB_ECOLI | 0 | 0.8 | 4 | 1 | Protein elaB. |
| P62707_GPMA_ECOLI | 0 | 0.7 | 149 | 2 | 2,3-bisphosphoglycerate-dependent phosphoglycerate mutase |
| P0A786_PYRB_ECOLI | 0 | 0.7 | 12 | 2 | Aspartate carbamoyltransferase catalytic chain |
| P00448_SODM_ECOLI | 0 | 0.7 | 66 | 2 | Superoxide dismutase [Mn] |
| P0ACW4_YDCA_ECOLI | 0 | 0.7 | 24 | 2 | Uncharacterized protein ydcA |
| P33219_YEBF_ECOLI | 0 | 0.7 | 15 | 2 | Protein yebF |
| P30139_THIG_ECOLI | 0 | 0.7 | 3 | 1 | Thiazole biosynthesis protein thiG. |
| P0ACF0_DBHA_ECOLI | 0 | 0.6 | 59 | 2 | DNA-binding protein HU-alpha |
| P0ACF4_DBHB_ECOLI | 0 | 0.6 | 32 | 2 | DNA-binding protein HU-beta |
| P76009_EMTA_ECOLI | 0 | 0.6 | 8 | 2 | Endotype membrane-bound lytic murein transglycosylase A |
| P0A7M6_RL29_ECOLI | 0 | 0.6 | 37 | 2 | 50S ribosomal protein L29. |
| P02930_TOLC_ECOLI | 0 | 0.6 | 14 | 2 | Outer membrane protein tolC |
| P0AG51_RL30_ECOLI | 0 | 0.6 | 4 | 1 | 50S ribosomal protein L30. |
| P0A6Y1_IHFB_ECOLI | 0 | 0.5 | 11 | 2 | Integration host factor subunit beta |
| P0ADE6_YGAU_ECOLI | 0 | 0.5 | 13 | 2 | Uncharacterized protein ygaU. |
| P68679_RS21_ECOLI | 0 | 0.5 | 10 | 1 | 30S ribosomal protein S21. |
| P37194_SLP_ECOLI | 0 | 0.5 | 12 | 1 | Outer membrane protein slp |
| P0AB38_YCFM_ECOLI | 0 | 0.5 | 5 | 1 | Uncharacterized protein ycfM. |
| P31828_PQQL_ECOLI | 0 | 0.4 | 11 | 2 | Probable zinc protease pqqL |
| P0AA04_PTHP_ECOLI | 0 | 0.4 | 23 | 2 | Phosphocarrier protein HPr |
| P0A7U7_RS20_ECOLI | 0 | 0.4 | 18 | 2 | 30S ribosomal protein S20. |
| P0AF70_YJEI_ECOLI | 0 | 0.4 | 6 | 2 | Uncharacterized protein yjeI |
| P62768_YAEH_ECOLI | 0 | 0.4 | 3 | 1 | UPF0325 protein yaeH. |
| P64503_YEBV_ECOLI | 0 | 0.4 | 7 | 1 | Uncharacterized protein yebV. |
| P31063_YEDD_ECOLI | 0 | 0.4 | 3 | 1 | Uncharacterized lipoprotein yedD |

| | | | | | |
|---|---|---|---|---|---|
| P0ACF8_HNS_ECOLI | 0 | 0.3 | 18 | 2 | DNA-binding protein H-NS |
| P09169_OMPT_ECOLI | 0 | 0.3 | 125 | 2 | Protease 7 |
| P0A7J3_RL10_ECOLI | 0 | 0.3 | 12 | 2 | 50S ribosomal protein L10 |
| P0A7U3_RS19_ECOLI | 0 | 0.3 | 11 | 2 | 30S ribosomal protein S19. |
| P36995_CSPB_ECOLI | 0 | 0.3 | 3 | 1 | Cold shock-like protein cspB |
| P0AEZ9_MOAB_ECOLI | 0 | 0.3 | 3 | 1 | Molybdenum cofactor biosynthesis protein B. |
| P0A912_PAL_ECOLI | 0 | 0.2 | 11 | 2 | Peptidoglycan-associated lipoprotein |
| P0A7J7_RL11_ECOLI | 0 | 0.2 | 17 | 2 | 50S ribosomal protein L11. |
| P0ADY7_RL16_ECOLI | 0 | 0.2 | 35 | 2 | 50S ribosomal protein L16. |
| P0A877_TRPA_ECOLI | 0 | 0.2 | 46 | 2 | Tryptophan synthase alpha chain |
| P0A8B5_YBAB_ECOLI | 0 | 0.2 | 16 | 2 | UPF0133 protein ybaB. |
| P0A707_IF3_ECOLI | 0 | 0.2 | 9 | 1 | Translation initiation factor IF-3. |
| P34210_OMPP_ECOLI | 0 | 0.2 | 4 | 1 | Outer membrane protease ompP |
| P0A7W7_RS8_ECOLI | 0 | 0.2 | 5 | 1 | 30S ribosomal protein S8. |
| P04949_FLIC_ECOLI | 0 | 0.1 | 918 | 2 | Flagellin. |
| P61175_RL22_ECOLI | 0 | 0.1 | 30 | 2 | 50S ribosomal protein L22. |
| P0ACG1_STPA_ECOLI | 0 | 0.1 | 12 | 2 | DNA-binding protein stpA |
| P76002_YCGK_ECOLI | 0 | 0.1 | 13 | 2 | Uncharacterized protein ycgK |
| P64451_YDCL_ECOLI | 0 | 0.1 | 11 | 2 | Uncharacterized lipoprotein ydcL |
| P0AC02_YFIO_ECOLI | 0 | 0.1 | 16 | 2 | UPF0169 lipoprotein yfiO |
| P0ADZ0_RL23_ECOLI | 0 | 0.1 | 4 | 1 | 50S ribosomal protein L23. |
| P0AA10_RL13_ECOLI | 0 | 0 | 13 | 2 | 50S ribosomal protein L13. |
| P0AEK4_FABI_ECOLI | 0 | 0 | 6 | 1 | Enoyl-[acyl-carrier-protein] reductase [NADH] |
| P04128_FIMA1_ECOLI | 0 | 0 | 4 | 1 | Type-1 fimbrial protein, A chain |
| P12281_MOEA_ECOLI | 0 | 0 | 3 | 1 | Molybdopterin biosynthesis protein moeA. |
| P0AG59_RS14_ECOLI | 0 | 0 | 11 | 1 | 30S ribosomal protein S14. |
| P0A8E7_YAJQ_ECOLI | 0 | 0 | 4 | 1 | UPF0234 protein yajQ. |
| P77625_YFBT_ECOLI | 0 | 0 | 5 | 1 | Phosphatase yfbT |
| P0AB46_YMGD_ECOLI | 0 | 0 | 3 | 1 | Uncharacterized protein ymgD |
| P0A972_CSPE_ECOLI | 0 | -0.1 | 15 | 2 | Cold shock-like protein cspE |
| P0C0L2_OSMC_ECOLI | 0 | -0.1 | 27 | 2 | Peroxiredoxin osmC |
| P02413_RL15_ECOLI | 0 | -0.1 | 25 | 2 | 50S ribosomal protein L15. |
| P0AG55_RL6_ECOLI | 0 | -0.1 | 38 | 2 | 50S ribosomal protein L6. |

| | | | | | |
|---|---|---|---|---|---|
| P0A7S3_RS12_ECOLI | 0 | -0.1 | 9 | 2 | 30S ribosomal protein S12. |
| P36659_CBPA_ECOLI | 0 | -0.1 | 3 | 1 | Curved DNA-binding protein. |
| P00935_METB_ECOLI | 0 | -0.1 | 3 | 1 | Cystathionine gamma-synthase |
| P0A903_NLPB_ECOLI | 0 | -0.1 | 7 | 1 | Lipoprotein 34 |
| P0AAV6_YBGS_ECOLI | 0 | -0.1 | 4 | 1 | Uncharacterized protein ybgS |
| P0AFG6_ODO2_ECOLI | 0 | -0.2 | 15 | 2 | Dihydrolipoyllysine-residue succinyltransferase component of |
| P0AG44_RL17_ECOLI | 0 | -0.2 | 20 | 2 | 50S ribosomal protein L17. |
| P0C018_RL18_ECOLI | 0 | -0.2 | 38 | 2 | 50S ribosomal protein L18. |
| P0A7M2_RL28_ECOLI | 0 | -0.2 | 7 | 2 | 50S ribosomal protein L28. |
| P60422_RL2_ECOLI | 0 | -0.2 | 35 | 2 | 50S ribosomal protein L2. |
| P0A7T3_RS16_ECOLI | 0 | -0.2 | 19 | 2 | 30S ribosomal protein S16. |
| P0A7T7_RS18_ECOLI | 0 | -0.2 | 15 | 2 | 30S ribosomal protein S18. |
| P69776_LPP_ECOLI | 0 | -0.2 | 6 | 1 | Major outer membrane lipoprotein |
| P0A935_MLTA_ECOLI | 0 | -0.2 | 6 | 1 | Membrane-bound lytic murein transglycosylase A |
| P69829_PTSN_ECOLI | 0 | -0.2 | 4 | 1 | Nitrogen regulatory protein |
| P21513_RNE_ECOLI | 0 | -0.2 | 5 | 1 | Ribonuclease E |
| P0ACY1_YDJA_ECOLI | 0 | -0.2 | 3 | 1 | Protein ydjA. |
| P0AEK2_FABG_ECOLI | 0 | -0.3 | 13 | 2 | 3-oxoacyl-[acyl-carrier-protein] reductase |
| P0A7K2_RL7_ECOLI | 0 | -0.3 | 10 | 2 | 50S ribosomal protein L7/L12 |
| P0A7W1_RS5_ECOLI | 0 | -0.3 | 39 | 2 | 30S ribosomal protein S5. |
| P45464_YRAM_ECOLI | 0 | -0.3 | 10 | 2 | Uncharacterized protein yraM. |
| P52697_6PGL_ECOLI | 0 | -0.3 | 3 | 1 | 6-phosphogluconolactonase |
| P0A8P3_FETP_ECOLI | 0 | -0.3 | 8 | 1 | Probable Fe(2+)-trafficking protein. |
| P0A8U6_METJ_ECOLI | 0 | -0.3 | 4 | 1 | Met repressor |
| P17117_NFSA_ECOLI | 0 | -0.3 | 4 | 1 | Oxygen-insensitive NADPH nitroreductase |
| P0A7K6_RL19_ECOLI | 0 | -0.3 | 15 | 1 | 50S ribosomal protein L19. |
| P77499_SUFC_ECOLI | 0 | -0.3 | 3 | 1 | Probable ATP-dependent transporter sufC. |
| P75818_YBJP_ECOLI | 0 | -0.3 | 5 | 1 | Uncharacterized lipoprotein ybjP |
| P0AB55_YCII_ECOLI | 0 | -0.3 | 5 | 1 | Protein yciI. |
| P0AD07_YECF_ECOLI | 0 | -0.3 | 3 | 1 | Uncharacterized protein yecF. |
| P0ADU2_YGIN_ECOLI | 0 | -0.3 | 5 | 1 | Protein ygiN. |
| P0ADX5_YHFG_ECOLI | 0 | -0.3 | 3 | 1 | Uncharacterized protein yhfG. |
| P0A986_CSPI_ECOLI | 0 | -0.4 | 7 | 2 | Cold shock-like protein cspI |

| | | | | | |
|---|---|---|---|---|---|
| P75937_FLGE_ECOLI | 0 | -0.4 | 67 | 2 | Flagellar hook protein flgE. |
| P60438_RL3_ECOLI | 0 | -0.4 | 17 | 2 | 50S ribosomal protein L3. |
| P0A7R5_RS10_ECOLI | 0 | -0.4 | 13 | 2 | 30S ribosomal protein S10. |
| P0A7R9_RS11_ECOLI | 0 | -0.4 | 21 | 2 | 30S ribosomal protein S11. |
| P39177_USPG_ECOLI | 0 | -0.4 | 30 | 2 | Universal stress protein G. |
| P0ADX7_YHHA_ECOLI | 0 | -0.4 | 72 | 2 | Uncharacterized protein yhhA |
| P0ADN2_YIFE_ECOLI | 0 | -0.4 | 15 | 2 | UPF0438 protein yifE. |
| P0A908_MIPA_ECOLI | 0 | -0.4 | 14 | 1 | MltA-interacting protein |
| P0A7X3_RS9_ECOLI | 0 | -0.4 | 4 | 1 | 30S ribosomal protein S9. |
| P0ADA5_YAJG_ECOLI | 0 | -0.4 | 12 | 1 | Uncharacterized lipoprotein yajG |
| P76004_YCGM_ECOLI | 0 | -0.4 | 4 | 1 | Uncharacterized protein ycgM. |
| Q47702_YFEK_ECOLI | 0 | -0.4 | 5 | 1 | Uncharacterized protein yfeK |
| P64614_YHCN_ECOLI | 0 | -0.4 | 3 | 1 | UPF0379 protein yhcN |
| P00350_6PGD_ECOLI | 0 | -0.5 | 37 | 2 | 6-phosphogluconate dehydrogenase, decarboxylating |
| P0A9Y6_CSPC_ECOLI | 0 | -0.5 | 30 | 2 | Cold shock-like protein cspC |
| P04036_DAPB_ECOLI | 0 | -0.5 | 10 | 2 | Dihydrodipicolinate reductase |
| P0A9B2_G3P1_ECOLI | 0 | -0.5 | 258 | 2 | Glyceraldehyde-3-phosphate dehydrogenase A |
| P28635_METQ_ECOLI | 0 | -0.5 | 15 | 2 | D-methionine-binding lipoprotein metQ |
| P24182_ACCC_ECOLI | 0 | -0.5 | 10 | 1 | Biotin carboxylase |
| P05459_PDXB_ECOLI | 0 | -0.5 | 11 | 1 | Erythronate-4-phosphate dehydrogenase |
| P08179_PUR3_ECOLI | 0 | -0.5 | 4 | 1 | Phosphoribosylglycinamide formyltransferase |
| P0ADZ4_RS15_ECOLI | 0 | -0.5 | 4 | 1 | 30S ribosomal protein S15. |
| P0AAB8_USPD_ECOLI | 0 | -0.5 | 3 | 1 | Universal stress protein D. |
| P0ADP9_YIHD_ECOLI | 0 | -0.5 | 5 | 1 | Protein yihD. |
| P21179_CATE_ECOLI | 0 | -0.6 | 11 | 2 | Catalase HPII |
| P0A9C5_GLNA_ECOLI | 0 | -0.6 | 32 | 2 | Glutamine synthetase |
| P05793_ILVC_ECOLI | 0 | -0.6 | 108 | 2 | Ketol-acid reductoisomerase |
| P69441_KAD_ECOLI | 0 | -0.6 | 78 | 2 | Adenylate kinase |
| P0ACJ0_LRP_ECOLI | 0 | -0.6 | 6 | 2 | Leucine-responsive regulatory protein. |
| P61889_MDH_ECOLI | 0 | -0.6 | 143 | 2 | Malate dehydrogenase |
| P0A799_PGK_ECOLI | 0 | -0.6 | 114 | 2 | Phosphoglycerate kinase |
| P0AG30_RHO_ECOLI | 0 | -0.6 | 12 | 2 | Transcription termination factor rho |
| P0A7R1_RL9_ECOLI | 0 | -0.6 | 29 | 2 | 50S ribosomal protein L9. |

| | | | | | |
|---|---|---|---|---|---|
| P0A805_RRF_ECOLI | 0 | -0.6 | 26 | 2 | Ribosome recycling factor |
| P0A7S9_RS13_ECOLI | 0 | -0.6 | 16 | 2 | 30S ribosomal protein S13. |
| P0AA25_THIO_ECOLI | 0 | -0.6 | 10 | 2 | Thioredoxin-1 |
| P76172_YNFD_ECOLI | 0 | -0.6 | 9 | 2 | Uncharacterized protein ynfD |
| P69503_APT_ECOLI | 0 | -0.6 | 6 | 1 | Adenine phosphoribosyltransferase |
| P15770_AROE_ECOLI | 0 | -0.6 | 5 | 1 | Shikimate dehydrogenase |
| P06129_BTUB_ECOLI | 0 | -0.6 | 15 | 1 | Vitamin B12 transporter btuB |
| P0A9P0_DLDH_ECOLI | 0 | -0.6 | 5 | 1 | Dihydrolipoyl dehydrogenase |
| P0A6Q3_FABA_ECOLI | 0 | -0.6 | 5 | 1 | 3-hydroxydecanoyl-[acyl-carrier-protein] dehydratase |
| P0AEM4_FLGM_ECOLI | 0 | -0.6 | 3 | 1 | Negative regulator of flagellin synthesis |
| P0AAI3_FTSH_ECOLI | 0 | -0.6 | 4 | 1 | Cell division protease ftsH |
| P69797_PTNAB_ECOLI | 0 | -0.6 | 4 | 1 | PTS system mannose-specific EIIAB component |
| P0AE08_AHPC_ECOLI | 0 | -0.7 | 106 | 2 | Alkyl hydroperoxide reductase subunit C |
| P0AB91_AROG_ECOLI | 0 | -0.7 | 24 | 2 | Phospho-2-dehydro-3-deoxyheptonate aldolase, Phe-sensitive |
| P33235_FLGK_ECOLI | 0 | -0.7 | 21 | 2 | Flagellar hook-associated protein 1 |
| P29744_FLGL_ECOLI | 0 | -0.7 | 53 | 2 | Flagellar hook-associated protein 3 |
| P24216_FLID_ECOLI | 0 | -0.7 | 36 | 2 | Flagellar hook-associated protein 2 |
| P0A6T1_G6PI_ECOLI | 0 | -0.7 | 33 | 2 | Glucose-6-phosphate isomerase |
| P0A825_GLYA_ECOLI | 0 | -0.7 | 151 | 2 | Serine hydroxymethyltransferase |
| P23869_PPIB_ECOLI | 0 | -0.7 | 50 | 2 | Peptidyl-prolyl cis-trans isomerase B |
| P61714_RISB_ECOLI | 0 | -0.7 | 17 | 2 | 6,7-dimethyl-8-ribityllumazine synthase |
| P0A7L0_RL1_ECOLI | 0 | -0.7 | 39 | 2 | 50S ribosomal protein L1. |
| P02358_RS6_ECOLI | 0 | -0.7 | 36 | 2 | 30S ribosomal protein S6 |
| P33570_TKT2_ECOLI | 0 | -0.7 | 27 | 2 | Transketolase 2 |
| P0ACX3_YDHR_ECOLI | 0 | -0.7 | 6 | 2 | Protein ydhR |
| P0AF36_YIIU_ECOLI | 0 | -0.7 | 9 | 2 | Uncharacterized protein yiiU. |
| P0AF93_YJGF_ECOLI | 0 | -0.7 | 33 | 2 | UPF0076 protein yjgF. |
| P0ADV7_YRBC_ECOLI | 0 | -0.7 | 33 | 2 | Protein yrbC |
| P0AET8_HDHA_ECOLI | 0 | -0.7 | 3 | 1 | 7-alpha-hydroxysteroid dehydrogenase |
| P0AFG0_NUSG_ECOLI | 0 | -0.7 | 5 | 1 | Transcription antitermination protein nusG. |
| P31663_PANC_ECOLI | 0 | -0.7 | 5 | 1 | Pantothenate synthetase |
| P0A7J0_RIBB_ECOLI | 0 | -0.7 | 6 | 1 | 3,4-dihydroxy-2-butanone 4-phosphate synthase |
| P00956_SYI_ECOLI | 0 | -0.7 | 9 | 1 | Isoleucyl-tRNA synthetase |

| | | | | | |
|---|---|---|---|---|---|
| P76001_YCGJ_ECOLI | 0 | -0.7 | 3 | 1 | Uncharacterized protein ycgJ |
| P0ADI7_YECD_ECOLI | 0 | -0.7 | 4 | 1 | Isochorismatase family protein yecD. |
| P00509_AAT_ECOLI | 0 | -0.8 | 82 | 2 | Aspartate aminotransferase |
| P0A9G6_ACEA_ECOLI | 0 | -0.8 | 71 | 2 | Isocitrate lyase |
| P0A6F5_CH60_ECOLI | 0 | -0.8 | 94 | 2 | 60 kDa chaperonin |
| P08200_IDH_ECOLI | 0 | -0.8 | 130 | 2 | Isocitrate dehydrogenase [NADP] |
| P0A7A9_IPYR_ECOLI | 0 | -0.8 | 29 | 2 | Inorganic pyrophosphatase |
| P08997_MASY_ECOLI | 0 | -0.8 | 93 | 2 | Malate synthase A |
| P06996_OMPC_ECOLI | 0 | -0.8 | 26 | 2 | Outer membrane protein C |
| P02931_OMPF_ECOLI | 0 | -0.8 | 27 | 2 | Outer membrane protein F |
| P02359_RS7_ECOLI | 0 | -0.8 | 36 | 2 | 30S ribosomal protein S7. |
| P0A870_TALB_ECOLI | 0 | -0.8 | 55 | 2 | Transaldolase B |
| P0A6D7_AROK_ECOLI | 0 | -0.8 | 5 | 1 | Shikimate kinase 1 |
| P69910_DCEB_ECOLI | 0 | -0.8 | 3 | 1 | Glutamate decarboxylase beta |
| P0A858_TPIS_ECOLI | 0 | -0.8 | 11 | 1 | Triosephosphate isomerase |
| P0AD12_YEEZ_ECOLI | 0 | -0.8 | 5 | 1 | Protein yeeZ |
| P37659_YHJU_ECOLI | 0 | -0.8 | 5 | 1 | Uncharacterized protein yhjU. |
| P0AGE6_YIEF_ECOLI | 0 | -0.8 | 8 | 1 | Uncharacterized protein yieF. |
| P68206_YJBJ_ECOLI | 0 | -0.8 | 9 | 1 | UPF0337 protein yjbJ. |
| P39315_YTFG_ECOLI | 0 | -0.8 | 7 | 1 | Uncharacterized oxidoreductase ytfG |
| P0ABT2_DPS_ECOLI | 0 | -0.9 | 69 | 2 | DNA protection during starvation protein |
| P0AAI9_FABD_ECOLI | 0 | -0.9 | 31 | 2 | Malonyl CoA-acyl carrier protein transacylase |
| P0A796_K6PF1_ECOLI | 0 | -0.9 | 19 | 2 | 6-phosphofructokinase isozyme 1 |
| P0A817_METK_ECOLI | 0 | -0.9 | 28 | 2 | S-adenosylmethionine synthetase |
| P09373_PFLB_ECOLI | 0 | -0.9 | 172 | 2 | Formate acetyltransferase 1 |
| P0AFM4_PSIF_ECOLI | 0 | -0.9 | 28 | 2 | Phosphate starvation-inducible protein psiF |
| P76177_YDGH_ECOLI | 0 | -0.9 | 101 | 2 | Protein ydgH |
| P0A955_ALKH_ECOLI | 0 | -0.9 | 9 | 1 | KHG/KDPG aldolase |
| P0A6D3_AROA_ECOLI | 0 | -0.9 | 5 | 1 | 3-phosphoshikimate 1-carboxyvinyltransferase |
| P0ABA6_ATPG_ECOLI | 0 | -0.9 | 4 | 1 | ATP synthase gamma chain |
| P08191_FIMH_ECOLI | 0 | -0.9 | 3 | 1 | Protein fimH |
| P0A715_KDSA_ECOLI | 0 | -0.9 | 24 | 1 | 2-dehydro-3-deoxyphosphooctonate aldolase |
| P0A794_PDXJ_ECOLI | 0 | -0.9 | 5 | 1 | Pyridoxine 5'-phosphate synthase |

| | | | | | |
|---|---|---|---|---|---|
| P0A6L0_DEOC_ECOLI | 0 | -1 | 23 | 2 | Deoxyribose-phosphate aldolase |
| P0AET2_HDEB_ECOLI | 0 | -1 | 132 | 2 | Protein hdeB |
| P0AD61_KPYK1_ECOLI | 0 | -1 | 50 | 2 | Pyruvate kinase I |
| P0A744_MSRA_ECOLI | 0 | -1 | 11 | 2 | Peptide methionine sulfoxide reductase msrA |
| P0A7Z0_RPIA_ECOLI | 0 | -1 | 21 | 2 | Ribose-5-phosphate isomerase A |
| P0A7V0_RS2_ECOLI | 0 | -1 | 32 | 2 | 30S ribosomal protein S2. |
| P0A7V8_RS4_ECOLI | 0 | -1 | 23 | 2 | 30S ribosomal protein S4. |
| P23721_SERC_ECOLI | 0 | -1 | 49 | 2 | Phosphoserine aminotransferase |
| P09158_SPEE_ECOLI | 0 | -1 | 8 | 2 | Spermidine synthase |
| P21889_SYD_ECOLI | 0 | -1 | 15 | 2 | Aspartyl-tRNA synthetase |
| P0A8M0_SYN_ECOLI | 0 | -1 | 19 | 2 | Asparaginyl-tRNA synthetase |
| P45955_YBGF_ECOLI | 0 | -1 | 13 | 2 | Uncharacterized protein ybgF |
| P33362_YEHZ_ECOLI | 0 | -1 | 36 | 2 | Uncharacterized protein yehZ |
| P0A991_ALF1_ECOLI | 0 | -1 | 3 | 1 | Fructose-bisphosphate aldolase class 1 |
| P22256_GABT_ECOLI | 0 | -1 | 3 | 1 | 4-aminobutyrate aminotransferase |
| P06715_GSHR_ECOLI | 0 | -1 | 14 | 1 | Glutathione reductase |
| P0AG07_RPE_ECOLI | 0 | -1 | 7 | 1 | Ribulose-phosphate 3-epimerase |
| P0A9P4_TRXB_ECOLI | 0 | -1 | 11 | 1 | Thioredoxin reductase |
| P66948_YFGC_ECOLI | 0 | -1 | 9 | 1 | TPR repeat-containing protein yfgC |
| P39180_AG43_ECOLI | 0 | -1.1 | 513 | 2 | Antigen 43 |
| P0A6F9_CH10_ECOLI | 0 | -1.1 | 18 | 2 | 10 kDa chaperonin |
| P24171_DCP_ECOLI | 0 | -1.1 | 18 | 2 | Peptidyl-dipeptidase dcp |
| P0A9Q9_DHAS_ECOLI | 0 | -1.1 | 96 | 2 | Aspartate-semialdehyde dehydrogenase |
| P0A6P9_ENO_ECOLI | 0 | -1.1 | 249 | 2 | Enolase |
| P0A9C3_GALM_ECOLI | 0 | -1.1 | 20 | 2 | Aldose 1-epimerase |
| P00946_MANA_ECOLI | 0 | -1.1 | 13 | 2 | Mannose-6-phosphate isomerase |
| P25665_METE_ECOLI | 0 | -1.1 | 85 | 2 | 5-methyltetrahydropteroyltriglutamate--homocysteine methyltr |
| P38489_NFNB_ECOLI | 0 | -1.1 | 32 | 2 | Oxygen-insensitive NAD(P)H nitroreductase |
| P31133_POTF_ECOLI | 0 | -1.1 | 33 | 2 | Putrescine-binding periplasmic protein |
| P37744_RMLA1_ECOLI | 0 | -1.1 | 9 | 2 | Glucose-1-phosphate thymidylyltransferase 1 |
| P0A7Z4_RPOA_ECOLI | 0 | -1.1 | 20 | 2 | DNA-directed RNA polymerase subunit alpha |
| P0A9T0_SERA_ECOLI | 0 | -1.1 | 55 | 2 | D-3-phosphoglycerate dehydrogenase |
| P0A836_SUCC_ECOLI | 0 | -1.1 | 62 | 2 | Succinyl-CoA synthetase beta chain |

| | | | | | |
|---|---|---|---|---|---|
| P0AGE9_SUCD_ECOLI | 0 | -1.1 | 10 | 2 | Succinyl-CoA ligase [ADP-forming] subunit alpha |
| P0A8L1_SYS_ECOLI | 0 | -1.1 | 26 | 2 | Seryl-tRNA synthetase |
| P0A9M5_XGPT_ECOLI | 0 | -1.1 | 8 | 2 | Xanthine phosphoribosyltransferase |
| P75694_YAHO_ECOLI | 0 | -1.1 | 31 | 2 | UPF0379 protein yahO |
| P05194_AROD_ECOLI | 0 | -1.1 | 9 | 1 | 3-dehydroquinate dehydratase |
| P63284_CLPB_ECOLI | 0 | -1.1 | 9 | 1 | Chaperone protein clpB |
| P76015_DHAK_ECOLI | 0 | -1.1 | 3 | 1 | PTS-dependent dihydroxyacetone kinase, dihydroxyacetone-bind |
| P77454_GLSA1_ECOLI | 0 | -1.1 | 4 | 1 | Glutaminase 1 |
| P04079_GUAA_ECOLI | 0 | -1.1 | 6 | 1 | GMP synthase [glutamine-hydrolyzing] |
| P15288_PEPD_ECOLI | 0 | -1.1 | 13 | 1 | Aminoacyl-histidine dipeptidase |
| P08244_PYRF_ECOLI | 0 | -1.1 | 3 | 1 | Orotidine 5'-phosphate decarboxylase |
| P33643_RLUD_ECOLI | 0 | -1.1 | 3 | 1 | Ribosomal large subunit pseudouridine synthase D |
| P0A9Q7_ADHE_ECOLI | 0 | -1.2 | 28 | 2 | Aldehyde-alcohol dehydrogenase |
| P0AE52_BCP_ECOLI | 0 | -1.2 | 15 | 2 | Putative peroxiredoxin bcp |
| P0A6G7_CLPP_ECOLI | 0 | -1.2 | 10 | 2 | ATP-dependent Clp protease proteolytic subunit |
| P0AC53_G6PD_ECOLI | 0 | -1.2 | 9 | 2 | Glucose-6-phosphate 1-dehydrogenase |
| P06988_HISX_ECOLI | 0 | -1.2 | 22 | 2 | Histidinol dehydrogenase |
| P0A705_IF2_ECOLI | 0 | -1.2 | 18 | 2 | Translation initiation factor IF-2. |
| P0ADG7_IMDH_ECOLI | 0 | -1.2 | 25 | 2 | Inosine-5'-monophosphate dehydrogenase |
| P21165_PEPQ_ECOLI | 0 | -1.2 | 10 | 2 | Xaa-Pro dipeptidase |
| P15254_PUR4_ECOLI | 0 | -1.2 | 40 | 2 | Phosphoribosylformylglycinamidine synthase |
| P07813_SYL_ECOLI | 0 | -1.2 | 14 | 2 | Leucyl-tRNA synthetase |
| P27302_TKT1_ECOLI | 0 | -1.2 | 39 | 2 | Transketolase 1 |
| P35340_AHPF_ECOLI | 0 | -1.2 | 6 | 1 | Alkyl hydroperoxide reductase subunit F |
| P0A6R0_FABH_ECOLI | 0 | -1.2 | 8 | 1 | 3-oxoacyl-[acyl-carrier-protein] synthase 3 |
| P33195_GCSP_ECOLI | 0 | -1.2 | 3 | 1 | Glycine dehydrogenase [decarboxylating] |
| P13035_GLPD_ECOLI | 0 | -1.2 | 19 | 1 | Aerobic glycerol-3-phosphate dehydrogenase |
| P67910_HLDD_ECOLI | 0 | -1.2 | 7 | 1 | ADP-L-glycero-D-manno-heptose-6-epimerase |
| P06721_METC_ECOLI | 0 | -1.2 | 6 | 1 | Cystathionine beta-lyase |
| P0AF24_NAGD_ECOLI | 0 | -1.2 | 3 | 1 | Protein nagD. |
| P0AFX7_RSEA_ECOLI | 0 | -1.2 | 3 | 1 | Sigma-E factor negative regulatory protein. |
| P0A927_TSX_ECOLI | 0 | -1.2 | 7 | 1 | Nucleoside-specific channel-forming protein tsx |
| P0A8F0_UPP_ECOLI | 0 | -1.2 | 3 | 1 | Uracil phosphoribosyltransferase |

| | | | | | |
|---|---|---|---|---|---|
| P0A6A3_ACKA_ECOLI | 0 | -1.3 | 30 | 2 | Acetate kinase |
| P0AB71_ALF_ECOLI | 0 | -1.3 | 56 | 2 | Fructose-bisphosphate aldolase class 2 |
| P00888_AROF_ECOLI | 0 | -1.3 | 10 | 2 | Phospho-2-dehydro-3-deoxyheptonate aldolase, Tyr-sensitive |
| P0ABK5_CYSK_ECOLI | 0 | -1.3 | 163 | 2 | Cysteine synthase A |
| P0A9D8_DAPD_ECOLI | 0 | -1.3 | 67 | 2 | 2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-succinyltrans |
| P00370_DHE4_ECOLI | 0 | -1.3 | 22 | 2 | NADP-specific glutamate dehydrogenase |
| P31658_HCHA_ECOLI | 0 | -1.3 | 12 | 2 | Chaperone protein hchA |
| P0AB80_ILVE_ECOLI | 0 | -1.3 | 15 | 2 | Branched-chain-amino-acid aminotransferase |
| P30125_LEU3_ECOLI | 0 | -1.3 | 16 | 2 | 3-isopropylmalate dehydrogenase |
| P18843_NADE_ECOLI | 0 | -1.3 | 23 | 2 | NH(3)-dependent NAD(+) synthetase |
| P36938_PGM_ECOLI | 0 | -1.3 | 24 | 2 | Phosphoglucomutase |
| P0A8G6_WRBA_ECOLI | 0 | -1.3 | 46 | 2 | Flavoprotein wrbA |
| P0ADS9_YGGN_ECOLI | 0 | -1.3 | 8 | 2 | Uncharacterized protein yggN. |
| P76142_YNEA_ECOLI | 0 | -1.3 | 45 | 2 | Uncharacterized protein yneA |
| P0A9Q1_ARCA_ECOLI | 0 | -1.3 | 4 | 1 | Aerobic respiration control protein arcA |
| P0ABJ1_CYOA_ECOLI | 0 | -1.3 | 7 | 1 | Ubiquinol oxidase subunit 2 |
| P0ABP8_DEOD_ECOLI | 0 | -1.3 | 6 | 1 | Purine nucleoside phosphorylase deoD-type |
| P0A9S5_GLDA_ECOLI | 0 | -1.3 | 3 | 1 | Glycerol dehydrogenase |
| P0A6W5_GREA_ECOLI | 0 | -1.3 | 4 | 1 | Transcription elongation factor greA |
| P27298_OPDA_ECOLI | 0 | -1.3 | 9 | 1 | Oligopeptidase A |
| P08178_PUR5_ECOLI | 0 | -1.3 | 11 | 1 | Phosphoribosylformylglycinamidine cyclo-ligase |
| P60651_SPEB_ECOLI | 0 | -1.3 | 3 | 1 | Agmatinase |
| P11875_SYR_ECOLI | 0 | -1.3 | 4 | 1 | Arginyl-tRNA synthetase |
| P64506_YEBY_ECOLI | 0 | -1.3 | 4 | 1 | Uncharacterized protein yebY |
| P08660_AK3_ECOLI | 0 | -1.4 | 22 | 2 | Lysine-sensitive aspartokinase 3 |
| P0A6K6_DEOB_ECOLI | 0 | -1.4 | 14 | 2 | Phosphopentomutase |
| P0A953_FABB_ECOLI | 0 | -1.4 | 38 | 2 | 3-oxoacyl-[acyl-carrier-protein] synthase 1 |
| P09832_GLTD_ECOLI | 0 | -1.4 | 17 | 2 | Glutamate synthase [NADPH] small chain |
| P0ACE7_HINT_ECOLI | 0 | -1.4 | 11 | 2 | HIT-like protein hinT. |
| P06986_HIS8_ECOLI | 0 | -1.4 | 12 | 2 | Histidinol-phosphate aminotransferase |
| P0AAN9_IRAP_ECOLI | 0 | -1.4 | 9 | 2 | Protein iraP. |
| P06959_ODP2_ECOLI | 0 | -1.4 | 14 | 2 | Dihydrolipoyllysine-residue acetyltransferase component of p |
| P08839_PT1_ECOLI | 0 | -1.4 | 83 | 2 | Phosphoenolpyruvate-protein phosphotransferase |

| | | | | | |
|---|---|---|---|---|---|
| P69783_PTGA_ECOLI | 0 | -1.4 | 79 | 2 | Glucose-specific phosphotransferase enzyme IIA component |
| P0A7D4_PURA_ECOLI | 0 | -1.4 | 44 | 2 | Adenylosuccinate synthetase |
| P33221_PURT_ECOLI | 0 | -1.4 | 24 | 2 | Phosphoribosylglycinamide formyltransferase 2 |
| P0A7V3_RS3_ECOLI | 0 | -1.4 | 18 | 2 | 30S ribosomal protein S3. |
| P07395_SYFB_ECOLI | 0 | -1.4 | 19 | 2 | Phenylalanyl-tRNA synthetase beta chain |
| P0A850_TIG_ECOLI | 0 | -1.4 | 80 | 2 | Trigger factor |
| P37685_ALDB_ECOLI | 0 | -1.4 | 3 | 1 | Aldehyde dehydrogenase B |
| P0AAI5_FABF_ECOLI | 0 | -1.4 | 12 | 1 | 3-oxoacyl-[acyl-carrier-protein] synthase 2 |
| P0AC62_GLRX3_ECOLI | 0 | -1.4 | 9 | 1 | Glutaredoxin-3 |
| P0A9D2_GST_ECOLI | 0 | -1.4 | 10 | 1 | Glutathione S-transferase |
| P04391_OTC1_ECOLI | 0 | -1.4 | 8 | 1 | Ornithine carbamoyltransferase chain I |
| P37095_PEPB_ECOLI | 0 | -1.4 | 5 | 1 | Peptidase B |
| P15640_PUR2_ECOLI | 0 | -1.4 | 19 | 1 | Phosphoribosylamine--glycine ligase |
| P30177_YBIB_ECOLI | 0 | -1.4 | 8 | 1 | Uncharacterized protein ybiB. |
| P0ABH7_CISY_ECOLI | 0 | -1.5 | 20 | 2 | Citrate synthase |
| P0AC59_GLRX2_ECOLI | 0 | -1.5 | 34 | 2 | Glutaredoxin-2 |
| P62399_RL5_ECOLI | 0 | -1.5 | 38 | 2 | 50S ribosomal protein L5. |
| P16659_SYP_ECOLI | 0 | -1.5 | 11 | 2 | Prolyl-tRNA synthetase |
| P12758_UDP_ECOLI | 0 | -1.5 | 28 | 2 | Uridine phosphorylase |
| P0AED0_USPA_ECOLI | 0 | -1.5 | 42 | 2 | Universal stress protein A. |
| P0A8W8_YFBU_ECOLI | 0 | -1.5 | 12 | 2 | UPF0304 protein yfbU. |
| P0A6N8_EFPL_ECOLI | 0 | -1.5 | 7 | 1 | Elongation factor P-like protein. |
| P0AAB6_GALF_ECOLI | 0 | -1.5 | 6 | 1 | UTP--glucose-1-phosphate uridylyltransferase |
| P31057_PANB_ECOLI | 0 | -1.5 | 3 | 1 | 3-methyl-2-oxobutanoate hydroxymethyltransferase |
| P05055_PNP_ECOLI | 0 | -1.5 | 10 | 1 | Polyribonucleotide nucleotidyltransferase |
| P18335_ARGD_ECOLI | 0 | -1.6 | 10 | 2 | Acetylornithine/succinyldiaminopimelate aminotransferase |
| P00968_CARB_ECOLI | 0 | -1.6 | 29 | 2 | Carbamoyl-phosphate synthase large chain |
| P17169_GLMS_ECOLI | 0 | -1.6 | 16 | 2 | Glucosamine--fructose-6-phosphate aminotransferase [isomeriz |
| P76558_MAO2_ECOLI | 0 | -1.6 | 15 | 2 | NADP-dependent malic enzyme |
| P0A749_MURA_ECOLI | 0 | -1.6 | 9 | 2 | UDP-N-acetylglucosamine 1-carboxyvinyltransferase |
| P0A9M8_PTA_ECOLI | 0 | -1.6 | 18 | 2 | Phosphate acetyltransferase |
| P04693_TYRB_ECOLI | 0 | -1.6 | 9 | 2 | Aromatic-amino-acid aminotransferase |
| P0ADE8_YGFZ_ECOLI | 0 | -1.6 | 18 | 2 | tRNA-modifying protein ygfZ. |

| | | | | | |
|---|---|---|---|---|---|
| P76223_YNJB_ECOLI | 0 | -1.6 | 28 | 2 | Protein ynjB. |
| P76344_YODA_ECOLI | 0 | -1.6 | 285 | 2 | Metal-binding protein yodA |
| P0A6L2_DAPA_ECOLI | 0 | -1.6 | 8 | 1 | Dihydrodipicolinate synthase |
| P0AES4_GYRA_ECOLI | 0 | -1.6 | 3 | 1 | DNA gyrase subunit A |
| P40874_MTOX_ECOLI | 0 | -1.6 | 7 | 1 | N-methyl-L-tryptophan oxidase |
| P0A763_NDK_ECOLI | 0 | -1.6 | 12 | 1 | Nucleoside diphosphate kinase |
| P37760_RFBD_ECOLI | 0 | -1.6 | 3 | 1 | dTDP-4-dehydrorhamnose reductase |
| P0AD33_YFCZ_ECOLI | 0 | -1.6 | 3 | 1 | UPF0381 protein yfcZ. |
| P0ABB4_ATPB_ECOLI | 0 | -1.7 | 26 | 2 | ATP synthase subunit beta |
| P00864_CAPP_ECOLI | 0 | -1.7 | 54 | 2 | Phosphoenolpyruvate carboxylase |
| P0A6Y8_DNAK_ECOLI | 0 | -1.7 | 77 | 2 | Chaperone protein dnaK |
| P0AEP3_GALU_ECOLI | 0 | -1.7 | 15 | 2 | UTP--glucose-1-phosphate uridylyltransferase |
| P0A717_KPRS_ECOLI | 0 | -1.7 | 9 | 2 | Ribose-phosphate pyrophosphokinase |
| P0AFG8_ODP1_ECOLI | 0 | -1.7 | 43 | 2 | Pyruvate dehydrogenase E1 component |
| P0AG67_RS1_ECOLI | 0 | -1.7 | 34 | 2 | 30S ribosomal protein S1. |
| P07639_AROB_ECOLI | 0 | -1.7 | 4 | 1 | 3-dehydroquinate synthase |
| P0AES0_GSP_ECOLI | 0 | -1.7 | 6 | 1 | Bifunctional glutathionylspermidine synthetase/amidase |
| P00962_SYQ_ECOLI | 0 | -1.7 | 7 | 1 | Glutaminyl-tRNA synthetase |
| P0A6P1_EFTS_ECOLI | 0 | -1.8 | 68 | 2 | Elongation factor Ts |
| P0A9A6_FTSZ_ECOLI | 0 | -1.8 | 12 | 2 | Cell division protein ftsZ. |
| P37689_GPMI_ECOLI | 0 | -1.8 | 53 | 2 | 2,3-bisphosphoglycerate-independent phosphoglycerate mutase |
| P00909_TRPC_ECOLI | 0 | -1.8 | 16 | 2 | Tryptophan biosynthesis protein trpCF |
| P0A6E9_BIOD2_ECOLI | 0 | -1.8 | 4 | 1 | Putative dethiobiotin synthetase |
| P0A6F1_CARA_ECOLI | 0 | -1.8 | 9 | 1 | Carbamoyl-phosphate synthase small chain |
| P0ABU5_ELBB_ECOLI | 0 | -1.8 | 4 | 1 | Enhancing lycopene biosynthesis protein 2 |
| P08142_ILVB_ECOLI | 0 | -1.8 | 9 | 1 | Acetolactate synthase isozyme 1 large subunit |
| P52643_LDHD_ECOLI | 0 | -1.8 | 3 | 1 | D-lactate dehydrogenase |
| P09424_MTLD_ECOLI | 0 | -1.8 | 8 | 1 | Mannitol-1-phosphate 5-dehydrogenase |
| P30850_RNB_ECOLI | 0 | -1.8 | 3 | 1 | Exoribonuclease 2 |
| P0A879_TRPB_ECOLI | 0 | -1.8 | 14 | 1 | Tryptophan synthase beta chain |
| P0A8A0_YEBC_ECOLI | 0 | -1.8 | 5 | 1 | UPF0082 protein yebC. |
| P45578_LUXS_ECOLI | 0 | -1.9 | 16 | 2 | S-ribosylhomocysteine lyase |
| P05020_PYRC_ECOLI | 0 | -1.9 | 22 | 2 | Dihydroorotase |

| | | | | | |
|---|---|---|---|---|---|
| P07012_RF2_ECOLI | 0 | -1.9 | 10 | 2 | Peptide chain release factor 2 |
| P00960_SYGA_ECOLI | 0 | -1.9 | 14 | 2 | Glycyl-tRNA synthetase alpha subunit |
| Q46829_BGLA_ECOLI | 0 | -1.9 | 3 | 1 | 6-phospho-beta-glucosidase bglA |
| P25524_CODA_ECOLI | 0 | -1.9 | 4 | 1 | Cytosine deaminase |
| P0AD57_ISPB_ECOLI | 0 | -1.9 | 4 | 1 | Octaprenyl-diphosphate synthase |
| P0A7C6_PEPE_ECOLI | 0 | -1.9 | 6 | 1 | Peptidase E |
| P0A7F6_SPED_ECOLI | 0 | -1.9 | 3 | 1 | S-adenosylmethionine decarboxylase proenzyme |
| P29217_YCEH_ECOLI | 0 | -1.9 | 4 | 1 | UPF0502 protein yceH |
| P0A6E4_ASSY_ECOLI | 0 | -2 | 28 | 2 | Argininosuccinate synthase |
| P09831_GLTB_ECOLI | 0 | -2 | 35 | 2 | Glutamate synthase [NADPH] large chain |
| P23857_PSPE_ECOLI | 0 | -2 | 45 | 2 | Phage shock protein E |
| P07650_TYPH_ECOLI | 0 | -2 | 11 | 2 | Thymidine phosphorylase |
| P16703_CYSM_ECOLI | 0 | -2 | 5 | 1 | Cysteine synthase B |
| P0A6B7_ISCS_ECOLI | 0 | -2 | 10 | 1 | Cysteine desulfurase |
| P0A729_YCEF_ECOLI | 0 | -2 | 3 | 1 | Maf-like protein yceF. |
| P0A8T7_RPOC_ECOLI | 0 | -2.1 | 23 | 2 | DNA-directed RNA polymerase subunit beta' |
| P0AAC0_USPE_ECOLI | 0 | -2.1 | 16 | 2 | Universal stress protein E. |
| P0A6A8_ACP_ECOLI | 0 | -2.1 | 7 | 1 | Acyl carrier protein |
| P08192_FOLC_ECOLI | 0 | -2.1 | 3 | 1 | Bifunctional protein folC |
| P0A6T3_GAL1_ECOLI | 0 | -2.1 | 9 | 1 | Galactokinase |
| P00547_KHSE_ECOLI | 0 | -2.1 | 6 | 1 | Homoserine kinase |
| P07623_META_ECOLI | 0 | -2.1 | 4 | 1 | Homoserine O-succinyltransferase |
| P0AEZ3_MIND_ECOLI | 0 | -2.1 | 3 | 1 | Septum site-determining protein minD |
| P36683_ACON2_ECOLI | 0 | -2.2 | 56 | 2 | Aconitate hydratase 2 |
| P0ABB0_ATPA_ECOLI | 0 | -2.2 | 13 | 2 | ATP synthase subunit alpha |
| P0A6M8_EFG_ECOLI | 0 | -2.2 | 191 | 2 | Elongation factor G |
| P37192_GATY_ECOLI | 0 | -2.2 | 23 | 2 | Tagatose-1,6-bisphosphate aldolase gatY |
| P0AFF6_NUSA_ECOLI | 0 | -2.2 | 16 | 2 | Transcription elongation protein nusA |
| P0A7D7_PUR7_ECOLI | 0 | -2.2 | 35 | 2 | Phosphoribosylaminoimidazole-succinocarboxamide synthase |
| P37759_RFBB_ECOLI | 0 | -2.2 | 11 | 2 | dTDP-glucose 4,6-dehydratase |
| P11447_ARLY_ECOLI | 0 | -2.2 | 6 | 1 | Argininosuccinate lyase |
| P06983_HEM3_ECOLI | 0 | -2.2 | 9 | 1 | Porphobilinogen deaminase |
| P16456_SELD_ECOLI | 0 | -2.2 | 9 | 1 | Selenide, water dikinase |

| | | | | | |
|---|---|---|---|---|---|
| P0ACY3_YEAG_ECOLI | 0 | -2.2 | 4 | 1 | Uncharacterized protein yeaG. |
| P0ABF6_CDD_ECOLI | 0 | -2.3 | 12 | 2 | Cytidine deaminase |
| P13033_GLPB_ECOLI | 0 | -2.3 | 6 | 1 | Anaerobic glycerol-3-phosphate dehydrogenase subunit B |
| P0A6C8_ARGB_ECOLI | 0 | -2.4 | 12 | 2 | Acetylglutamate kinase |
| P0A6N1_EFTU_ECOLI | 0 | -2.4 | 250 | 2 | Elongation factor Tu |
| P21599_KPYK2_ECOLI | 0 | -2.4 | 33 | 2 | Pyruvate kinase II |
| P22259_PPCK_ECOLI | 0 | -2.4 | 30 | 2 | Phosphoenolpyruvate carboxykinase [ATP] |
| P0A8V2_RPOB_ECOLI | 0 | -2.4 | 26 | 2 | DNA-directed RNA polymerase subunit beta |
| P0AC33_FUMA_ECOLI | 0 | -2.4 | 5 | 1 | Fumarate hydratase class I, aerobic |
| P23893_GSA_ECOLI | 0 | -2.4 | 3 | 1 | Glutamate-1-semialdehyde 2,1-aminomutase |
| P06999_K6PF2_ECOLI | 0 | -2.4 | 6 | 1 | 6-phosphofructokinase isozyme 2 |
| P09151_LEU1_ECOLI | 0 | -2.4 | 6 | 1 | 2-isopropylmalate synthase |
| P0A993_F16P_ECOLI | 0 | -2.5 | 9 | 2 | Fructose-1,6-bisphosphatase |
| P0A6F3_GLPK_ECOLI | 0 | -2.5 | 57 | 2 | Glycerol kinase |
| P00934_THRC_ECOLI | 0 | -2.5 | 71 | 2 | Threonine synthase |
| P23908_ARGE_ECOLI | 0 | -2.5 | 4 | 1 | Acetylornithine deacetylase |
| P15639_PUR9_ECOLI | 0 | -2.5 | 4 | 1 | Bifunctional purine biosynthesis protein purH |
| P25906_YDBC_ECOLI | 0 | -2.5 | 3 | 1 | Putative oxidoreductase ydbC |
| P76536_YFEX_ECOLI | 0 | -2.5 | 10 | 1 | Uncharacterized protein yfeX. |
| P05791_ILVD_ECOLI | 0 | -2.6 | 10 | 2 | Dihydroxy-acid dehydratase |
| P30126_LEUD_ECOLI | 0 | -2.6 | 20 | 2 | 3-isopropylmalate dehydratase small subunit |
| P0AEZ1_METF_ECOLI | 0 | -2.6 | 19 | 2 | 5,10-methylenetetrahydrofolate reductase |
| P78067_YNJE_ECOLI | 0 | -2.6 | 64 | 2 | Putative thiosulfate sulfurtransferase ynjE |
| P22106_ASNB_ECOLI | 0 | -2.6 | 4 | 1 | Asparagine synthetase B [glutamine-hydrolyzing] |
| P24186_FOLD_ECOLI | 0 | -2.7 | 5 | 1 | Bifunctional protein folD |
| P37747_GLF_ECOLI | 0 | -2.7 | 5 | 1 | UDP-galactopyranose mutase |
| P69828_PTKA_ECOLI | 0 | -2.8 | 11 | 2 | Galactitol-specific phosphotransferase enzyme IIA component |
| P60723_RL4_ECOLI | 0 | -2.9 | 12 | 1 | 50S ribosomal protein L4. |
| P25516_ACON1_ECOLI | 0 | -3 | 29 | 2 | Aconitate hydratase 1 |
| P00561_AK1H_ECOLI | 0 | -3 | 29 | 2 | Bifunctional aspartokinase/homoserine dehydrogenase I |
| P0A915_OMPW_ECOLI | 0 | -3 | 19 | 1 | Outer membrane protein W |
| P37188_PTKB_ECOLI | 0 | -3 | 3 | 1 | Galactitol-specific phosphotransferase enzyme IIB component |
| P13029_CATA_ECOLI | 0 | -3.1 | 22 | 1 | Peroxidase/catalase HPI |

| | | | | | |
|---|---|---|---|---|---|
| Q46857_DKGA_ECOLI | 0 | -3.1 | 9 | 1 | 2,5-diketo-D-gluconic acid reductase A |
| P0A9T4_TAS_ECOLI | 0 | -3.1 | 3 | 1 | Protein tas. |
| P0A9W3_YJJK_ECOLI | 0 | -3.1 | 13 | 1 | Uncharacterized ABC transporter ATP-binding protein yjjK. |
| P0AGD3_SODF_ECOLI | 0 | -3.3 | 13 | 1 | Superoxide dismutase [Fe] |
| P27550_ACSA_ECOLI | 0 | -3.4 | 51 | 2 | Acetyl-coenzyme A synthetase |
| P37191_GATZ_ECOLI | 0 | -3.4 | 25 | 2 | Putative tagatose 6-phosphate kinase gatZ |
| P0A6A6_LEU2_ECOLI | 0 | -3.5 | 38 | 2 | 3-isopropylmalate dehydratase large subunit |
| Q59385_ATCU_ECOLI | 0 | -3.6 | 23 | 1 | Copper-transporting P-type ATPase |
| P21888_SYC_ECOLI | 0 | -3.7 | 11 | 2 | Cysteinyl-tRNA synthetase |
| P30178_YBIC_ECOLI | 0 | -3.8 | 12 | 1 | Uncharacterized oxidoreductase ybiC |
| P0A6H5_HSLU_ECOLI | 0 | -4 | 7 | 1 | ATP-dependent hsl protease ATP-binding subunit hslU |
| P32132_TYPA_ECOLI | 0 | -4.9 | 15 | 2 | GTP-binding protein typA/BipA |
| P77739_YNIA_ECOLI | 0 | -7 | 6 | 1 | Uncharacterized protein yniA. |

# Vita

William "Judson" Hervey, IV, was born on January 18, 1979 in Wheeling, the first capital of West Virginia. From childhood through adolescence, Judson was raised in the small town of Wellsburg, WV. He frequently spent time with his father on the family farm, which fostered an interest in living things from a young age. Judson studied in the Brooke County public school system, attending Wellsburg Primary West, Wellsburg Middle School, and Brooke High School. When not engaged in his studies, Judson worked part-time at the historic Drover's Inn and Tavern in Wellsburg from May 1996 to June 2002.

In 1997, Judson graduated from Brooke High and began studying at West Liberty State College (now West Liberty University). While at West Liberty, he participated in a research project propagating an heirloom tomato seed repository under Dr. Roger G. Seeber and Dr. Robert Kreisberg. Work on this project led to an interest in research. Not soon thereafter, Judson changed his major from a pre-professional track to Biology.

After graduation from West Liberty in May 2002, he accepted an offer to study in the multidisciplinary Graduate School of Genome Science and Technology, a joint program between the University of Tennessee-Knoxville and Oak Ridge National Laboratory. While completing his Doctorate of Philosophy degree in 2009, Judson received a National Research Council Research Associateship from the National Academies of Science. He plans to conduct post-doctoral research in proteomics at the Naval Research Laboratory in Washington, D. C. In any free time, Judson plans to resume his study of music and the bass guitar.