

UCLA

UCLA Electronic Theses and Dissertations

Title

High throughput analysis of proteome perturbations induced by radiation, radiomitigators and chemotherapeutics

Permalink

<https://escholarship.org/uc/item/00r178f9>

Author

Erde, Jonathan

Publication Date

2012

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

High throughput analysis of proteome perturbations
induced by radiation, radiomitigators and chemotherapeutics

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

by

Jonathan Erde

2012

ABSTRACT OF THE DISSERTATION

High throughput analysis of proteome perturbations
induced by radiation, radiomitigators and chemotherapeutics

by

Jonathan Erde

Doctor of Philosophy in Biochemistry and Molecular Biology

University of California, Los Angeles, 2012

Professor Joseph Loo, Chair

Proteomics has flourished as a result of major advances in computational power, data storage technologies, and network architecture, all of which enable the acquisition, handling, and processing of enormous data sets. As mass spectrometers continue to break resolution and sensitivity barriers, the limitations of qualitative and quantitative proteomic experiments will no longer result from instrumentation constraints, but rather from the quality and availability of robust sample preparation methods and post-processing algorithms. My dissertation focuses on the development of technologies in experimental proteomic workflows that precede and that follow mass spectrometric analysis - areas that are integral to the future of proteomics.

Enhanced Filter Aided Sample Preparation (eFASP) incorporates Tween-20 for plastic passivation, and n-octylglucoside and deoxycholic acid for digestion enhancement. In *E. coli*, eFASP reduces sample losses associated with normal FASP

by almost 300%, and significantly increases hydrophobic protein representation when compared to normal FASP and to a routine precipitation-based sample preparation.

eFASP is validated in the analysis of the ultra-stable proteinaceous sheath of *Methanospirillum hungatei*. eFASP and label-free quantitative analysis revealed a 40.7 kDa polypeptide (Q2FRN9) encoded by Mhun_2271 as the predominant sheath component by 10-fold.

eFASP is combined with sample fractionation methods and software tools for the quantitative analysis of bacterial and human samples. Hydrophilic interaction chromatography (HILIC) enables small-scale fractionation for less complex samples: *Syntrophus aciditrophicus*, *Anaerobaculum hydrogeniformans*, and melanoma-derived exosomes. *S. aciditrophicus* and *A. hydrogeniformans* are profiled for proteome shifts resulting changes in carbon source. Treatment of melanoma cells with B-Raf inhibitor, PLX-4720, results in the packaging of two tyrosine kinases, VEGFR and PDGFR- β , into secreted exosomes. Upregulation of these receptor tyrosine kinases has been found to be a result of the release of feedback inhibition following treatment with MAPK targeted therapeutics.

Complex whole-cell proteome analysis is effected by high-resolution electrostatic-repulsion hydrophilic interaction chromatography (ERLIC) fractionation of eFASP-processed samples. eFASP-ERLIC facilitated the quantitative analysis of normal and ATM-deficient human lymphoblastoid cell lines (LCLs) exposed to ionizing radiation (IR) and to YEL2, a potent radiomitigator. Exposure to IR resulted in the durable upregulation of NRF2-oxidative stress response components in Normal LCLs, while thrombin signaling and pro-survival proteins were upregulated in ATM-deficient

LCLs. YEL2 treatment upregulated DNA repair components and enriched pro-survival pathways, which implicate the involvement of multiple processes in YEL2-mediated radioresistance.

The dissertation of Jonathan Erde is approved.

Robert Clubb

Thomas Graeber

Catherine Clarke

Jing Huang

Joseph A. Loo, Committee Chair

University of California, Los Angeles

2012

TABLE OF CONTENTS

PREFACE	vii
VITA	.xi
CHAPTER 1: Enhanced FASP (eFASP) incorporates surfactants and passivation into FASP to increase proteome coverage and sample recovery for quantitative proteomic experiments.....	1
References	51
CHAPTER 2: Data Extraction and Correction (DECO): software for the accurate merging of fractionated LC-MSE data, and the extraction and quantification of proteins identified in the ProteinLynx Global Server pipeline	54
References	67
CHAPTER 3: Quantitative proteomic characterization of the para-crystalline sheath encapsulating <i>Methanospirillum hungatei</i> strain JF1: identity of the major structural components.....	68
References	92
CHAPTER 4: Comprehensive proteome characterization of the newly discovered Anaerobaculum hydrogeniformans strain OS-1, and the modulation of catabolic pathways to efficiently grow with glucose, serine, or histidine	95
References	131
CHAPTER 5: <i>Syntrophus aciditrophicus</i> adapts its metabolic machinery for growth with benzoate, crotonate, or cyclohexane carboxylate	135
References	154
CHAPTER 6: Differential exosome packaging by melanoma cells treated with Plexxikon (PLX-4720): probing a potentially rich source of protein and nucleic acid biomarkers for diagnostics and therapeutic response monitoring	156
References	176
CHAPTER 7: The early and latent effects of ionizing radiation on lymphoblastoid cells with functional and non-functional ATM	179
References	212
CHAPTER 8: The immune cell responses to the radiomitigator drug, YEL2 and to ATM-deficiency.....	216
References	256

PREFACE

Proteomics has flourished as a result of major advances in computational power, data storage technologies, and network architecture, all of which enable the acquisition, handling, and processing of enormous data sets. As mass spectrometers continue to break resolution and sensitivity barriers, the limitations of qualitative and quantitative proteomic experiments will no longer result from instrumentation constraints, but rather from the quality and availability of robust sample preparation methods and post-processing algorithms. My dissertation focuses on the development of technologies in experimental proteomic workflows that precede and that follow mass spectrometric analysis - areas that are integral to the future of proteomics.

In Chapter 1 and Chapter 2, I present a new sample preparation method that increases proteome coverage, and then validate its effectiveness using a notoriously difficult sample. Chapter 1 describes the current state of proteomic sample preparation and the recently proposed Filter Aided Sample Preparation (FASP) method that revolutionizes the concept of sample preparation. I present my improvements to the FASP method, collectively referred to as enhanced FASP (eFASP), which include plastic passivation, digestion-enhancing surfactants, and alternative reduction and alkylation reagents. The enhanced features of eFASP increase proteome coverage and membrane protein representation, and quantitatively increase sample recovery. In Chapter 2, the effectiveness of eFASP is illustrated by the successful quantitative and qualitative MS analysis of the para-crystalline proteinaceous sheath of *Methanospirillum hungatei* strain JF1. MS analysis of the sheath components is a significant

achievement due to its extreme stability, which has hindered composition analysis since the 1970s. While eFASP is useful for the comprehensive characterization of difficult samples, its application to comparative label-free quantitative proteomics is the focus of the remainder of my dissertation.

Quantitative characterization of protein changes resulting from exposure to stimuli can benefit many areas of biological research and drug discovery. The application of eFASP, and the development of ancillary software and fractionation tools, for the quantitative proteomic analysis of bacterial and mammalian samples are described in the remaining chapters. Chapter 3 demonstrates one of the major stumbling blocks to the accurate quantitative analysis of samples that require fractionation before LC-MS analysis. I created a software strategy for overcoming this challenge and programmed its implementation in the software suite Data Extraction and Correction (DECO). DECO maintains additional tools aimed at increasing researcher access to experimental data and batch processing that are imperative for large-scale proteomic experiments.

In Chapter 4 and Chapter 5, hydrophilic interaction chromatography (HILIC) is used to fractionate eFASP-produced peptides to reduce sample complexity for the proteomic investigation of two anaerobic bacteria, *Syntrophus aciditrophicus* and *Anaerobaculum hydrogeniformans*, that are grown on a variety of carbon sources. These bacteria are integral to the degradation of hydrocarbons and the global cycling of carbon. *A. hydrogeniformans* is a recently discovered bacterium that produces hydrogen at levels approaching the thermodynamic limit, which makes the elucidation of

pathways important to bioenergetics research and biohydrogen production development.

Chapters 6, 7, and 8 focus on the proteomic effects of chemotherapeutic drugs, radiation exposure, and radiomitigator drugs on human cells. Chapter 6 employs eFASP-HILIC to explore exosome packaging by melanoma cells treated with the B-Raf kinase inhibitor, Plexxikon. Differential proteome signatures in drug-treated tumor-derived exosomes provide a collection of proteins that may be useful for monitoring drug response, for identifying factors involved in metastasis, and for further characterizing Plexxikon pharmacodynamics. Chapter 7 presents the characterization of immune cell response to ionizing radiation. eFASP is utilized in conjunction with extensive fractionation by electrostatic-repulsion hydrophilic interaction chromatography (ERLIC), which enables whole-cell proteome analysis. Detailing cellular responses to radiation highlights radio-responsive pathways integral to radioresistance and damage repair. In Chapter 8, eFASP-ERLIC facilitates the interrogation of immune cells treated with the potent radiomitigator drug, YEL2. In addition to YEL2-induced changes in DNA repair components, pro-survival pathways are also modulated, which implicates the involvement of multiple processes in YEL2-mediated radio-resistance.

Acknowledge the people who have provided samples for analysis. Richard Gatti and Shareef Nahas provided control and experimental lymphoblastoid cell line samples that had been treated with radiomitigators and ionizing radiation. Robert Schiestl and Yelena Rivina for providing the radiomitigator compound, YEL2. Michael McInerney and Matt Maune provided *S. aciditrophicus* samples and *A. hydrogeniformans* samples

grown under a variety of carbon source conditions. Robert Gunsalus and Anne Henstra who provided isolated sheath samples from *M. hungatei*.

This work was funded by the NIH/NAIAD for the UCLA Center for Biological Radioprotectors (U19 AI067769; PI: W. McBride; Co-PI: R. Gatti and J. Loo), the UCLA Cellular and Molecular Biology Training Program (T32 GM007185; to J. Erde), the Department of Energy Biosciences Division grant award DE-FG02-08ER64689 to RPG, and the UCLA-DOE Institute of Genomics and Proteomics to JL and RPG.

VITA

2005	Bachelor of Science, Biochemistry and Molecular Biology Loyola Marymount University
2005 – 2007	Immunosciences Lab., Inc. Beverly Hills, CA
2007 – 2012	Department of Chemistry and Biochemistry University of California, Los Angeles

Publications and Presentations

Journal Articles

Erde J, Loo RO, Loo JA. “Enhanced FASP (eFASP) incorporates surfactants and passivation into FASP to increase proteome coverage and sample recovery for quantitative proteomic experiments.” *J. Proteome Res.* Submitted.

Kani K, Faca VM, Hughes LD, Zhang W, Fang Q, Shahbaba B, Luethy R, Erde J, Schmidt J, Pitteri SJ, Zhang Q, Katz JE, Gross ME, Plevritis SK, McIntosh MW, Jain A, Hanash S, Agus DB, Mallick P. “Quantitative proteomic profiling identifies protein correlates to EGFR kinase inhibition.” *Mol Cancer Ther.* 2012 May;11(5):1071-81.

Vojdani A, Erde J. “Regulatory T cells, a potent immunoregulatory target for CAM researchers: modulating tumor immunity, autoimmunity and alloreactive immunity (III).” *Evid Based Complement Alternat Med.* 2006 Sep;3(3):309-16.

Vojdani A, Erde J. “Regulatory T Cells, a Potent Immunoregulatory Target for CAM Researchers: Modulating Allergic and Infectious Disease Pathology (II).” *Evid Based Complement Alternat Med.* 2006 Jun;3(2):209-15.

Vojdani A, Erde J. “Regulatory T cells, a potent immunoregulatory target for CAM researchers: the ultimate antagonist (I).” *Evid Based Complement Alternat Med.* 2006 Mar;3(1):25-30.

Oral Presentations

“Measuring the Cellular Effects of Ionizing Radiation Exposure and Radiomitigator Treatment using Label-free Quantitative Proteomics.” Presented at the Cellular and Molecular Biology Training Grant 3rd Year Seminar Series, Los Angeles, CA, May 29, 2012.

“Quantifying the Cellular Effects of Ionizing Radiation and Radioprotectors.” Presented at the Cellular and Molecular Biology Training Grant 2nd Year Seminar Series, Los Angeles, CA, January 31, 2011.

“Quantifying the Cellular Effects of Ionizing Radiation and Radiomitigators on Immune Cells.” Presented at the Molecular Biology Institute Lake Arrowhead Retreat, Lake Arrowhead, CA, October 16, 2011.

“Quantifying the Cellular Effects of Ionizing Radiation and Radioprotectors”
Presented at the UCLA Center for Medical Countermeasures against Radiation Monthly
Program Meeting, Los Angeles, CA, July 25, 2011.

“Quantifying the Cellular Effects of Ionizing Radiation.”
Presented at the UCLA Department of Biological Chemistry Annual Retreat, Los Angeles, CA,
April 30, 2010.

Poster Presentations (presenter underlined)

Erde J, Ytterberg AJ, Nahas SA, Rivina Y, Schiestl RH, Gatti RA, Loo JA. “Measuring the Cellular Effects of Ionizing Radiation Exposure and Radioprotector Treatment in Immune Cells using Quantitative Proteomics.” Presented at the 60th ASMS Conference on Mass Spectrometry and Allied Topics, Vancouver, Canada, May 20-24, 2012.

Harris H, ErdeJ, Hughes C, Gethings L, Vissers J, Langridge J, Loo JA. “Getting More from Less: Quantitative Proteomics using LC-IMS-MSE.” Presented at US HUPO 8th Annual Conference, San Francisco, CA, March 4-7, 2012.

Erde J, Ytterberg AJ, Nahas SA, Rivina Y, Schiestl RH, Gatti RA, Loo JA. “Quantifying the effects of ionizing radiation and radioprotectors on immune cells using OFFGeLC-MSE.” Presented at the Human Proteome Organization 10th Annual World Congress, Geneva, Switzerland, September 3-7, 2011.

Erde J, Ytterberg AJ, Nahas SA, Gatti RA, Loo JA. “Quantifying the Effects of Ionizing Radiation and Radioprotectors on Immune Cells using OFFGeLC-MSE.” Presented at the 59th ASMS Conference on Mass Spectrometry and Allied Topics, Denver, CO, June 5-9, 2011.

Loo R, Francoleon D, Erde J, Henstra A, Rohlin L, Gunsalus RP, Daly M, Loo JA. “Identification and Quantification of Archaeon “Ghost” Proteins.” Presented at US HUPO 7th Annual Conference, Raleigh, NC, March 20-23, 2011.

Erde J, Ytterberg AJ, Nahas SA, Gatti RA, Loo JA. “Quantifying the Effects of Ionizing Radiation and Radioprotectors on Lymphoblastoid Cells using OFFGeLC-MS-E.” Presented at the 58th ASMS Conference on Mass Spectrometry and Allied Topics, Salt Lake City, UT, May 23-27, 2010.

Erde J, Ytterberg AJ, Nahas SA, Gatti RA, Loo JA. “The Effects of Ionizing Radiation on Lymphoblastoid Cells.” Presented at the Human Proteome Organization 8th Annual World Congress, Toronto, Canada, September 26-30, 2009

CHAPTER 1

Enhanced FASP (eFASP) incorporates surfactants and passivation into FASP to increase proteome coverage and sample recovery for quantitative proteomic experiments.

ABSTRACT

The integrity of quantitative proteomic experiments depends on the reliability and the robustness of the protein extraction, solubilization, and digestion methods utilized. Combinations of detergents, chaotropes, and mechanical means can achieve successful protein preparation; however, the methods utilized to purify a sample of these contaminants, in addition to the salts, nucleic acids, and lipids already in the sample, can result in significant sample losses and incomplete contaminant removal. A recently introduced method for proteomic sample preparation, Filter Aided Sample Preparation (FASP), cleverly circumvents many of the challenges associated with traditional protein purification methods, but it is associated with significant sample loss.

Presented here is the enhanced FASP (eFASP) approach that incorporates alternative reagents into the traditional FASP protocol, which does not use urea during digestion, in order to improve the sensitivity, recovery and proteomic coverage of processed samples. Utilization of 0.2% deoxycholic acid during eFASP digestion increases tryptic digestion efficiency for both cytosolic and membrane proteins, while obviating the need for additional cleanup steps associated with sodium deoxycholate usage. Passivation of all plastic components of the Microcon Filter Unit, with 5%

Tween-20, reduces peptide loss by 300% by blocking sites to which peptides or proteins can tightly bind. Additionally, an express eFASP method is presented, which utilizes TCEP and 4-VP for in-solution alkylation of proteins, prior to sample application to the Microcon Filter Unit, which increases alkylation specificity and can significantly reduce processing time.

TABLE OF CONTENTS

INTRODUCTION.....	3
Filter aided sample preparation for proteomics.....	3
Quantitative proteomics	6
Increasing proteome coverage with surfactants	11
EXPERIMENTAL.....	13
Processing a standard protein mixture with DTT, TCEP, IAN, and 4-VP.....	15
Enhanced filter aided sample preparation experimental conditions	17
eFASP without Microcon filter unit passivation	17
eFASP with Microcon filter unit passivation	18
Sample preparation for LC-MS ^E analysis	20
Analysis by nanoLC-MS ^E	20
Data processing and protein identification	21
RESULTS AND DISCUSSION	22
DCA is effective in protein extraction and solubilization	22
DCA enhances trypsin digestion without introducing non-volatile salt	24
Solubilization and digestion comparison.....	26
Identification comparison	27
Qualitative comparison.....	27
Quantitative comparison	29
The effects of DCA concentration on digestion efficiency	34
Combination of surfactants	36
Passivation of Plastics	42
Express eFASP	46
CONCLUSION	50
REFERENCES.....	51

INTRODUCTION

At the heart of any proteomic experiment is the extraction and solubilization of proteins from a sample. This is generally achieved with a combination of detergents, chaotropes, and mechanical means. Salts, nucleic acids, lipids and denaturants must be removed for efficient downstream processing and MS analysis. Protein purification is generally accomplished by organic precipitation, affinity chromatography or electrophoresis. Purification methods are subject to individual biases, re-solubilization problems and reduced recovery. A recently introduced method for proteomic sample preparation, Filter Aided Sample Preparation (FASP) ^{1,2}, cleverly circumvents many of the challenges associated with protein purification; however, it is associated with significant sample loss. We present modifications to the FASP workflow that quantitatively improve sensitivity, recovery and proteomic coverage.

Filter aided sample preparation for proteomics

The usage of denaturants in sample preparation is imperative for efficient extraction and solubilization of proteins from a sample matrix. Anionic detergents (lauryl sulfate and lauryl sarcosine) and chaotropes (urea and thiourea) are common components of lysis buffers which function differently, but achieve the same goal. While playing an essential role in sample preparation these, denaturants can have deleterious effects on downstream processing and analysis.

Separation of denaturants from a solution can be achieved in many ways. One of the most common methods is the selective precipitation of proteins with large volumes of an organic, usually acetone, in the presence of reducing agents, acids, or

carrier compounds. Anionic detergents can be removed by precipitation at low pH, or by selective binding to materials specific to that detergent. Dialysis can remove detergents and salts from a sample. Each purification method has its own problems associated with sample loss and incomplete interference removal.

Manza et al. presented a novel method for sample purification and preparation in a spin filter unit, enabling the removal of SDS prior to enzymolysis ¹. SDS and contaminants (salts, nucleic acids, and lipids) are removed by buffer exchange, and then the proteins remaining in the spin filter can be reduced, alkylated and digested. Resulting peptides are liberated by centrifugation, and may be directly analyzed by LC-MS. While conceptually promising, the method is subject to high sample loss and incomplete SDS ². In 2009, Wisniewski et al. described a modification of this method, Filter Aided Sample Preparation (FASP), which incorporates urea for the complete removal of SDS ³.

The usage of urea enables two important advances. First, urea dissociates SDS from proteins and reduces the concentration of SDS below its critical micelle concentration, which together allows buffer exchange to completely remove SDS from the sample. Second, proteins remain denatured in the presence of urea, enabling the usage of Microcon Filter Units with pore sizes up to 30 kDa. Using spin filters with larger pore sizes significantly reduces the time required for each buffer exchange step ^{3,4}.

One concern for the future of FASP was the discontinuation of the Millipore Micron Centrifugal Filters. Wisniewski et al. compared the retention of denatured proteins and the FASP performance of a number alternative ultrafiltration units ⁴. Vivapore and

Millipore ultrafiltration units performed similarly for FASP of cell lysates; however, the Millipore units outperformed Vivapore units in FASP tailored for glycoproteins ^{1,2,4}. This study estimated approximate sample loss near 50%, measured by UV absorbance; however, they were able to identify well over 6,500 peptides and 2500 proteins.

Additions to the FASP workflow have been aimed at enhancing performance and enabling selectivity. Coupling FASP to StageTips (Thermo) strong anion exchange (SAX) can desalt the sample, remove residual urea used during digestion, selectively enrich peptide populations, and enable small-scale fractionation. This FASP-SAX was used to comprehensively analyze membrane enriched samples and to globally map phosphorylation sites in *C. elegans* ^{1,5-7}.

Glycopeptide enrichment usually requires lectin binding to a solid support and the use of affinity columns. FASP can be modified to allow glycopeptide enrichment without the need for binding solid supports, using lectins larger than the MW cutoff of a spin filter unit. Lectins will bind glycopeptides and prevent their passage through the filter, while non-glycopeptides will pass through. This strategy was successful for glycopeptide enrichment and comprehensive mapping of the mammalian N-glycoproteome ^{2,4,8}.

The unique ability of filter units to separate differentially sized species is the basis for using multiple enzymes for consecutive digestion of a protein sample, MED-FASP ^{3,9}. This method is based on findings that tryptic digestion alone leaves many larger peptides intact which can be further digested by another enzyme to enhance sequence coverage ^{3,4,10}. In the FASP workflow, larger peptides, and residual protein material, are unable to pass through the filter, and may be digested by a secondary enzyme.

Peptides from each digestion step may be collected as separate fractions, representing orthogonal peptide populations, for analysis. Tryptic digestion of materials remaining in the filter, after Lys-C digestion and peptide isolation, yields a 40% increase in protein and phosphorylation sites identified than using trypsin alone.

Quantitative proteomics

In proteomic experiments, the identification of proteins in an unknown sample is based on the most reasonable collection of proteins that can account for the peptide sequences observed. The abundance of identified proteins is of equal importance, and may be determined based on various metrics associated with constituent peptides. Protein abundance data can be compared across experimental samples to determine biological effects on gene expression and protein degradation. There are two major approaches for peptide and protein quantification: isotope label dependent, and label independent.

Quantification by isotope labeling

Isotope labeling of samples allows for the acquisition, and quantification, of samples simultaneously. Heavy carbon, oxygen, hydrogen, or nitrogen, can be incorporated into proteins or peptides of specified samples to produce peptide populations that can be easily differentiated based on mass differences. Peptide elution and ionization efficiency are not influenced by isotopic composition, so in-spectrum comparisons of ion intensities for the light and heavy forms of a peptide can be used to

measure relative abundance between differentially labeled samples. Isotope labels can be incorporated either metabolically or chemically.

Stable isotope labeling by amino acids in cell culture (SILAC) is a metabolic labeling strategy in which heavy isotopes of arginine and lysine are incorporated into the cellular proteome to distinguish and relatively quantify the light and the heavy forms of a peptide ^{4,11,12}. Cells are cultured in the presence or absence of isotope coded amino acids (¹³C₆-Arg and ¹³C₆-Lys). The heavy amino acids are labeled with ¹³C, instead of the naturally occurring ¹²C, and will yield a 6 Da mass shift compared with the peptide containing the light version of the amino acid. Full incorporation of the heavy amino acids takes about five cell doublings, such that all the naturally occurring ¹²C₆-Arg and ¹²C₆-Lys in the proteome are replaced by ¹³C₆-Arg and ¹³C₆-Lys, respectively. This occurs through cell division, protein synthesis and protein degradation.

Trypsin cleaves at the carboxyl-termini of Arg and Lys residues; therefore, using Arg and Lys together as labeling amino acids assures that all peptides will be labeled. Using ¹³C isotopes of amino acids is preferential to other labeling strategies (e.g. deuterium or ¹⁵N) because ¹³C labeled peptides and ¹²C peptides are chemically identical, will co-elute from the reverse-phase chromatography column and will maintain the same ionization efficiency.

An alternative to metabolic incorporation of isotope labels is chemical tagging. Isobaric Tag for Relative and Absolute Quantitation (iTRAQ) allows for the simultaneous identification and quantification of proteins from up to 8 different sources ¹³. The method utilizes covalent labeling of the N-terminus and sidechain amines of peptides with isotope-coded tags of varying masses. A distinct advantage is that there is no

need for assuring full incorporation in cell culture, allowing for non-cell culture samples to be analyzed and quantified. Other isotope labeling reagents include cysteine specific isotope-coded affinity tags (ICAT) ¹⁴, and deuterated alkylators like acrylamide ¹⁵ and 2-vinylpyridine ¹⁶.

Label-free quantification

Experimental limitations introduced by dependency on isotope labeling can be overcome by label-free methods for protein quantification. Quantification and comparison of samples requires separate LC-MS analysis followed by comparison of peptide features. The most used peptide features for label-free quantification are spectral counts and ion volume.

Spectral counting is one of the earliest methods for label-free peptide quantification. In these experiments, the sum of fragment ion spectra is correlated with peptide abundance ¹⁷. The spectral count values for peptide belonging to a protein, across samples, gives an indication for relative protein abundance. A similar method, exponentially modified protein abundance index (emPAI), utilizes protein coverage as a gauge of abundance for comparative analysis ¹⁸.

The utilization of peptide ion volume, as measured by summed peptide intensity over the course of elution, is a more accurate metric for quantification. However, the prevalent form of data acquisition, data dependent acquisition (DDA), does not preserve complete elution profiles and is insufficient for accurate and robust label-free quantification. During a DDA experiment, a great deal of time is spent acquiring fragment ion data for selected precursor ions, during which time no precursor ion data

can be collected. This results in precursor ion elution profile dead zones, which can encompass regions of maximal peptide elution and hamper the accuracy and completeness of comparative quantification. In order to circumvent this problem, a newer mode of data acquisition is utilized, data independent acquisition (DIA or MS^E).

LC-MS^E is a strategy for the acquisition of precursor and fragment ion data that increases the depth of analysis and enables absolute quantification of proteins from complex peptide mixtures. MS^E is non-selective method that continually collects precursor ion and fragment ion data in parallel by alternating between low and high collision energy ¹⁹. This feature enables the collection of complete peptide ion elution profiles for relative quantification between samples as well as absolute quantification.

The comprehensive nature of LC-MS^E allows for the determination of absolute quantities of proteins in complex samples ^{20,21}. A known quantity of pre-digested protein can be utilized as an internal standard that is spiked into a sample prior to LC-MS^E analysis. The average MS signal for the 3 most intensely ionizing peptides (Hi3) for each protein is compared to the Hi3 of the internal standard, of known concentration, in order to determine absolute quantities. The Hi3 for a protein may also be used for relative protein quantification across samples. This acquisition method is advantageous because it does not require isotope labeling and will afford a comprehensive qualitative and quantitative view of the cellular proteome. Isotope labeled samples may be analyzed by LC-MS^E to enable much deeper proteome coverage, and increased MS data for quantification; however, the analysis will still be subject to the costs and pitfalls of isotope labeling.

Isotope-label versus label-free quantitative analysis

The utilization of isotopes in quantitative proteomic experiments enables multiplexed sample acquisition. Concurrent acquisition reduces experimental duration, which enables the collection of additional technical replicates for better statistical significance. Collecting all isotope versions of a peptide at the same time reduces complications from inter-run instrument variations. This feature also allows for the quantification of peptides with incomplete elution profiles, which is a significant challenge to label-free DDA experiments. While advantageous in some circumstances, isotope labeling strategies are subject to high reagent costs, incomplete labeling, and increased sample complexity.

Label-free quantification strategies aim to circumvent the complications of isotope labeling while providing unique benefits. Label-free quantification involves the independent analysis of samples, which avoids the complexity issues associated with multiplexing. Mixing samples for multiplexed acquisition results in an increased spectral complexity that can hinder peptide identification. Independent analysis of samples avoids this problem, enabling better proteome coverage and more complete peptide information. While inter-run variability can be a potential issue, it is becoming less significant as chromatographic and electrospray ionization technologies become more stable. An additional benefit of independent acquisition, and isotope avoidance, is the reduction of sample handling and of sample source limitations associated with metabolic labeling. While maintaining distinct advantages over isotope labeling, DDA-based label-free quantitative proteomic analysis is challenged by inter-run differences in peak selection and peak elution profiles. The development of MS^E overcomes these

significant challenges by collecting complete elution profiles of all precursor ions, which permits robust qualitative and quantitative proteomic analysis of unknown samples.

Increasing proteome coverage with surfactants

Sodium deoxycholate (NaDOC) is a bile salt and surfactant that aids in the solubilization of lipids in the alimentary canal, and is secreted along with cholate and pancreatic enzymes, e.g., chymotrypsin and trypsin. In 2006 and 2008, two groups proposed the usage of NaDOC in proteomic workflows to enhance trypsin digestion and increase membrane protein representation²²⁻²⁶. Concentrations up to 10% (w/v) NaDOC are tolerable by trypsin, with concentrations beginning as low as 0.01% yielding improvements in digestion efficiency and sequence coverage^{22,25,27}. NaDOC is mass-spectrometry compatible because it can be quantitatively removed by acidification and phase transfer.

Removal of NaDOC is accomplished by precipitation under low pH conditions. Insoluble surfactant is removed by Phase Transfer (PT) to a peptide immiscible solvent phase, ethyl acetate (EA), by thorough phase mixing and centrifugation²². The peptide-rich surfactant-free aqueous layer is then carefully removed. PT avoids peptide losses associated with NaDOC removal by precipitation and centrifugation²⁸.

PT to an EA layer provides an additional benefit, since many common contaminants are soluble in EA, including (SDS n-octylglucoside, NP-40, Triton X-100), as well as unidentified contaminants²⁹. During PT surfactant removal, a variety of these trace contaminants will be removed. While incorporation of NaDOC enhances trypsin digestion, inclusion of non-volatile sodium will interfere with downstream sample

processing, fractionation, and analysis, unless an additional step is added for desalting the sample.

The purpose of this work is to increase peptide recovery from the FASP method, and to increase digestion efficiency to provide quantitatively reproducible sample digests with good coverage of hydrophobic species (Figure 1-1). The goals of this work are threefold: (i) to introduce the usage of removable surfactants to the FASP workflow to enhance trypsin digestion, (ii) to present deoxycholic acid as an alternative surfactant to NaDOC to eliminate desalting steps prior to salt-sensitive downstream processing, (iii) to assess the utility of Microcon Filter Unit passivation for to decrease sample losses during FASP, (iv) and to determine the compatibility of FASP with alternative reduction and alkylation reagents, TCEP and 4-VP, which can significantly reduce sample handling and processing time. Digests from FASP and eFASP performed with various modifications are analyzed by Q-TOF MS operated in data-independent MS^E mode. MS^E data affords identification and quantification information for peptides and proteins in each sample, as well as complete elution profiles for all eluting species. Custom post-processing software calculates the isoelectric point and Grand Average of Hydropathicity index (GRAVY) values for proteins and peptides. The combination of MS^E data with physiochemical information on identified species allows for the comprehensive evaluation and comparison of sample preparation methods.

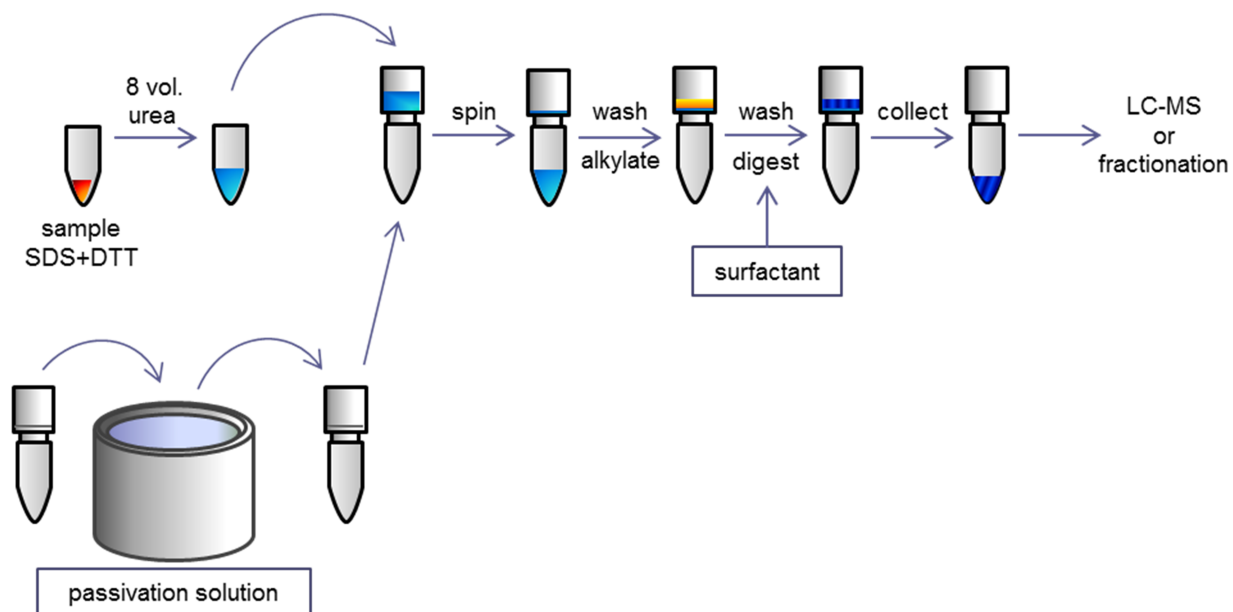


Figure 1-1. Incorporation of enhancements to the FASP workflow. Samples are prepared in 4% SDS, and then diluted with 8M urea to enable the dissociation of SDS from proteins. Microcon Filter Units and collection tubes are passivated by overnight incubation with 5% Tween-20, followed by thorough washing in MS-grade water. The diluted sample is then applied to the passivated Microcon Filter Unit, and through successive buffer exchanges, SDS and contaminants can be removed. The proteins are then alkylated in the presence of urea, which is followed by successive buffer exchanges, to remove alkylation reagents and urea prior to digestion. Proteins are then digested by trypsin in the presence of surfactant, and the resulting peptides can be liberated by centrifugation. Removal of surfactant into an organic layer results in a pure peptide sample that can be analyzed by fractionated or directly analyzed by LC-MS.

EXPERIMENTAL

Extraction and SDS-PAGE with DCA

Lyophilized *Escherichia coli* strain K12 cells were suspended in MS-grade water. Cells were washed twice by pelleting at 1000 g for 2 minutes, removing the supernatant, and re-suspending in water. The suspended cells were transferred to four sample tubes and pelleted. Ice-cold Extraction Buffer A (12 mM deoxycholic acid, 12 mM N-

lauroylsarcosine, 10 mM TCEP, 200 mM ABC), Extraction Buffer B (12 mM deoxycholic acid, 12 mM N-lauroylsarcosine, 10 mM DTT, 200 mM ABC), or Extraction Buffer C (4% SDS, 10 mM DTT, 200 mM ABC) was added to specified sample tubes. Samples were sonicated on ice (0.01-0.03) in 30 second bursts five times, in between sonication steps, and then the samples were centrifuged at 16,000 rcf for 5 minutes. Debris was removed by centrifugation at 2,500 g for 5 minutes. The supernatant of each sample was transferred to a new tube for further processing. One of the samples processed with Extraction Buffer C underwent acetone precipitation and re-suspension in Extraction Buffer C with sonication before further processing.

Each sample (10 µg) was aliquoted into tubes for SDS-PAGE analysis. Samples were boiled for 5 minutes with NuPAGE® LDS Sample Buffer (4X, 10 ml - NP0007) and NuPAGE® Sample Reducing Agent (10X, 250 µl - NP0004), and spun down. The samples were loaded into lanes 2 through 5, and replicates loaded into lanes 7 through 10 of a NuPAGE® Novex 4-12% Bis-Tris Gel (1.0 mm, 12 well, Number, NP0322BOX). Proteins were separated on ice, at 200 V, with 160 mA per gel, for 90 minutes. The gel was removed from the cassette, and the proteins were fixed in 50% methanol / 7% acetic acid for 15 minutes. The gel was washed in water three times, and then incubated with GelCode Blue Reagent (Pierce, 24590) and processed by the manufacturer's instructions. The gel was visualized in a flatbed scanner. For better sensitivity, the gel was then rinsed, and silver stained with SilverQuest Silver Staining Kit (Invitrogen, LC6070) by the manufacturer's instructions.

Processing a standard protein mixture with DTT, TCEP, IAN, and 4-VP

A standard protein mixture was produced with α -casein (bovine, Sigma C-6780), β -casein (bovine, Sigma C-6905), enolase (yeast, Sigma E6126), apo-transferrin (human, Sigma T-4382), carbonic anhydrase (bovine, Sigma C-3934), and ribonuclease B (bovine, Sigma R-7884). Specified quantities of each of the six standard proteins were weighed out, and suspended with 200 mM ABC, to a total concentration of (12.5 mg/ml). Eight aliquots of the standard protein mixture (400 μ g) were alkylated and digested under different conditions.

The first sample was brought to 100 mM ABC, incubated with 22 mM DTT on a shaker at 50°C for 60 minutes, and then alkylated with 30 mM IAN on a shaker at 37°C for 30 minutes. The second sample was brought to 0.1% DCA and 100 mM ABC, incubated with 22 mM DTT on a shaker at 50°C for 60 minutes, and then alkylated with 30 mM IAN on a shaker at 37°C for 30 minutes. The third sample was brought to 0.1% DCA and 100 mM ABC, incubated with 22 mM DTT on a shaker at 50°C for 60 minutes, and then alkylated with 25 mM 4-VP on a shaker at 37°C for 30 minutes. The fourth sample was brought to 100 mM ABC, incubated with 5.5 mM TCEP on a shaker at 50°C for 60 minutes, and then alkylated with 30 mM IAN on a shaker at 37°C for 30 minutes. The fifth sample was brought to 0.1% DCA and 100 mM ABC, incubated with 5.5 mM TCEP on a shaker at 50°C for 60 minutes, and then alkylated with 30 mM IAN on a shaker at 37°C for 30 minutes. The sixth sample was brought to 100 mM ABC, incubated with 5.5 mM TCEP on a shaker at 50°C for 60 minutes, and then alkylated with 25 mM 4-VP on a shaker at 37°C for 30 minutes. The seventh sample was brought to 0.1% DCA and 100 mM ABC, incubated with 5.5 mM TCEP on a shaker at 50°C for

60 minutes, and then alkylated with 25 mM 4-VP on a shaker at 37°C for 30 minutes.

The eighth sample was brought to 0.1% DCA and 100 mM ABC, incubated with 5.5 mM TCEP and 22 mM DTT on a shaker at 50°C for 60 minutes, and then alkylated with 25 mM 4-VP on a shaker at 37°C for 30 minutes.

Alkylation reactions for all samples were quenched with 20 mM DTT. All samples were diluted 1:1 with either 100 mM ABC or 0.1% DCA in 100 mM ABC. 12.5 µg of trypsin (1 µg trypsin:30 µg protein) was added to each sample. Digestion proceeded for 12 hours on a shaker set to 37°C.

Each sample (10 µg) was aliquoted for SDS-PAGE analysis. Samples were boiled for 5 minutes with NuPAGE® LDS Sample Buffer (4X, 10 ml - NP0007) and NuPAGE® Sample Reducing Agent (10X, 250 µl - NP0004), and spun down. The protein ladder, SeeBlue® Plus2 Pre-Stained Standard, was loaded into lane 1 of a NuPAGE® Novex 4-12% Bis-Tris Gel (1.0 mm, 12 well, Number, NP0322BOX). The unprocessed standard protein mixture (10 µg) was loaded into lane 2. The experimental digests of the standard protein mixture (10 µg as calculated from the unprocessed concentration) were loaded into lanes 3 through 10. Proteins were separated on ice, at 200 V, with 160 mA per gel, for 90 minutes. The gel was removed from the cassette, and the proteins were fixed in 50% methanol / 7% acetic acid for 15 minutes. The gel was washed in water three times, and then incubated with GelCode Blue Reagent (Pierce, 24590) and processed by the manufacturer's instructions. The gel was visualized in a flatbed scanner. For better sensitivity, the gel was silver stained with SilverQuest Silver Staining Kit (Invitrogen, LC6070) by the manufacturer's instructions.

Enhanced filter aided sample preparation experimental conditions

eFASP without Microcon filter unit passivation

Aliquots of *E. coli* lysate (162.5 µg) were prepared in Solubilization Buffer A (4% SDS, 0.2% DCA, 7.5 mM TCEP, 200 mM ABC), and incubated on a heated block for 10 minutes at 90°C. The sample was sonicated for three 10-second intervals with a Sonic Dismembrator Model 100 (Fisher Scientific), and then spun at 16,000 rcf for 10 minutes in an Eppendorf centrifuge; this step was repeated once. The supernatant and any residual pellet was sonicated and allowed to cool to 37°C. One aliquot was reserved for processing by 2D Cleanup Kit and in-solution digestion (described below).

Eight volumes of Exchange Buffer A (8 M urea, 100 mM ABC, pH 8) was added to each sample to begin buffer exchange of SDS. Diluted samples were applied to Microcon Filter Units, and spun at 14,000 rcf for 10 minutes. The eluate was discarded, and 200 µl of Exchange Buffer A was added to the filter unit and then spun at 14,000 rcf for 10 minutes. This step was repeated twice.

Proteins were alkylated by the addition of Alkylation Buffer (8 M urea, 50 mM iodoacetamide, 100 mM ABC) to the Microcon Filter Unit, and incubated for 1 hour on a shaker at 37°C. DTT was added to a final concentration of 50 mM to deactivate residual IAN. To remove IAN, DTT and urea, one buffer exchange step was performed with Exchange Buffer A, followed by three buffer exchange steps with FASP Digestion Buffer (50 mM ABC) or eFASP Digestion Buffer (0.1% DCA, 0.1% nOG, 0.2% DCA, or 0.1%DCA and 0.1% nOG in 50 mM ABC) .

FASP or eFASP Digestion Buffer (100 µl) was added to each Microcon Filter Unit, followed by the addition of trypsin (3.25 µg, 1 µg trypsin:50 µg protein). Digestion

proceeded for 12 hours on a shaker set to 37°C. The Microcon Filter Unit was transferred to a new collection tube, and spun at 14,000 rcf for 10 minutes. To assure complete removal of peptides, 50 µl of 50 mM ABC buffer was added to the Microcon Filter Unit and collected by centrifugation, twice.

Sample preparation with the 2D Clean Up Kit

One aliquot of *E. coli* lysate in Solubilization Buffer A was processed with the 2D Clean Up Kit (GE) according to the manufacturers specifications. The precipitated *E. coli* proteins were suspended in 100 mM ABC, incubated with 22 mM DTT on a shaker at 50°C for 60 minutes, and then alkylated with 30 mM IAN on a shaker at 37°C for 30 minutes. The alkylation reaction was quenched with 20 mM DTT, and the sample was diluted to a final concentration 0.1% DCA and 100 mM ABC. Trypsin (3.25 µg, 1 µg trypsin:50 µg protein) was added to each sample, and digestion proceeded for 12 hours on a shaker set to 37°C.

eFASP with Microcon filter unit passivation

Microcon Filter Units (30,000 Da cutoff), and collection tubes were incubated with MS-grade water, 5% Tween-20, or 5% ammonium lauryl sulfate overnight. Filter units and collection tubes were removed and thoroughly rinsed with large volumes of MS-grade water.

Two *E. coli* lysates were prepared, one with Lysis Buffer A (4% SDS, 0.2% nOG, 7.5 mM TCEP, 200 mM ABC) and one with Lysis Buffer B (4% SDS, 50 mM TCEP 0.1% DCA, 0.1% nOG, 200 mM ABC). Aliquots of the *E. coli* lysates (162.5 µg from

Lysis Buffer A, or 50 µg from Lysis Buffer B) were incubated on a heated block for 10 minutes at 90°C, and then sonicated for three 10-second intervals with a Sonic Dismembrator Model 100 (Fisher Scientific), and then spun at 16,000 rcf for 10 minutes in an Eppendorf centrifuge; this step was repeated once. The supernatant and any residual pellet, was sonicated and allowed to cool to 37°C. Proteins were alkylated by the addition of 4-vinylpyridine (4-VP) at 25 mM final concentration, and incubated for 1 hour on a shaker at 37°C. DTT was added to a final concentration of 40 mM to deactivate residual 4-VP.

Eight volumes of Exchange Buffer B (8 M urea, 100 mM Tris-HCL, pH8) with no surfactant, or with surfactant (0.1% nOG, 0.1% DCA, 0.2% DCA, or 0.1%DCA and 0.1% nOG), was added to the digest to begin buffer exchange of SDS. The diluted sample was applied to a Passivated Microcon Filter Unit (described below), and spun at 14,000 rcf for 10 minutes. The eluate was discarded, and 200 µl of Exchange Buffer B was added to the filter unit and then spun at 14,000 rcf for 10 minutes. This step was repeated twice. To remove urea, three buffer exchange steps were performed with FASP Digestion Buffer (50 mM ABC) or eFASP digestion buffer (0.1% nOG, 0.1% DCA, 0.2% DCA, or 0.1%DCA and 0.1% nOG, in 50 mM ABC).

FASP Digestion Buffer or eFASP Digestion Buffer (100 µl) was added to a Microcon Filter Unit, followed by the addition of 3.25 µg of trypsin (1 µg trypsin:50 µg protein). Digestion proceeded for 12 hours on a shaker set to 37°C. The Microcon Filter Unit was transferred to a new Passivated Collection Tube, and spun at 14,000 rcf for 10 minutes. To assure complete removal of peptides, 50 µl of 50 mM ABC buffer was added to the Microcon Filter Unit, and collected by centrifugation, two times.

Sample preparation for LC-MS^E analysis

Removal of DCA or nOG from the digests was accomplished by acidification and phase transfer to an ethyl acetate layer ²². Instead of removing the aqueous layer as in Masuda et al. ²², the organic layer was removed, and then two more rounds of addition and removal of ethyl acetate facilitate the removal of surfactant from the solution, similar to the strategy Yeung et al. employ for the removal of nOG ²⁹.

Collected peptides were dried down in a SpeedVac and re-suspended in 50% methanol, three times. Dried peptides were suspended in modified Mobile Phase A (96.8% H₂O/3% ACN/0.2% formic acid).

Analysis by nanoLC-MS^E

nanoUPLC of peptides was performed on an Waters nanoACQUITY UPLC system equipped with a nanoACQUITY UPLC Symmetry C18 180 µm x 20 mm trap column (Waters, 186003514), and a nanoACQUITY UPLC BEH C18 75 µm x 150 mm reversed phase analytical column (Waters, 186003543). The aqueous mobile phase (Mobile Phase A) contained 0.1% formic acid in water and the organic mobile phase (Mobile Phase B) contained 0.1% formic acid in acetonitrile. 5 µl of the sample was loaded onto the trap with and washed with 99% A and 1% B for 5 minutes, and then injected onto the column with 3% Mobile Phase B. Peptides were eluted from the column with a gradient of 3-40% B over 60 or 90 minutes at 300 nl/min, followed by a 5 minute rinse with 95% B. The column was re-equilibrated to 97% A for 20 minutes.

The lock mass compound, [Glu¹]fibrinopeptide (0.5 pmol/µl, 25% ACN, 0.1% formic acid), and the analyte were delivered at 300 nl/min to the NanoLockSpray dual

electrospray ion source (Waters, UK), which was equipped with a microcapillary for the lockmass compound, and a PicoTip Emitter (new Objective) for the analyte. Two electrosprays are produced that may be selectively sampled by positioning of the baffle.

The ion source was interfaced with a Waters Synapt qTof HDMS or a Waters Xevo qTof MS for mass spectrometric analysis. Acquisition was performed in positive nanoESI mode and the instrument was operated in V-mode with an average resolution of 9500. Accurate mass data for precursor and fragment ions were collected over the 100-2000 m/z range, with the mode of acquisition alternating between low and elevated energy at 1-second intervals, for the duration of analysis. The lock mass channel was sampled every 30 seconds for calibration of MS and MS^E data.

Data processing and protein identification

Raw data files were processed in ProteinLynx Global Server 2.5.2 (PLGS, Waters) to provide an inventory of precursor ions along with their respective fragment ions, to search against the HAMAP *Escherichia coli* database (March 2011, entries added for common contaminants and protein standards) for peptide and protein identifications³⁰. Database search settings included a minimum of 3 peptides per protein identification, two missed cleavages, a 4% false positive rate, fixed carbamidomethyl or S-pyridylethyl modification of cysteine residues, and variable oxidation modification of methionine residues. The Waters Expression Informatics component of the PLGS software was utilized to quantify data^{20,21,31}.

Custom software calculates peptide and protein hydrophobicity based on the hydropathicity of constituent amino acids derived from Kyte and Doolittle³², calculates

protein and peptide isoelectric points based on acidic and basic residues, and extracts protein and peptide level data for comparative analysis.

RESULTS AND DISCUSSION

DCA is effective in protein extraction and solubilization

A variety of surfactants and detergents have been tested, at various concentrations, for effects on trypsin and Lys-C digestion. NaDOC is commonly used as a component of lysis buffers, e.g., RIPA buffer, and more recently it has been found to enhance protein digestion by trypsin and Lys-C ^{22,25}. NaDOC can be quantitatively removed from a digest by acidification and transfer to an ethyl acetate layer.

Masuda et al proposed to replace the standard SDS lysis buffer with one that combined a high concentration of NaDOC with sodium lauroylsarcosine (NaLS). NaLS is a strong anion detergent that is not used commonly, but has been shown to be amenable to removal by acidification and phase transfer ²⁴. The NaDOC/NaLS lysis buffer proved effective in extracting and solubilizing proteins from membrane samples, and, after dilution, useful for enhancing subsequent trypsin digestion of those proteins ²⁴. After removal of NaDOC and NaLS, the sample must be desalted before post-processing and MS analysis.

To evaluate the utility of free deoxycholic acid (DCA), we incorporated it into extraction buffers similar to those used by Masuda et al. ²⁴. Three extraction buffers were used. Two extraction buffers were composed of 12 mM DCA, 12 mM ammonium lauroylsarcosine (AmLS), and either 10 mM TCEP or 10 mM DTT. The Laemmli extraction buffer had 2% SDS with 10 mM DTT. One sample was processed with each

extraction buffer. An additional sample was processed with Laemmli buffer, precipitated with acetone, and then re-suspended in Laemmli buffer. Each sample was run on an SDS-PAGE gel and silver stained (Figure 1-2).

Samples extracted with DCA and AmLS show high extraction and solubilization efficiency across a large MW range, with particularly good solubilization of high MW proteins, over 115 kDa. Laemmli buffered samples fared well, and, surprisingly, acetone precipitation did not have a large effect on the protein yield, although a 1D SDS-PAGE gel only yields information on the more abundant proteins present. Thus the alternative formulations of NaDOC and NaLS are just as powerful for protein extraction and solubilization, and have the added benefit of not adding nonvolatile salts to the solution.

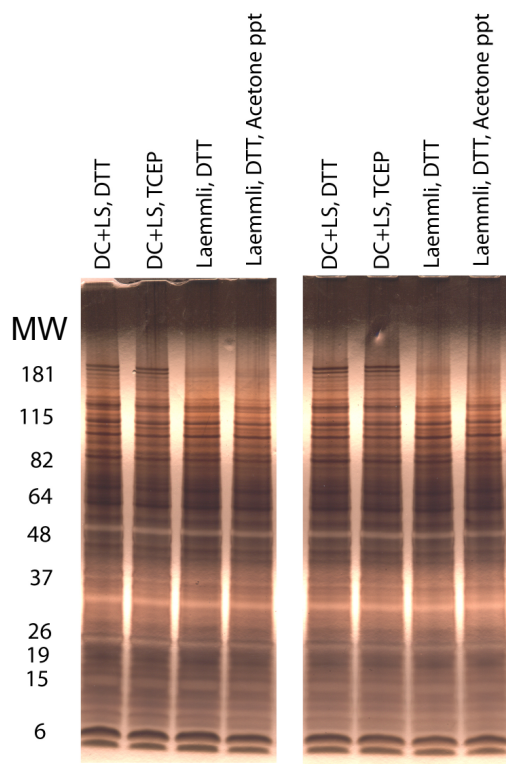


Figure 1-2. Deoxycholic acid (DCA) and ammonium lauryl sarcosine can replace SDS in lysis buffers.

DCA enhances trypsin digestion without introducing non-volatile salt

Digestion of proteins with trypsin is a critical step in proteomic workflows, especially for quantitative proteomic experiments. After proteins are isolated from harsh denaturing agents (SDS, urea, thiourea) by precipitation, dialysis, or affinity chromatography, they must be reduced and alkylated. DTT and TCEP are commonly used to break disulfide bonds in preparation for alkylation. Iodoacetamide is the most widely used alkylator of cysteine residues, with a minority of studies utilizing vinylpyridines (4-VP or 2-VP). Since these reagents react with proteins, including digestion enzymes, they should be removed or diluted prior to enzyme addition.

While NaDOC has been shown effective for trypsin digestion enhancement, there are no data assessing similar efficacy for DCA. In order to test the effect of DCA on trypsin activity, partially diluted alkylated samples were used to emulate adverse conditions. A standard protein mixture was created with 6 proteins with varying molecular weight and cysteine content. The standard protein mixture was suspended in either ABC or 0.1% DCA in ABC, and then processed with different combinations of reducing agent and alkylating agent. All samples were diluted 1:1 before addition of trypsin. Aliquots of each digest were run on an SDS-PAGE gel and proteins were visualized by silver staining (Figure 1-3).

Combinations of reducing agent and alkylating agent are detrimental to trypsin activity, and require adequate dilution, or complete removal, prior to trypsin addition (Figure 1-3). Under these inhibitory conditions, enough trypsin is active to digest a good portion of the proteins present, but a significant amount of protein remains undigested. The presence of 0.1% DCA rescued poor digestion efficiency in samples containing

combinations of reducing agent and alkylating agent. The combination of TCEP with DTT resulted in the poorest digestion efficiency, which is probably a result of strong trypsin inhibition, beyond the point at which DCA enhancement can be useful; however, digestion of carbonic anhydrase, α -casein, and β -casein is still better than in those samples digested in the absence of 0.1% DCA. Therefore, the usage of DCA enables enhanced trypsin activity and does not contribute non-volatile salts to the sample.

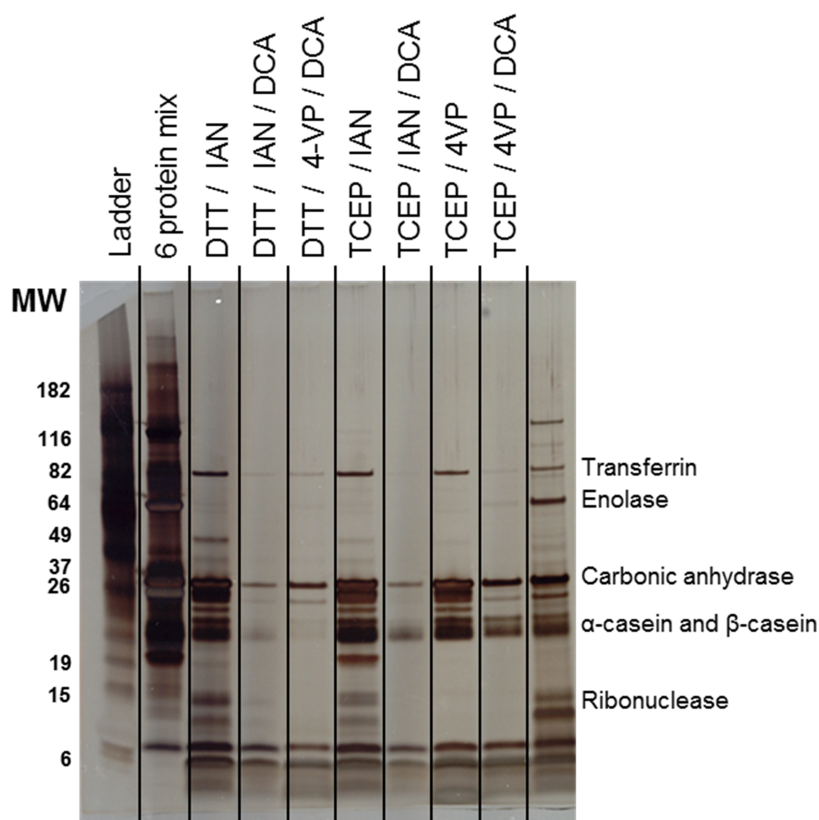


Figure 1-3. A standard protein mixture was reduced and alkylated with combinations of DTT, TCEP, IAN, and 4-VP, prior to trypsin digestion in the presence or absence of 0.1% DCA. The protein ladder was loaded into lane 1, the unprocessed protein mixture (10 μ g) was loaded into lane 2, and the experimental protein mixture digestions (10 μ g) were loaded into lanes 3 through 10, of an SDS-PAGE gel.

Solubilization and digestion comparison

One of the drawbacks of in-solution digestion is the requirement for sample clean-up to remove residual contaminants prior to LC-MS analysis. The FASP workflow offers the ability to effectively purify protein samples from many kinds of contaminants, before and after alkylation.

Urea is utilized in the FASP protocol through alkylation, and then it may be removed prior to enzymolysis. 2M urea used during digestion in the FASP protocol increases peptide IDs (11,137 vs. 13,345), but doesn't exert great changes in the number of protein identifications or protein classes, e.g., membrane and nuclear proteins³. Usage of urea during digestion increases the chance of carbamylation reactions and necessitates additional steps for its removal from the resulting digest. We find poor sensitivity when FASP is performed with ABC as the digestion buffer, and hypothesize that trypsin enhancing surfactants can be incorporated into the FASP workflow to increase performance and sensitivity, while avoiding chromatographic cleanup steps (Figure S1).

DCA and nOG greatly enhance in-solution and in-gel trypsin digestion^{22,33}. An additional benefit of these surfactants with respect to FASP is the ability to digest proteins at higher temperatures without the risk of carbamylation of cysteine, lysine, or arginine residues. The incorporation of removable surfactants, DCA and nOG, into the FASP workflow is examined for effects on digestion efficiency and sample loss.

An *E. coli* lysate was prepared, and aliquots were processed by precipitation and in-solution digestion, by FASP, or by eFASP with experimental digestion buffers. The 2D Clean Up Kit (GE) was used for protein precipitation, and the resulting proteins were

reduced, alkylated, and digested in-solution with 0.1% DCA. Microcon Filter Units were used for FASP and eFASP processing of the remaining samples. FASP processed samples were digested with ABC, and eFASP processed samples were digested with 0.1% nOG, or 0.1% DCA in ABC. All digests were analyzed by LC-MS^E and compared.

Identification comparison

FASP processing resulted in large sample loss compared to processing by 2DCleanUpKit-0.1DCA, eFASP-0.1nOG, and eFASP-0.1DCA (Table 1-1). FASP-0.1DCA yielded the best identification results, with 13% and 27% more protein identifications than FASP-0.1nOG and 2DCleanUpKit-0.1DCA, respectively. While eFASP-0.1nOG does not perform as well as eFASP-0.1DCA, both were significantly better than FASP-ABC and 2D Clean Up Kit, in spite of the fact twice as much theoretical material was injected for the latter two in an attempt to more closely compare the samples qualitatively. FASP-0.1DCA allows for a 30% increase in Pass 1 peptide IDs (the highest quality peptides identified by PLGS that contain no variable modifications) compared to the 2DCK-0.1DCA. These indicate that digestion in FASP is greatly hindered when detergents, chaotropes, and surfactants are not present.

Qualitative comparison

Samples processed by eFASP-0.1DCA maintain increased sensitivity and sample purity compared to samples processed by 2DCleanUpKit-0.1DCA. Base peak intensity (BPI) chromatogram comparisons between samples processed by eFASP-0.1DCA or by 2DCleanUpKit-0.1DCA assesses efficacy of sample cleaning and general

peptide abundance patterns (Figure 1-4). Although twice as much material was analyzed from the 2DCleanupKit-0.1DCA processed sample, the magnitude of chromatographic features was far below those of FASP-0.1DCA. In the first half of the BPI chromatogram from the 2DCleanupKit-0.1DCA sample large interference peaks are present that indicate inefficient removal of contaminants by precipitation. The eFASP-0.1DCA sample also displays more chromatographic features with higher intensities, particularly in the later, more hydrophobic, region of the gradient.

Comparison of eFASP performed in the presence of 0.1% DCA and 0.1% nOG reveals differences that indicate differential modes of digestion enhancement and degrees of increased efficiency (Figure 1-5). Both eFASP-0.1nOG and eFASP-0.1DCA produce exceptionally clean digests. The eFASP-0.1DCA sample maintains more BPI chromatographic features in the middle and later portions of the gradient, and overall greater peak intensities than eFASP-0.1nOG. The disparity between eluting peak intensity widens with increasing organic. FASP-0.1DCA maintains better representation of hydrophobic species, and, perhaps, functions on a wider range of proteins than nOG.

In addition to the number of proteins identified, the percent of proteins with trans-membrane helices (TMH) is affected by the different protocols (Table 1-1a). 9.1% of protein identifications by eFASP-0.1DCA have at least 1 TMH, while FASP-0.1nOG and 2dCleanUpKit trail with 8.3% and 7%. One protein with 11 TMHs is only identified by eFASP-0.1%DCA is the drug efflux protein, Acriflavine resistance protein B (Uniprot AC P31224). The loss of membrane proteins by precipitation is highlighted by these data.

Quantitative comparison

LC-MS^E experiments collect precursor ion information across the elution gradient, which are inherently quantitative. This feature allows for quantitative comparisons based on peptide ion volume. Comparing samples based on protein quantification information will not give an idea as to how well the methods function for peptides. Using peptides for comparison allows for the assessment of a greater feature set covering a large dynamic range.

Data Extraction and Collation (DECO) collected the ion volume information for all Pass 1 peptides identified in samples processed by 2DCleanUpKit-0.1DCA, eFASP-0.1DCA, and eFASP-0.1nOG. The ion volumes for peptides identified across the samples were plotted against each other (Figure 1-6). Plotting the peptide ion volumes from eFASP-0.1nOG against those from eFASP-0.1DCA reveals dissimilarity (Figure 1-6a). There are groups of peptides which are more abundant under each condition, and while the overall correlation between the methods is not terrible ($R^2 = 0.733$), it adds more evidence to possibility that each surfactant has specificity for certain protein populations or features.

Plotting peptide ion volumes from FASP-0.1nOG or FASP-0.1DCA against those obtained from 2DCleanUpKit-0.1DCA, reveal striking differences (Figure 1-6b). While FASP-0.1DCA and FASP-0.1nOG may differ compared, they show much higher similarity when compared with 2DCleanUpKit-0.1DCA. Both eFASP-0.1DCA and eFASP-0.1nOG increase protein IDs and peptide IDs in comparison to 2DCleanupKit, but they also allow for large gains in the quantity of those peptides produced. The overall reduction of peptide ion volumes from 2DCleanUpKit-0.1DCA when compared to

eFASP-0.1DCA ($R^2 = 0.286$) indicates that eFASP avoids the significant challenges that precipitation methods are subject to: sample loss, re-solubilization, and contaminant persistence.

A.

<u>Prep Method</u>	<u>Digest Buffer</u>	<u>Quantity</u>	<u>Proteins</u>	<u>Peptides</u>		<u>Masses</u>
				<u>Pass 1</u>	<u>Pass 2</u>	
2DCleanupKit-0.1DCA	0.1% DCA, ABC	2 ug	344	1,965	4,576	29,972
FASP	ABC	2 ug	14	26	77	3,736
eFASP-0.1nOG	0.1% n-OG, ABC	1 ug	408	2,141	5,568	31,472
eFASP-0.1DCA	0.1% DCA, ABC	1 ug	471	2,678	6,145	31,744

B.

Proteins identified with specified number of TMHs				
<u># TMHs</u>	<u>2DCleanupKit-0.1DCA</u>	<u>FASP-ABC</u>	<u>eFASP-0.1nOG</u>	<u>eFASP-0.1DCA</u>
12	1	0	1	1
11	0	0	0	1
9	1	0	1	1
6	0	0	1	2
5	2	0	2	3
4	1	0	0	2
3	2	0	3	3
2	2	0	6	7
1	15	0	20	23
Total	24	0	34	43
% of all IDs	7.0	0	8.3	9.1

Table 1-1. Analysis of an *E. coli* lysate (162.5 µg) processed and digested by traditional and experimental methods. Lysate was either precipitated and digested in-solution in the presence of 0.1% DCA, or cleaned in a Microcon Filter Unit and digested in the presence of ABC (FASP), 0.1% nOG (eFASP-0.1nOG), or 0.1% DCA (eFASP-0.1DCA). Each digest (1 µg or 2 µg) was analyzed by LC-MS^E and searched for (a) protein and peptide content and (b) predicted transmembrane helices (TMH) in the identified protein sequences.

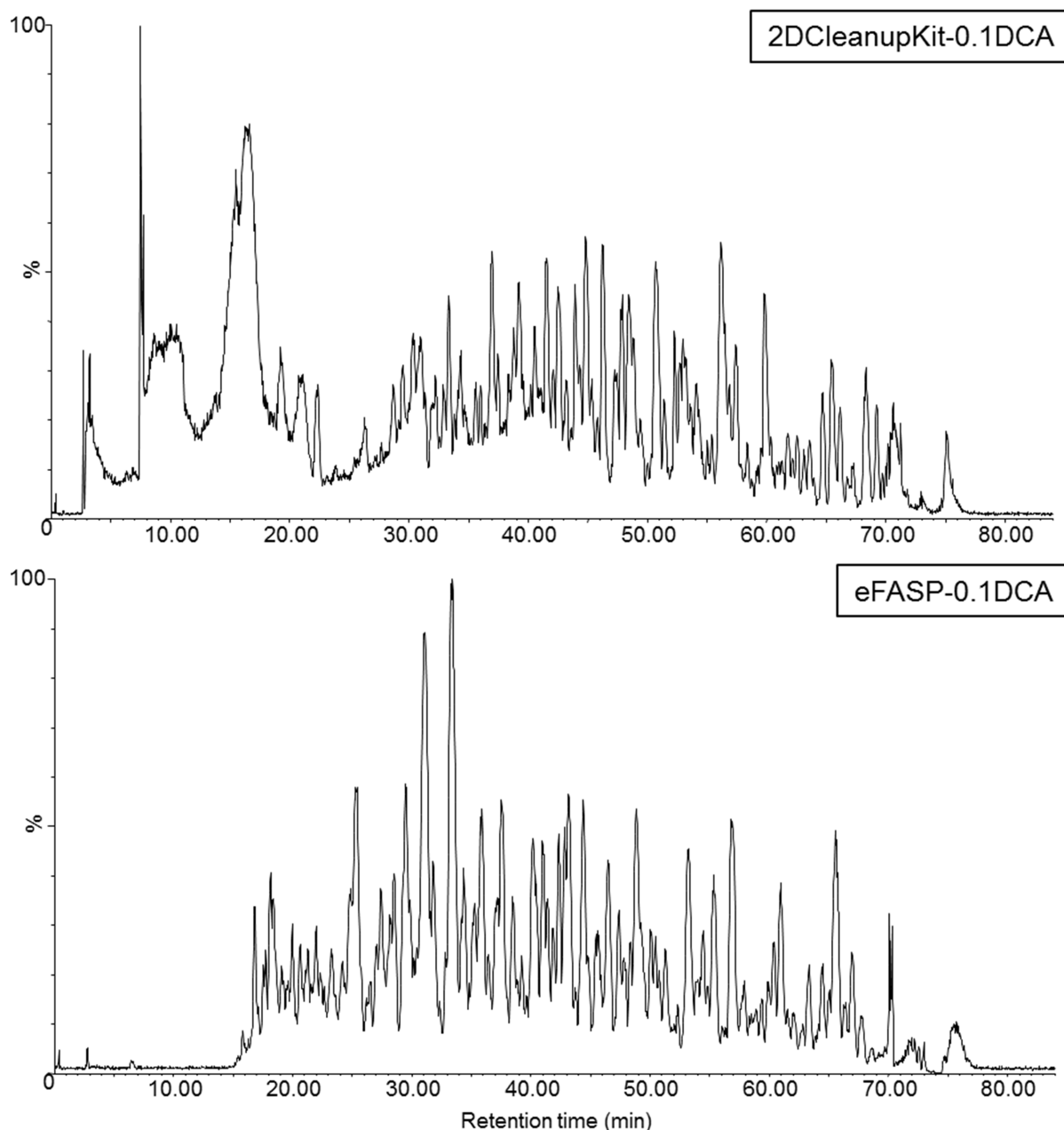


Figure 1-4. Increased digestion efficiency and sample cleanup is achieved with eFASP-0.1DCA. *E. coli* lysate was processed by 2DCleanupKit-0.1DCA or by eFASP-0.1DCA. Each digest was analyzed by LC-MS for complete precursor ion elution profiles. A 5 μ l injection of 2DCleanupKit-0.1DCA digest was required to obtain a base peak intensity chromatogram comparable to a 1 μ l injection of eFASP-0.1DCA digest. The maximum value of the Y-axis was set to 2.84×10^3 .

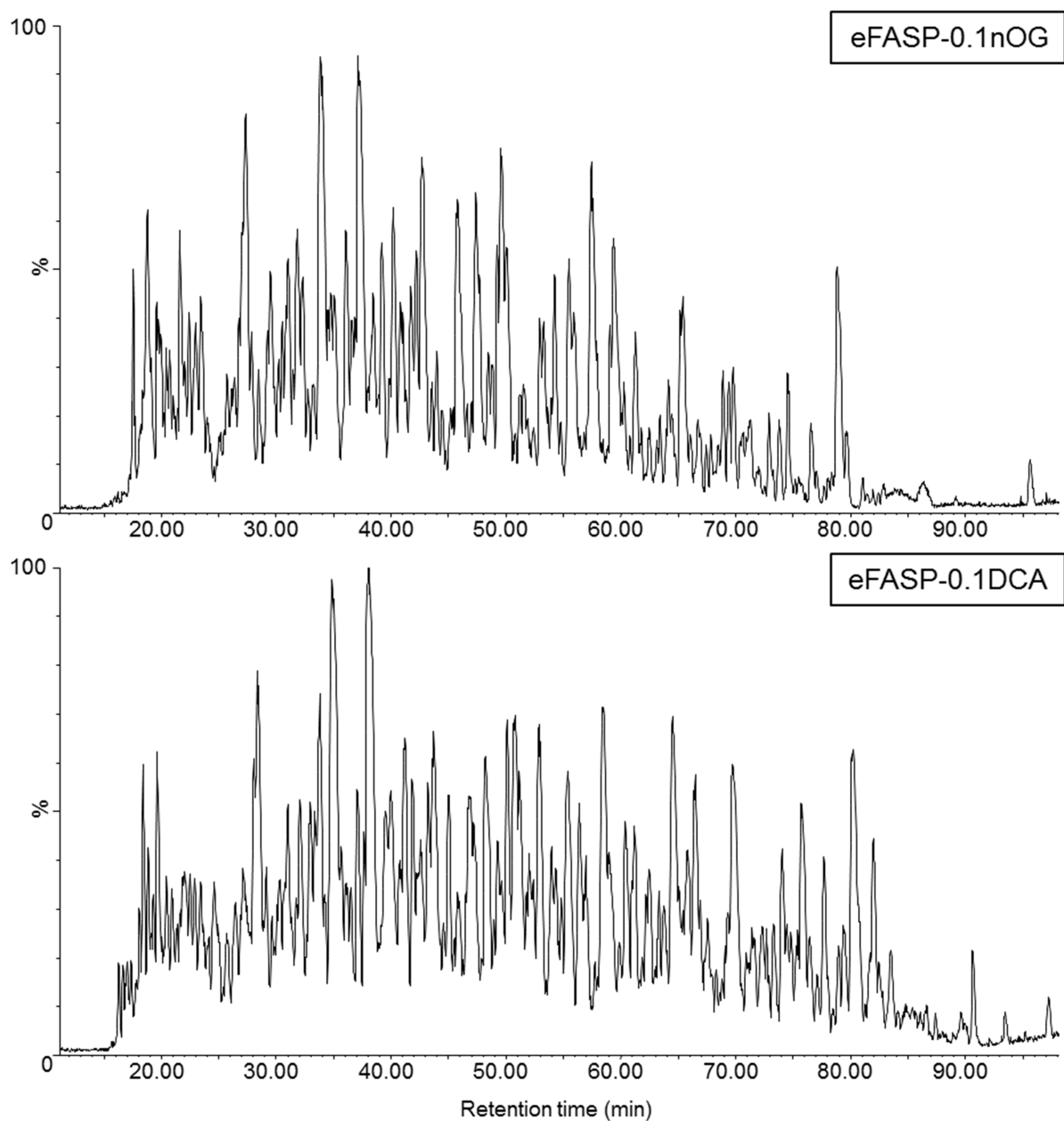


Figure 1-5. Increased digestion efficiency and chromatographic features for *E. coli* lysate processed by eFASP-0.1DCA than with eFASP-0.1nOG. Each digest (1 μ g) was analyzed by LC-MS, with a 60 minute LC gradient. The maximum value of the Y-axis was set to 2.68e3.

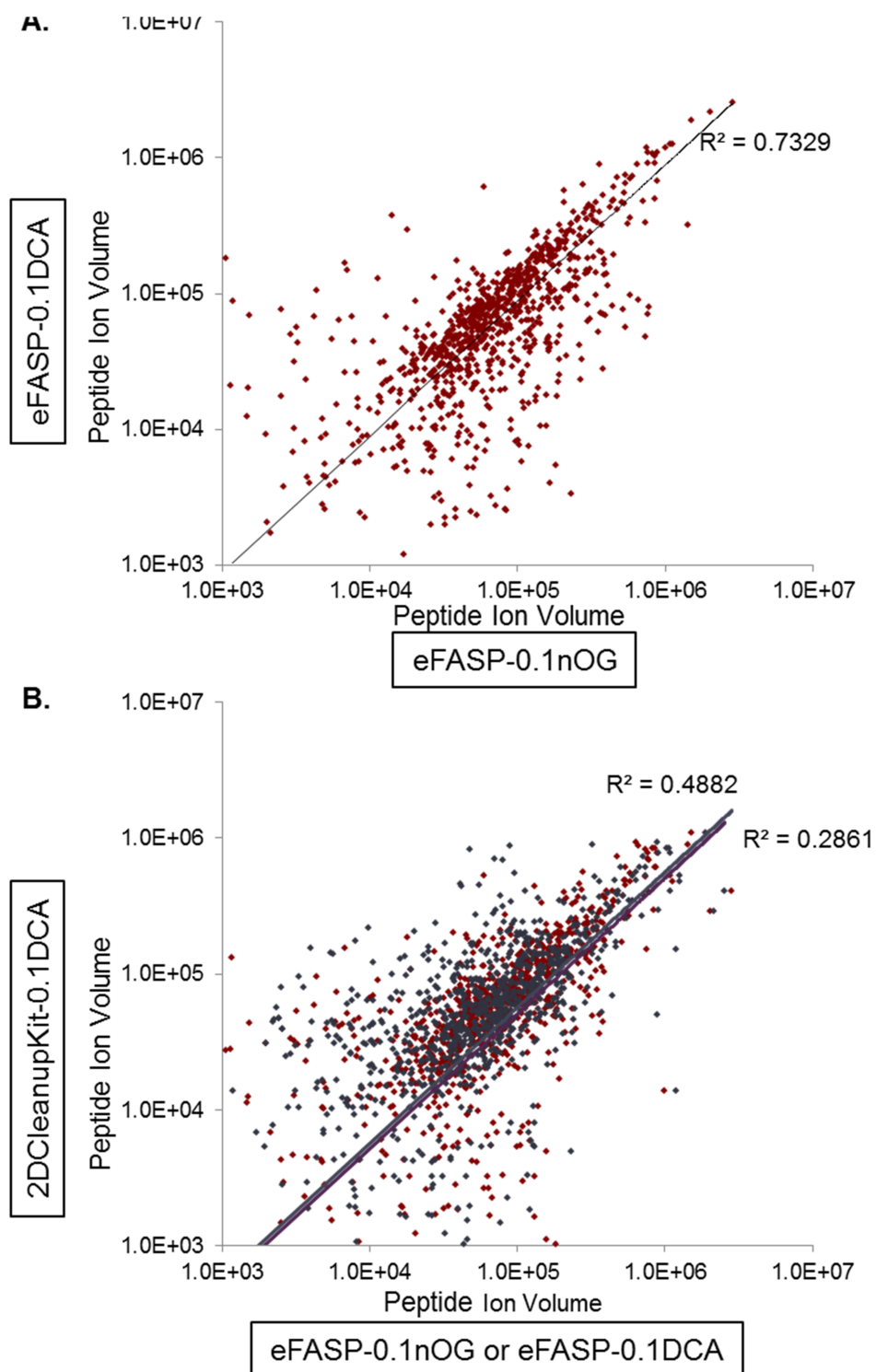


Figure 1-6. Quantitative comparison of digests from precipitation and from eFASP. *E. coli* lysate (50 μ g) was processed and digested by 2DCleanupKit-0.1DCA, eFASP-0.1nOG, or eFASP-0.1DCA, and the resulting digests were analyzed by LC-MS^E. Peptide ion volumes were compared and plotted for those peptides quantified in (a) FASP-0.1nOG and FASP-0.1DCA, and (b) 2D Clean Up Kit, FASP-0.1nOG and FASP-0.1DCA.

The effects of DCA concentration on digestion efficiency

The incorporation of 0.1% DCA into FASP enables large enhancements in quantitative proteome coverage. NaDOC is compatible with trypsin digestion over a wide range of concentrations, with benefits increasing from 0.01% to 1% NaDOC. Deoxycholic acid is less soluble than NaDOC, and we find 0.2% to be around the upper limit of solubility for DCA in ABC buffer. Incorporation of 0.2% DCA into the FASP workflow (eFASP-0.2DCA) was examined and compared to eFASP-0.1DCA.

LC-MS^E of *E. coli* samples processed by eFASP-0.1DCA and eFASP-0.2DCA highlight benefits realized with increased DCA. FASP-0.1DCA and FASP-0.2DCA identify a similar number of proteins and Pass 1 peptides (Table 1-2a). Quantitative comparison of peptides produced by eFASP-0.1DCA and eFASP-0.2DCA show high correlation ($R^2 = 0.992$), indicating a similar population of proteins affected while maintaining quantitative similarity (Figure 1-7). The main improvement afforded by FASP-0.2DCA is a 21% increase in the identification of proteins with trans-membrane helices (TMHs) as annotated by the TMHMM algorithm (Table 1-2b). A major component of carbohydrate translocation, PTS system glucose-specific EIICB component (Uniprot AC P69786), has 10 theoretical TMHs, and is only identified by eFASP-0.2DCA processing. Increasing DCA concentration from 0.1% to 0.2% in eFASP provides more effective solubilization of hydrophobic species, while not sacrificing hydrophilic protein representation or quantitative peptide yield.

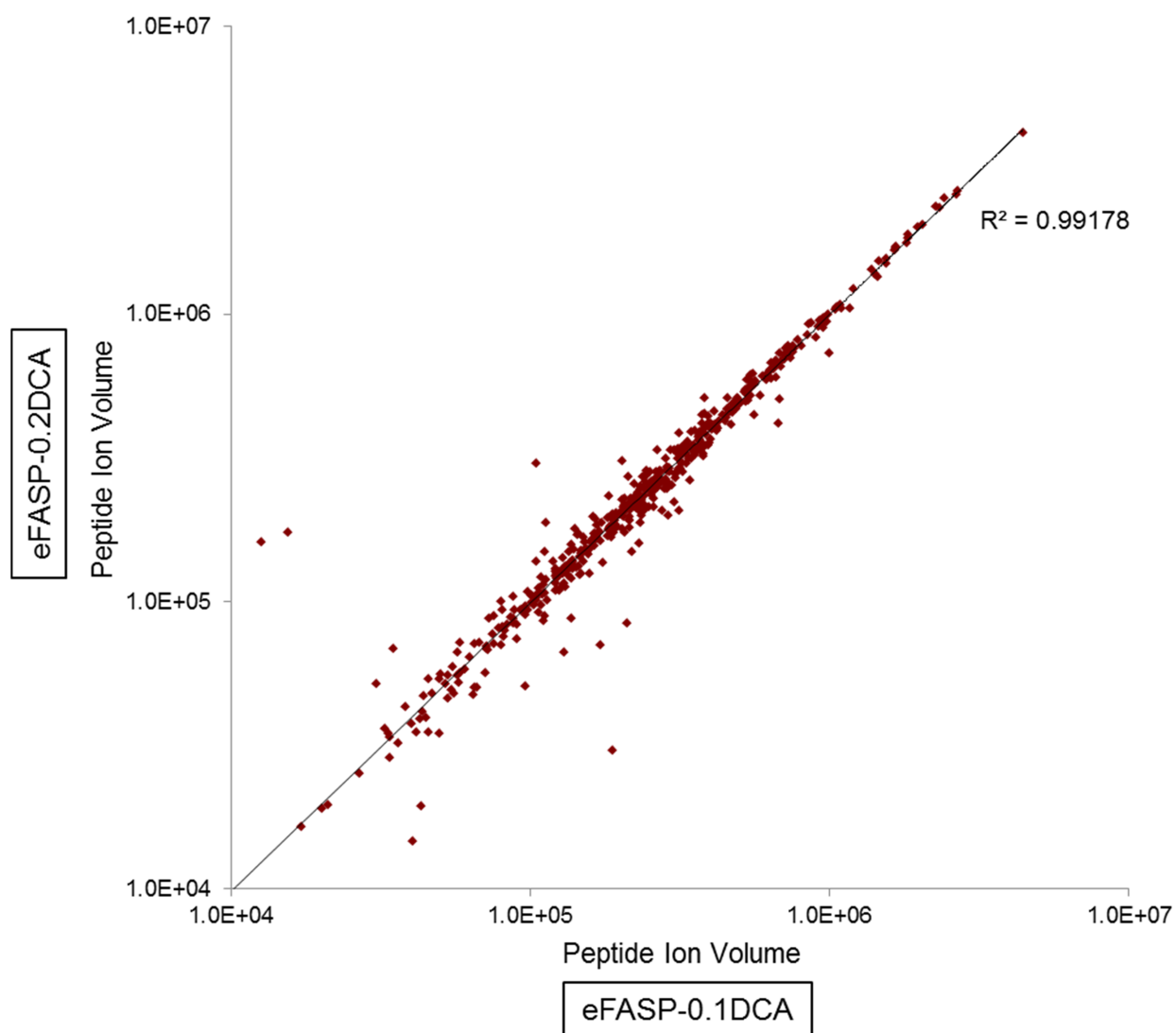


Figure 1-7. Quantitative comparison of peptide yield from eFASP processing of *E. coli* lysate (50 µg) with 0.1% DCA and 0.2% DCA in the digestion buffer. The resulting digests were analyzed by LC-MS^E. Peptide ion volumes were extracted, compared and plotted.

A.

<u>Prep Method</u>	<u>Digest Buffer</u>	<u>Proteins</u>	<u>Peptides</u>	
			<u>Pass 1</u>	<u>Pass 2</u>
eFASP-0.1nOG	0.1% nOG, ABC	202	989	3,766
eFASP-0.1DCA	0.1% DCA, ABC	253	1,304	4,947
eFASP-0.2DCA	0.2% DCA, ABC	256	1,392	4,842
eFASP-0.1DCA-0.1nOG	0.1% DCA, 0.1% nOG, ABC	249	1,394	4,689

B.

Proteins identified with specified number of TMHs			
<u># TMHs</u>	<u>eFASP-0.1DCA</u>	<u>eFASP-0.2DCA</u>	<u>eFASP-0.1DCA-0.1nOG</u>
12	1	1	1
10	0	1	0
9	1	1	0
5	2	1	1
4	1	1	0
3	1	1	1
2	2	2	3
1	10	15	12
Total	18	23	18
% of all IDs	7.1	9.0	7.2

Table 1-2. Analysis of an *E. coli* lysate (50 µg) processed by eFASP with digestion occurring in the presence of 0.1% nOG, 0.1% DCA, 0.2% DCA, or the combination of 0.1% nOG and 0.1% DCA. Each digest (400 ng) was analyzed by LC-MS^E and searched for (a) protein and peptide content and (b) predicted TMHs in the identified protein sequences.

Combination of surfactants

Qualitative and quantitative variation in proteins and peptides produced by FASP-0.1DCA and FASP-0.1nOG suggest that each surfactant has a bias toward certain populations of proteins. The combination of nOG and DCA during eFASP digestion (eFASP-0.1DCA-0.1nOG) was examined and compared to eFASP-0.1DCA, eFASP-0.1nOG, and eFASP-0.2DCA.

LC-MS^E analysis of *E. coli* samples processed by the three methods reveals close similarities between eFASP-0.1DCA and eFASP-0.1DCA-0.1nOG. The combination of 0.1% DCA with 0.1% nOG during digestion rescues the deficiencies associated with 0.1% nOG alone. eFASP-0.1DCA-0.1nOG yields peptide identifications that are slightly better than eFASP-0.1% DCA, and equivalent to eFASP-0.2DCA (Table 1-2a). Quantitative comparison reveals that the eFASP-0.1DCA-nOG has an ion volume profile that closely correlates with that of eFASP-0.1DCA ($R^2 = 0.986$), and eFASP-0.1DCA has already been shown to correlate well with eFASP-0.2DCA (Figure 1-7). eFASP-0.1DCA-0.1nOG has a less strong correlation with eFASP-0.1nOG ($R^2 = 0.815$) (Figure 1-8).

Proteins identified across the sample conditions reveal about a 50% overlap, but also unique differences (Figure 1-9). Although eFASP-0.1DCA-0.1nOG performs as well as eFASP-0.2DCA, there is a large portion of proteins that are only identified in eFASP-0.2DCA. Protein identification results from eFASP-0.2DCA are the most dissimilar to those from the other three conditions, even though the quantity of proteins and peptides identified are similar to eFASP-0.1DCA and eFASP-0.1DCA-0.1nOG.

The Grand Average of Hydropathy (GRAVY) value was calculated for all proteins identified from samples produced by the four eFASP variants (Figure 1-10a). GRAVY values represent the sum of hydropathy values, based on the Kyte-Doolittle scale, for all amino acids in a protein, divided by the amino acid length³². Comparison of proteins based on GRAVY value does not reveal any significant differences between eFASP-0.1DCA, eFASP-0.2DCA and eFASP-0.1DCA-0.1nOG. However, analysis of proteins unique to each condition reveals that in addition to a greater number of proteins

identified by eFASP-0.2DCA, there is also better representation of challenging hydrophobic proteins.

The quantitative and qualitative similarity between eFASP-0.1DCA-0.1nOG and eFASP-0.2DCA indicate that the two surfactants work synergistically to enhance trypsin digestion. However, eFASP-0.2DCA yields the best performance for hydrophobic protein representation. Since nOG can increase the solubility of DCA, these data are useful for enabling future studies of eFASP performance with higher concentrations of DCA during digestion.

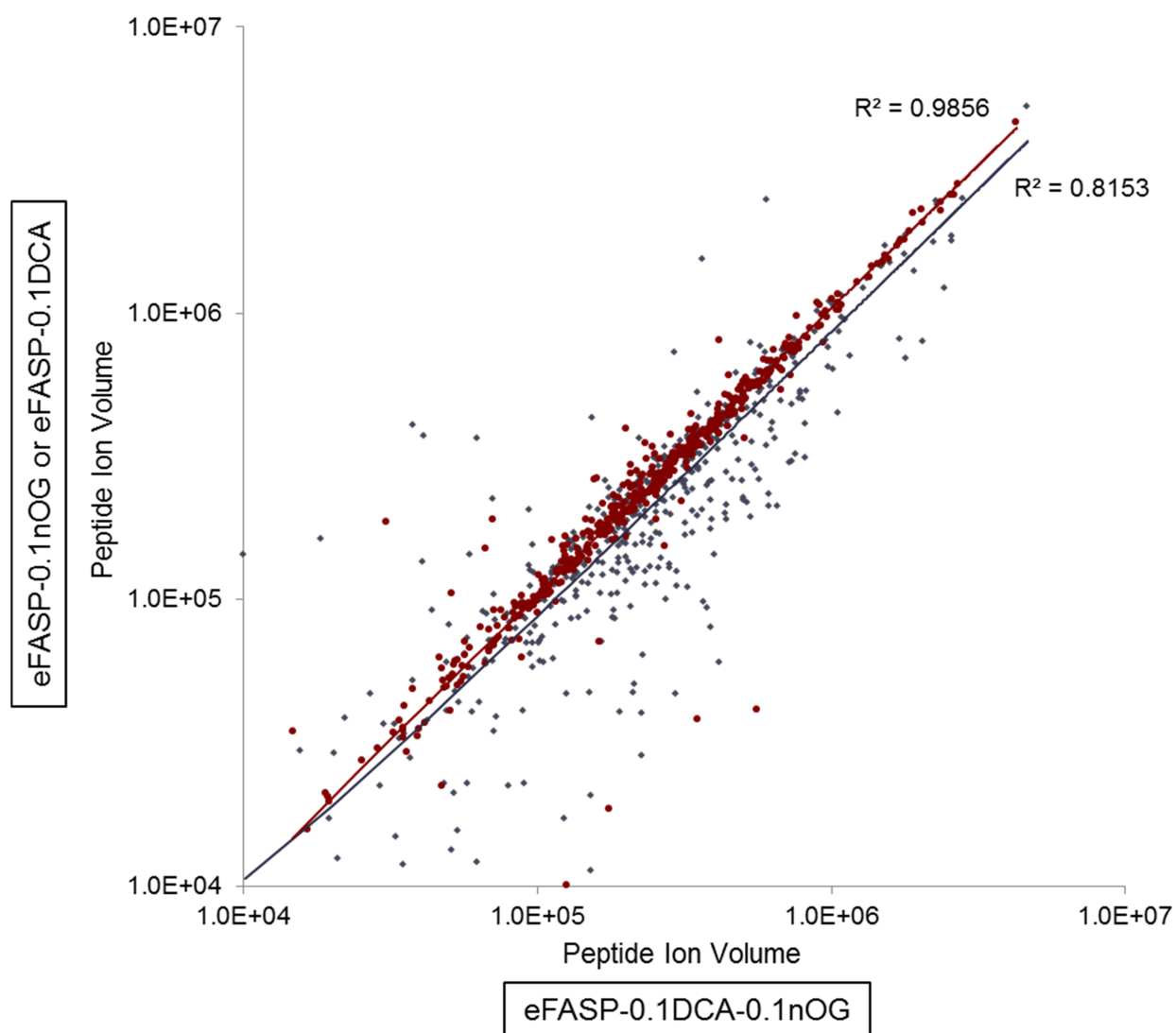


Figure 1-8. Quantitative comparison of peptide yields from eFASP processing of *E. coli* lysate (50 µg) with 0.1% DCA, 0.1nOG, or the combination of 0.1% DCA and 0.1% nOG in the digestion buffer. The resulting digests were analyzed by LC-MS^E. Peptide ion volumes were extracted, compared and plotted. eFASP-0.1DCA-nOG, and y-values represent peptide ion volumes from eFASP-0.1nOG (blue) or eFASP-0.1DCA (red).

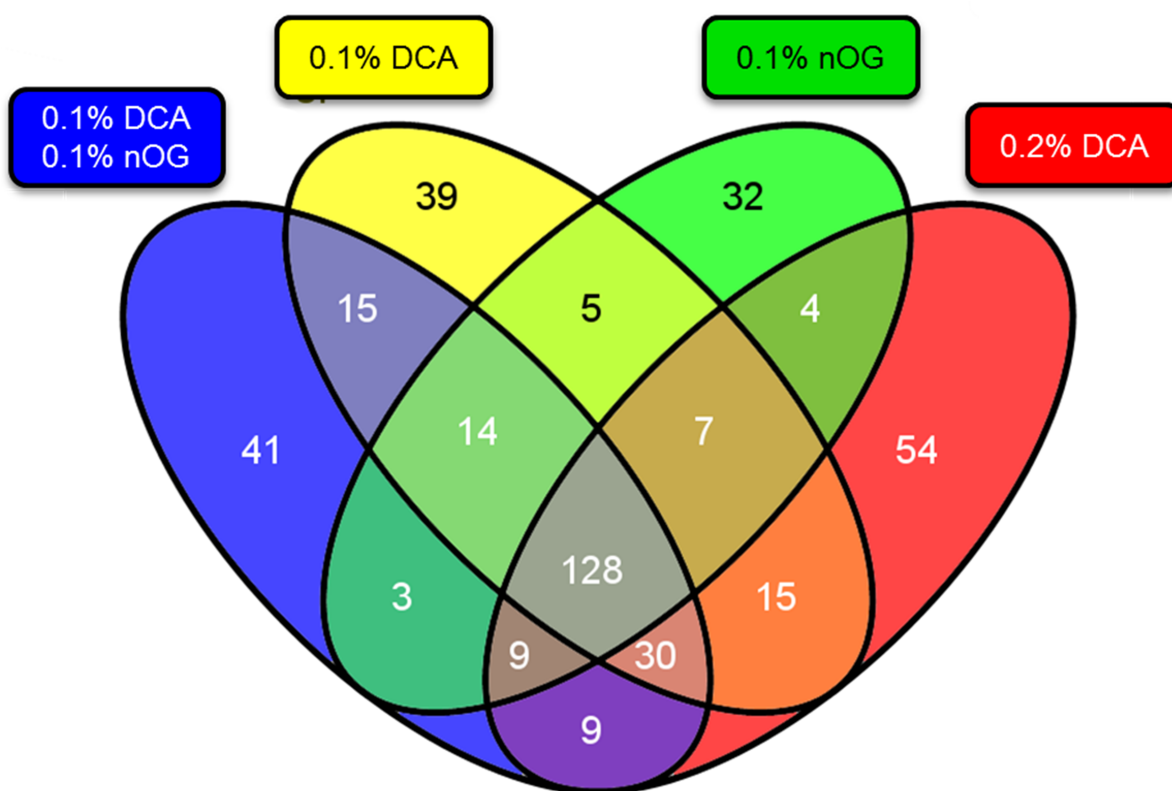


Figure 1-9. Protein identification overlap from eFASP processing of *E. coli* lysate with digestion buffers containing 0.1% nOG (green), 0.1% DCA (yellow), 0.2% DCA (red), and 0.1% DCA + 0.1% nOG (blue).

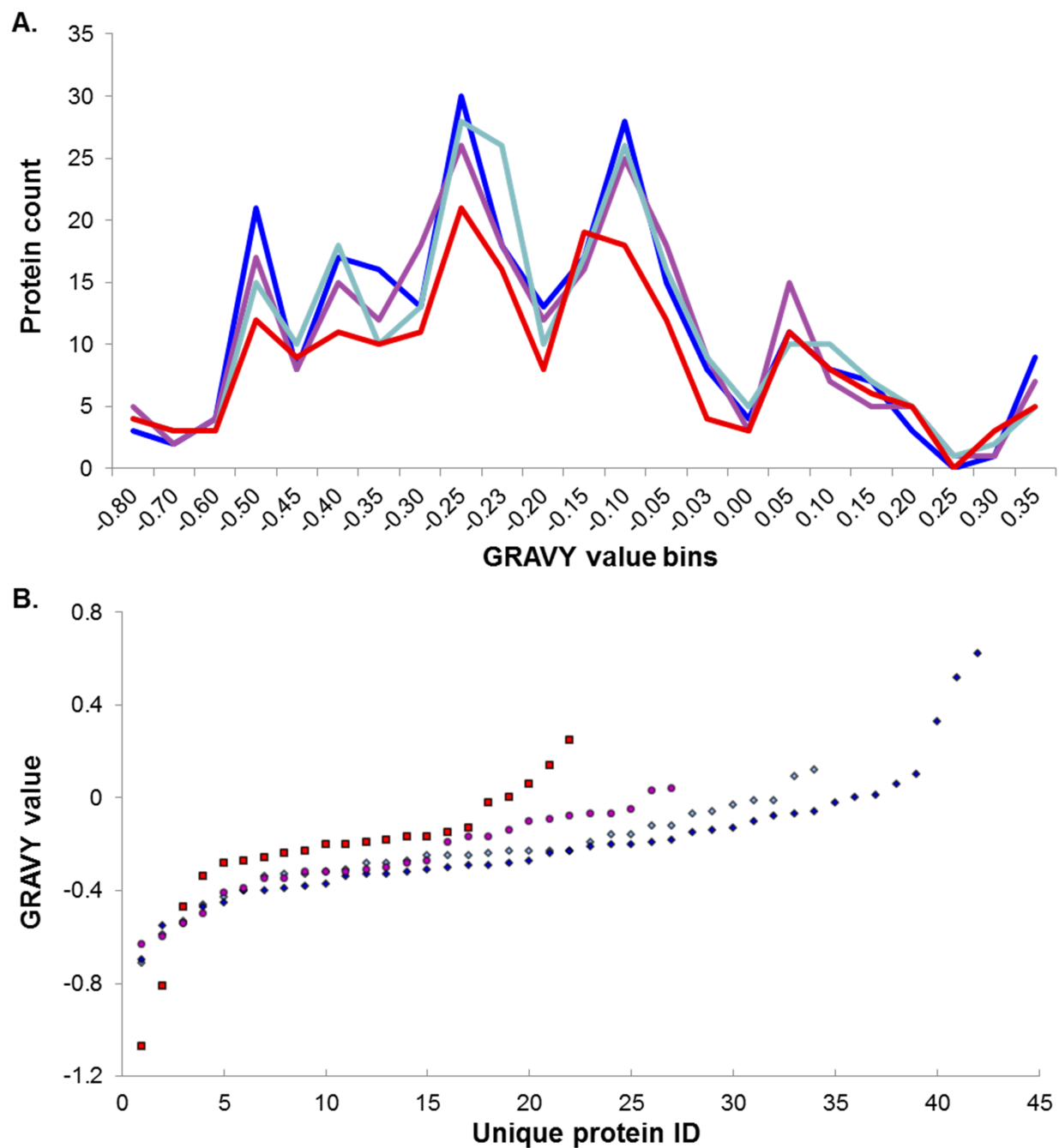


Figure 1-10. Analysis of GRAVY values for proteins identified in *E. coli* lysates processed by eFASP using different digestion buffers. GRAVY values were calculated for (a) all proteins identified in each digest, or (b) proteins unique to one digest condition. (a) GRAVY values for all proteins are binned and plotted for eFASP-0.1DCA (light blue), FASP-0.2DCA (dark blue), eFASP-0.1DCA-0.1nOG (magenta), and eFASP-0.1nOG (red). (b) GRAVY values are plotted for proteins unique to eFASP-0.1DCA (light blue), eFASP-0.2DCA (dark blue), eFASP-0.1DCA-nOG (magenta), and eFASP-0.1nOG (red).

Passivation of Plastics

Protein and peptide binding to surfaces can account for nonspecific losses during sample preparation. In proteomic experiments dealing with small amounts of material, this can be detrimental to qualitative and quantitative analyses. We hypothesize that nonspecific binding to the plastics and to the filter of the Microcon Filter Unit is a major factor in the estimated 50% sample loss associated with FASP.

Passivation of surfaces is not a novel concept, and has been utilized for various surfaces that can bind analytes, e.g., metal and plastic. Often manufacturers provide supplies produced with materials that prevent protein binding, e.g., Eppendorf Lo-bind tubes, or recommendations for reagents useful for passivating surfaces in their products, e.g., EDTA to passivate metal chromatography supports, or dimethyldichlorosilane to increase the hydrophobicity and slickness of glass surfaces. These options are not available for filter units, and the process of siliconization the plastic components of the filter unit is not practical and may be detrimental to the filter.

0.5% polyethylene glycol (PEG 20,000) and dextran T-70 have been utilized as carriers in FASP to increase peptide yields in the high nanogram and low microgram range without affecting LC-MS analysis³⁴. The improvement in peptide yield is around 30% and enabled the analysis of less than 1,000 cells prepared by laser capture microdissection of formalin fixed paraffin embedded tissues. Carriers are effective when starting material is below 10 µg; however, no additional benefit is realized with larger amounts of starting material³⁴.

Carrier compounds are only useful in areas where they are present. There are many surfaces of the Microcon Filter Unit that a sample can contact where carrier

compounds are excluded, including areas of the filter unit below the filter and the walls of the collection tube. Additionally, spontaneous decomposition of PEG carriers in FASP is a source of contamination in downstream LC-MS analyses³⁴. We propose to passivate all surfaces prior to sample application to the Microcon Filter Unit in order to avoid contamination and to reduce binding to all surfaces that come into contact with the protein and peptide solutions.

Millipore has recommended the usage of various reagents when using Amicon Centricon Concentrators with very dilute protein samples. Reagents include 1% BSA, 1% IgG, 1% powdered milk, 5% Tween-20, 5% SDS, 5% PEG, and 5% Triton-X (“Passivation of Amicon Centricon Concentrators for Improved Recovery”, Millipore, PC1003EN00, 1999). The best choices for proteomic work are Tween-20 and SDS, which will avoid proteinaceous contamination.

We utilized high-purity Tween-20 or ammonium lauryl sulfate (ALS) to passivate all surfaces of the Microcon Filter Units and the collection tubes. Filter units and collection tubes were incubated overnight in solutions of 5% Tween-20 or 5% ALS, and then thoroughly rinsed with large volumes of MS-grade water before use. Non-passivated (H₂O) or passivated (T20 or ALS) Microcon units and collection tubes were utilized to process 50 µg of *E. coli* by FASP or by eFASP-0.1DCA or eFASP-0.2DCA. Samples were analyzed by LC-MS in profile mode for qualitative analysis and LC-MS^E for identification and quantification.

LC-MS analysis revealed enhanced signal throughout the gradient for samples processed with passivated materials (Figure 1-11). eFASP-T20-0.1DCA outperforms eFASP-ALS-0.1DCA by a small margin, and both increase sensitivity about two-fold in

comparison to eFASP-H2O-0.1DCA. eFASP-T2O-0.1DCA also retains more hydrophobic species than both eFASP-H2O-0.1DCA and eFASP-ALS-0.1DCA.

In the absence of 0.2% DCA, the passivation of surfaces with Tween-20 prior to FASP increases protein identification by 250% and peptide identification by 300% (Table 1-3). In contrast, eFASP-0.2DCA processing with Tween-20 passivated materials improves protein and peptide identification by 12% and 7%, respectively, compared to eFASP-0.2DCA without passivated materials. Peptide ion volume correlation between eFASP-0.2DCA processing with passivated and non-passivated materials remains high ($R^2 = 0.954$).

Passivation agents, like Tween-20 and ALS, are binding to sites on the surfaces of the plastics, and effectively blocking proteins and peptides from binding. The reduction in passivation gains when DCA is present may be due to DCA acting as a passivation agent itself; however, it is not as effective as Tween-20 or ALS. This method of passivation prior to sample application assures that no contaminant species end up in the final peptide solution collected.

<u>Prep Method</u>	<u>Passivation</u>	<u>Digest Buffer</u>	<u>Proteins</u>	<u>Peptides</u>	
				<u>Pass 1</u>	<u>Pass 2</u>
FASP	No	ABC	21	85	229
FASP-T20	No	ABC	52	246	604
eFASP-0.2DCA	5% T-20	0.2% DCA, ABC	252	1,436	3,645
eFASP-0.2DCA-T20	5% T-20	0.2% DCA, ABC	284	1,531	3,671

Table 1-3. Microcon Filter Units and collection tubes were incubated with water or 5% Tween-20 overnight. Analysis of an *E. coli* lysate (50 µg) processed by eFASP with digestion occurring in the presence of ABC or 0.1% DCA. Each digest (100 ng) was analyzed by LC-MS^E and searched for (a) protein and peptide content and (b) predicted TMHs in the identified protein sequences.

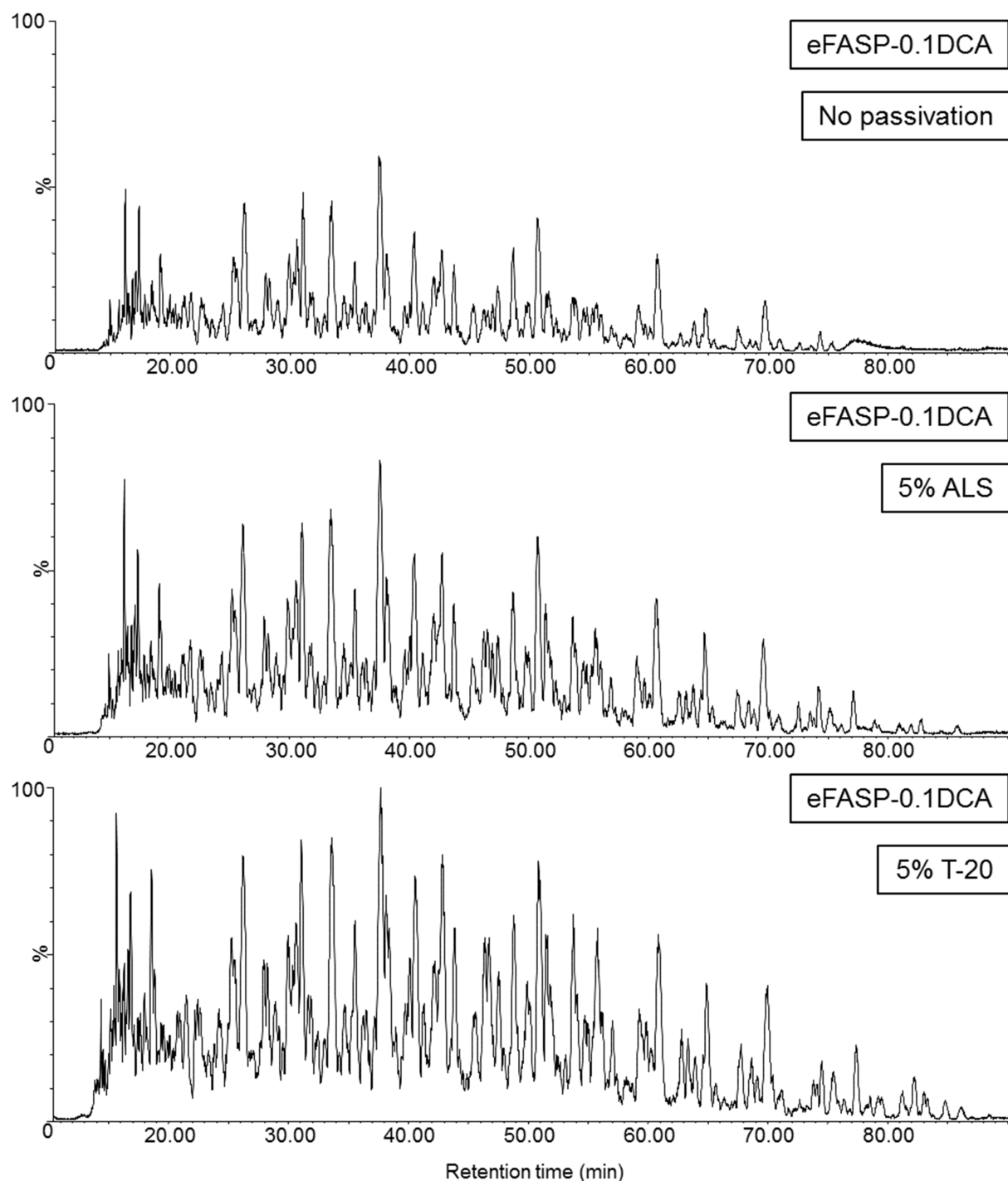


Figure 1-11. *E. coli* lysate was processed by eFASP-0.1DCA using Microcon Filter Units and collection tubes that were either not passivated or passivated by overnight incubation with 5% Tween-20 or 5% ALS. Each digest (200ng) was analyzed by LC-MS. The maximum y-axis value for each base peak intensity chromatogram is 2.11×10^3 .

Express eFASP

FASP includes two major buffer exchange steps before digestion. The first buffer exchange step displaces SDS with urea and removes DTT prior to alkylation. The second buffer exchange step removes residual iodoacetamide and urea prior to digestion. When using the YM-30 units, each step takes about 45 minutes. We propose the usage of alternative reduction and alkylation reagents in the FASP workflow, TCEP and 4-vinylpyridine, as a means to increase alkylation specificity and to enable in-solution alkylation (Figure 1-3, lane 9).

4-vinylpyridine is an alkylating agent with high specificity for cysteine residues ¹⁶. The alkylation reaction is fast and can occur at neutral pH ³⁵. Prolonged periods of incubation with 4-VP does not result in non-cysteine alkylation, and the pH for reactivity assures that all ϵ -amino groups on lysine residues are unreactive. 4-VP is an ideal alkylating agent to use because it is not inactivated by the reducing agent TCEP, allowing for reduction and alkylation to take place in solution. TCEP and 4-VP have been important for enhancing the resolution of basic proteins 2DE experiments ^{36,37}. Alkylation of samples prior to application to the spin filter would allow entry into the FASP workflow after the alkylation step, eliminating the first series of buffer exchanges, and essentially cutting time spent on buffer exchanges in half.

To examine the applicability of TCEP and 4-VP as replacements for DTT and IAN, two *E. coli* lysates were prepared with 4% SDS, and either 100 mM DTT or 7.5 mM TCEP. The lysate prepared with DTT was processed by eFASP-0.2DCA with IAN, as done previously. The lysate prepared with TCEP was alkylated in-solution with 4-VP and then processed by eFASP-0.2DCA.

LC-MS^E analysis of the sample processed by FASP-0.2DCA-IAN yielded 284 protein identifications and 1,531 first pass peptides, which is approximately 15% greater than those obtained for the eFASP-0.2DCA-4VP sample (Table 1-4). Further comparisons were done to address this disparity, since chromatograms for the samples were similar in elution pattern and intensity.

Peptides identified in both DTT/IAN and TCEP/4VP were quantitatively compared (Figure 1-12). While there is some deviance in more abundant species, the overall correlation of peptide ion volume profiles is very good ($R^2 = 0.972$). To assess sequence coverage of proteins in each sample, DECO was used to collect and process sequence coverage information for all proteins identified in each sample. The sequence coverage values were binned and tabulated for comparison (Figure 1-13). Samples processed with IAN/DTT or TCEP/4VP maintain relatively similar sequence coverage distributions.

In order to rule out poor alkylation due to inefficient reduction with 7.5 mM TCEP, data were compared to lysate prepared with 50 mM TCEP and processed by eFASP-0.2DCA. The TCEP/4VP sample results from this experiment were fairly similar to those from the previous experiment, with 101 S-pyridylethyl modified first pass peptides identified in this experiment, and 104 identified previously; however, both are far less than the 203 first pass peptides identified with carbamidomethyl modifications in the DTT/IAN sample from his experiment. The workflow parameters were also modified to check for incomplete alkylation. Changing the S-pyridylethyl modification from fixed to variable, or removing it completely, did not result in improved results. These observations, in light of the high similarity between the DTT/IAN and the TCEP/4VP

samples (chromatograms, sequence coverage, and identified masses), indicate possible computational complications for the proper identification or scoring of S-pyridylethyl cysteine sulfoxide-containing peptides and their fragments.

In conclusion, while there is some unresolved discrepancy between FASP-0.2DCA performed with DTT/IAN and with TCEP/4VP, the presented workflow is effective, reduces potential for side reactions by IAN or urea³⁸, and greatly reduces sample processing.

<u>Prep Method</u>	<u>Passivation</u>	<u>Reduction/Alkylation</u>	<u>Proteins</u>	<u>Peptides</u>	
				<u>Pass 1</u>	<u>Pass 2</u>
eFASP-0.2DCA-IAN	5% T-20	DTT / IAN	284	1,531	3,671
eFASP-0.2DCA-4VP	5% T-20	TCEP / 4VP	241	1,297	3,373

Table 1-4. *E. coli* lysates were prepared with solubilization buffer containing either DTT or TCEP. Lysate prepared with DTT was processed by eFASP with alkylation by iodoacetamide. Lysate prepared with TCEP was alkylated in-solution and then processed by eFASP, skipping the steps up to and including in-filter alkylation. Aliquots from each digest (100ng) were analyzed by LC-MS^E and searched for protein and peptide content.

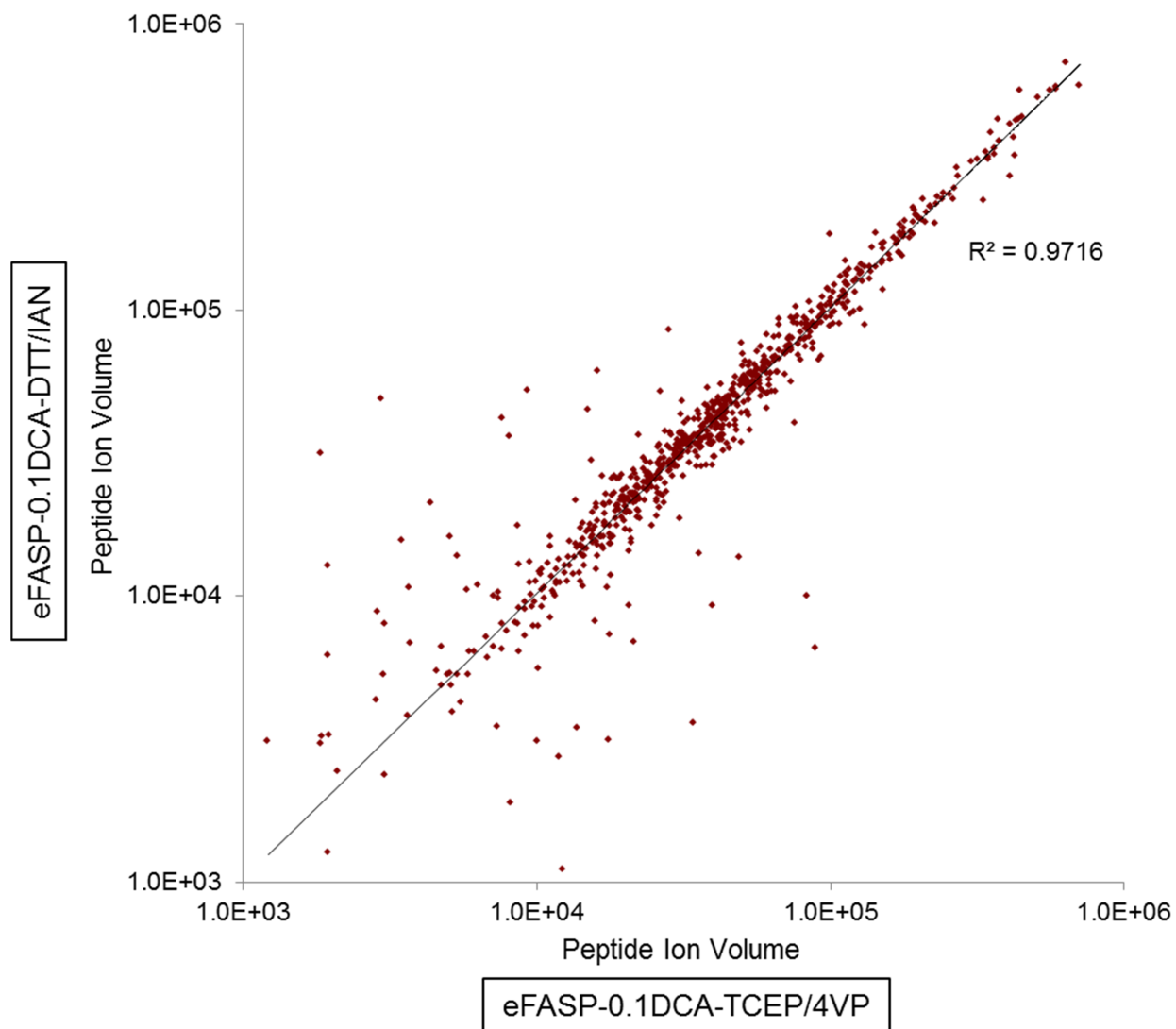


Figure 1-12. Quantitative comparison of peptide yield from eFASP-0.2DCA processing of *E. coli* lysate that was either reduced with DTT and alkylated in-filter with IAN or reduced with TCEP and alkylated in-solution with 4-VP. The resulting digests were analyzed by LC-MS^E. Peptide ion volumes were extracted, compared and plotted.

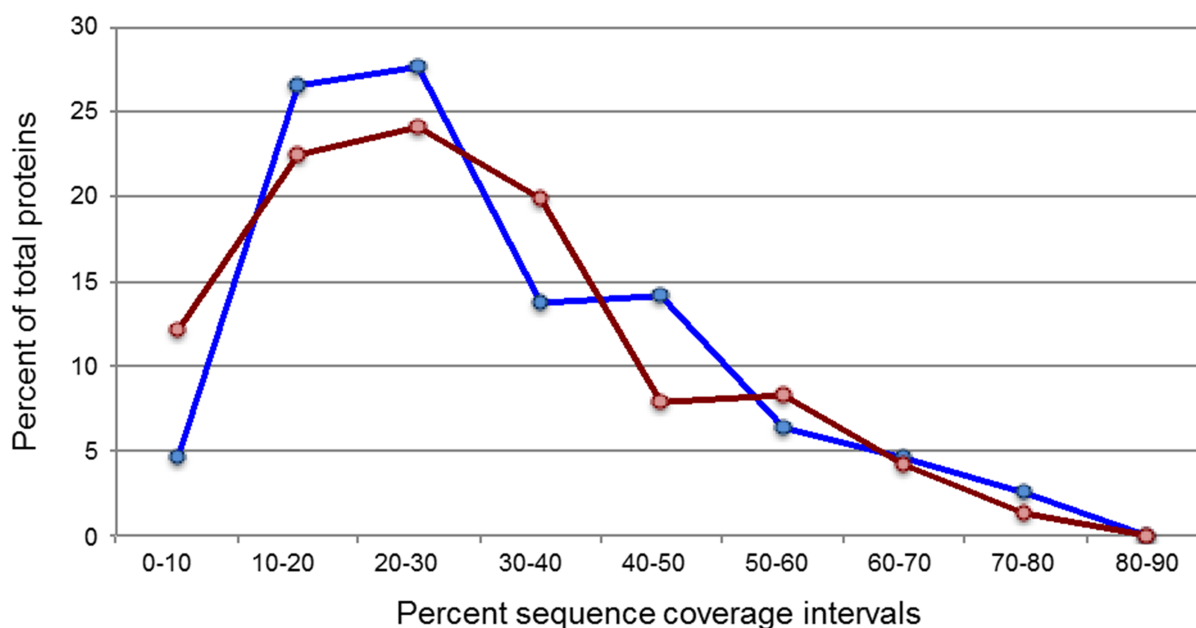


Figure 1-13. Sequence Coverage for *E. coli* samples processed by eFASP with reduction and alkylation accomplished with either DTT/IAN (blue) or TCEP/4VP (red).

CONCLUSION

The usage of 0.2% DCA in the digestion buffer during eFASP increases proteome coverage and reduces sample loss. The sensitivity of FASP is increased more than 10-fold when the ABC digestion buffer is replaced with 0.2% DCA. Comparison of eFASP-0.2%DCA with a sample processed with the 2D Clean Up Kit clearly illustrate the major problems with precipitation: sample loss and incomplete sample clean up. Sample sensitivity was also increased about 2-fold by the passivation of plastics with 5% Tween-20. We also introduce the incorporation of alternative reduction and alkylation reagents that can reduce experimental time.

REFERENCES

1. Manza, L. L., Stamer, S. L., Ham, A. J. L., Codreanu, S. G. & Liebler, D. C. Sample preparation and digestion for proteomic analyses using spin filters. *Proteomics* **5**, 1742–1745 (2005).
2. Wisniewski, J. R. & Mann, M. Spin filter–based sample preparation for shotgun proteomics. *Nat Methods* **6**, 785–786 (2009).
3. Wisniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nat Methods* **6**, 359–362 (2009).
4. Wisniewski, J. R., Zielinska, D. F. & Mann, M. Comparison of ultrafiltration units for proteomic and N-glycoproteomic analysis by the filter-aided sample preparation method. *Anal Biochem* **410**, 307–309 (2011).
5. Wisniewski, J. R., Zougman, A. & Mann, M. Combination of FASP and StageTip-based fractionation allows in-depth analysis of the hippocampal membrane proteome. *J Proteome Res* **8**, 5674–5678 (2009).
6. Wisniewski, J. R., Nagaraj, N., Zougman, A., Gnad, F. & Mann, M. Brain phosphoproteome obtained by a fasp-based method reveals plasma membrane protein topology. *J Proteome Res* **9**, 3280–3289 (2010).
7. Zielinska, D. F., Gnad, F., Jedrusik-Bode, M., Wisniewski, J. R. & Mann, M. *Caenorhabditis elegans* has a phosphoproteome atypical for metazoans that is enriched in developmental and sex determination proteins. *J Proteome Res* **8**, 4039–4049 (2012).
8. Zielinska, D. F., Gnad, F., Wisniewski, J. R. & Mann, M. Precision mapping of an *in vivo* n-glycoproteome reveals rigid topological and sequence constraints. *Cell* **141**, 897–907 (2010).
9. Wisniewski, J. R. & Mann, M. Consecutive proteolytic digestion in an enzyme reactor increases depth of proteomic and phosphoproteomic analysis. *Anal Chem* **84**, 2631–2637 (2012).
10. Tran, B. Q. *et al.* Addressing trypsin bias in large scale (phospho)proteome analysis by size exclusion chromatography and secondary digestion of large post-trypsin peptides. *J Proteome Res* **10**, 800–811 (2011).
11. Ong, S.-E. *et al.* Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* **1**, 376–386 (2002).
12. Deeb, S. J., D'Souza, R., Cox, J., Schmidt-Supprian, M. & Mann, M. Super-SILAC allows classification of diffuse large B-cell lymphoma subtypes by their protein expression profiles. *Mol Cell Proteomics* **11**, 77–89 (2012).

13. Ross, P. L. *et al.* Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics* **3**, 1154–1169 (2004).
14. Gygi, S. P. *et al.* Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* **17**, 994–999 (1999).
15. Sechi, S. A method to identify and simultaneously determine the relative quantities of proteins isolated by gel electrophoresis. *Rapid Commun Mass Spectrom* **16**, 1416–1424 (2002).
16. Sebastiano, R., Citterio, A., Lapadula, M. & Righetti, P. G. A new deuterated alkylating agent for quantitative proteomics. *Rapid Commun Mass Spectrom* **17**, 2380–2386 (2003).
17. Liu, H., Sadygov, R. G. & Yates, J. R. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem* **76**, 4193–4201 (2004).
18. Ishihama, Y. *et al.* Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol Cell Proteomics* **4**, 1265–1272 (2005).
19. Geromanos, S. J. *et al.* The detection, correlation, and comparison of peptide precursor and product ions from data independent LC-MS with data dependant LC-MS/MS. *Proteomics* **9**, 1683–1695 (2009).
20. Silva, J. C., Gorenstein, M. V., Li, G. Z., Vissers, J. P. & Geromanos, S. J. Absolute quantification of proteins by LCMS^E: a virtue of parallel MS acquisition. *Mol Cell Proteomics* **5**, 144–156 (2006).
21. Silva, J. C. *et al.* Simultaneous qualitative and quantitative analysis of the *Escherichia coli* proteome: a sweet tale. *Mol Cell Proteomics* **5**, 589–607 (2006).
22. Masuda, T., Tomita, M. & Ishihama, Y. Phase transfer surfactant-aided trypsin digestion for membrane proteome analysis. *J Proteome Res* **7**, 731–740 (2008).
23. Masuda, T., Sugiyama, N., Tomita, M. & Ishihama, Y. Microscale phosphoproteome analysis of 10,000 cells from human cancer cell lines. *Anal Chem* **83**, 7698–7703 (2011).
24. Masuda, T., Saito, N., Tomita, M. & Ishihama, Y. Unbiased quantitation of escherichia coli membrane proteome using phase transfer surfactants. *Mol Cell Proteomics* **8**, 2770–2777 (2009).
25. Zhou, J. *et al.* Evaluation of the application of sodium deoxycholate to proteomic analysis of rat hippocampal plasma membrane. *J Proteome Res* **5**, 2547–2553 (2006).

26. Lin, Y. *et al.* Sodium-deoxycholate-assisted tryptic digestion and identification of proteolytically resistant proteins. *Anal Biochem* **377**, 259–266 (2008).
27. Proc, J. L. *et al.* A quantitative study of the effects of chaotropic agents, surfactants, and solvents on the digestion efficiency of human plasma proteins by trypsin. *J Proteome Res* **9**, 5422–5437 (2010).
28. Lin, Y. *et al.* Evaluation and optimization of removal of an acid-insoluble surfactant for shotgun analysis of membrane proteome. *Electrophoresis* **31**, 2705–2713 (2010).
29. Yeung, Y. G., Nieves, E., Angeletti, R. H. & Stanley, E. R. Removal of detergents from protein digests for mass spectrometry analysis. *Anal Biochem* **382**, 135–137 (2008).
30. Li, G. Z. *et al.* Database searching and accounting of multiplexed precursor and product ion spectra from the data independent analysis of simple and complex peptide mixtures. *Proteomics* **9**, 1696–1719 (2009).
31. Silva, J. C. *et al.* Quantitative proteomic analysis by accurate mass retention time pairs. *Anal Chem* **77**, 2187–2200 (2005).
32. Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J Mol Biol* **157**, 105–132 (1982).
33. Katayama, H., Nagasu, T. & Oda, Y. Improvement of in-gel digestion protocol for peptide mass fingerprinting by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom* **15**, 1416–1421 (2001).
34. Wisniewski, J. R., Ostasiewicz, P. & Mann, M. High recovery FASP applied to the proteomic analysis of microdissected formalin fixed paraffin embedded cancer tissues retrieves known colon cancer markers. *J Proteome Res* **10**, 3040–3049 (2011).
35. Righetti, P. G. Real and imaginary artefacts in proteome analysis via two-dimensional maps. *J Chromatogr B* **841**, 14–22 (2006).
36. Bai, F., Liu, S. & Witzmann, F. A. A ‘de-streaking’ method for two-dimensional electrophoresis using the reducing agent tris(2-carboxyethyl)-phosphine hydrochloride and alkylating agent vinylpyridine. *Proteomics* **5**, 2043–2047 (2005).
37. Liu, S., Bai, F. & Witzmann, F. Destreaking strategies for two-dimensional electrophoresis. *Separation Methods In Proteomics* 207–217 (2006).
38. Boja, E. S. & Fales, H. M. Overalkylation of a protein digest with iodoacetamide. *Anal Chem* **73**, 3576–3582 (2001).

CHAPTER 2

Data Extraction and Correction (DECO): software for the accurate merging of fractionated LC-MS^E data, and the extraction and quantification of proteins identified in the ProteinLynx Global Server pipeline.

ABSTRACT

As with many data analysis software tools under development, there are issues that arise which require custom software to enable correct processing. The first problem concerns those peptides that can be identified in multiple fractions. Due to the way in which ProteinLynx Global Server (PLGS) handles fraction merging, and the requirement for absolute reproducibility of retention time between LC-MS^E runs, there are instances in which the same peptide, found in multiple fractions, are not correctly merged. In these cases, rather than the merge file containing one precursor ion for a peptide, which represents the summation of ion volumes for the peptide in all fractions it was identified in, the actual result is a merged file that contains information on two or more precursor ions, each with only a fraction of the ion volume data. One of the peaks will be identified correctly, and the other will not be identified, or will be identified as a much lower scoring peptide. This can have great effects on the accuracy of protein quantification, as well as the possibility of increased false positive protein identifications. This is an issue that must be handled properly for experiments involving peptide or protein fractionation prior to LC-MS analysis.

The second problem relates to data accessibility, and can be an issue for any post-processing done with proprietary software, or open-source software that requires tools to interpret or access processed data. This software enables the exportation of low level peptide data, as well as protein data, from PLGS and from Expression Analysis, which may be used for an analysis or visualization performed in other statistical software packages, like R, for the production of heatmaps, and other statistical tests of data quality or probability of differential regulation.

TABLE OF CONTENTS

INTRODUCTION.....	56
PLGS strategy: blind merge.....	58
IntelliMerge strategy	59
EXPERIMENTAL.....	62
RESULTS.....	63
Additional features of DECO.....	64
REFERENCES.....	67

INTRODUCTION

The population of peptides that do not resolve to a single fraction can challenge the quantitative analysis of experiments utilizing peptide separation methods, like isoelectric focusing (IEF), hydrophilic interaction chromatography, or ion exchange ¹⁻³. Even with high-resolution methods, like OFFGEL IEF, at least 20% of peptides will be found in more than one fraction ^{2,4,5}. For the purpose of protein identification, the smearing of peptides is not a large problem, as long as the resolution enables appropriate complexity reduction.

Protein quantification can be adversely affected by peptides that do not resolve to a single fraction. For example, a protein sample is digested, and its peptides are separated into fractions. Each fraction is then analyzed by LC-MS^E. The abundance of a peptide that is only identified in one fraction is equal to the ion volume of that peptide from the one fraction it was identified. The abundance of a peptide that is identified in more than one fraction is equal to the sum of the ion volumes of that peptide from each of the fractions in which it was identified. For a peptide that is found in four adjacent fractions, post-processing software will need to recognize that peptide in each of the four fractions, and then sum the ion volume for that peptide in each of the four fractions (Figure 2-1a). The complete accounting of ion volume data for fractionated samples assures that the resulting merged data set will be quantitatively accurate.

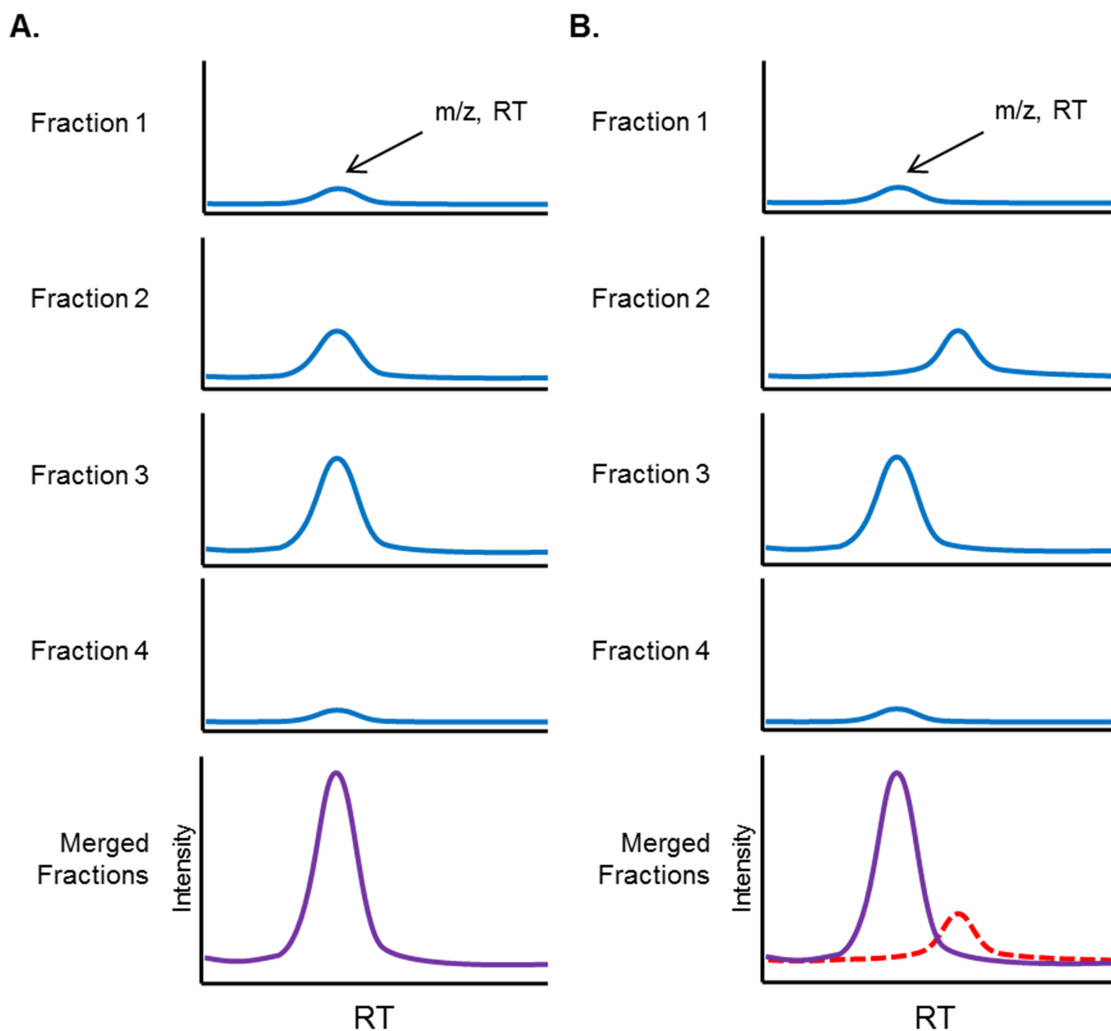


Figure 2-1. The PLGS (mergeFractions) method for merging data acquired from fractionated samples. (a) Under ideal conditions, this example peptide is identified in all four fractions with varying abundance (blue traces), and merging the data based on closeness in retention time and m/z enables a merged data set to properly represent the total ion volume of the peptide (purple trace). (b) Under real-world conditions, a peptide eluting at a different retention time in one fraction can evade merging, which results in the merged data set containing one peak that represents the peptide ion volume in three of the fractions (purple trace), and one peak (red dotted trace) that represents the peptide which fell outside the retention time window threshold. While the reduction in merged peptide ion volume may not seem significant here, in practice these effects can be much more dramatic and deleterious to quantification. Additionally, the extra precursor ion may be falsely identified.

PLGS strategy: blind merge

ProteinLynx Global Server (PLGS) is a software suite comprised of the tools necessary for the processing and database searching of MS^E data. PLGS handles fractionated sample data with a merge function (mergeFractions). mergeFractions merges processed data from all fractions to produce one XML file that contains all precursor ion and fragment ion data. To address smearing peptides, mergeFractions combines precursor ion data from different fractions when precursor ion masses are close enough, e.g., 7 ppm, and elute within a similar retention time window, e.g., 1 minute. For the example provided above, mergeFractions would produce an XML file with precursor ion data for peptides found in only one fraction and summed precursor ion data for any peptide found in multiple fractions. This merged XML file can then be searched in PLGS to produce a collection of peptides that reflect all peptides identified in the individual fractions, along with accurate ion volumes.

The mergeFractions strategy works well in an ideal world. In reality, there are many complications that can affect chromatography, and in turn, prevent like precursor ions from being properly merged (Figure 2-1b). Fractionation schemes rarely distribute peptides evenly across fractions; some fractions will have a much larger proportion of peptide species, and the complexity of the remaining fractions will range from medium to low. As complexity of peptides eluting from the column increases, the retention time of a given peptide can change, and in some cases this change can be large enough to prevent mergeFractions from correctly matching precursor ions for the same peptide between fractions. The failure to match these precursor ions results in a merged XML file that has more than one precursor ion for the same peptide, only one of which can be

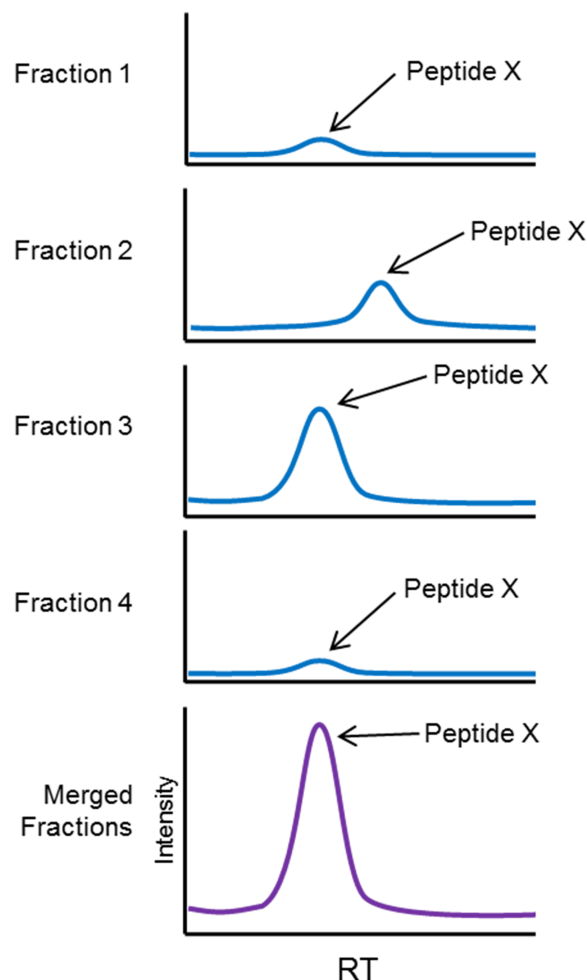
identified as the correct peptide. The remaining precursor ion may be incorrectly identified, or it may be unidentified. In either case, the correctly identified peptide will only have a partial ion volume that does not reflect its true abundance across the fractions. This will negatively affect protein quantification if the improperly merged peptide is one of the peptides used for quantification determination.

In addition to variations in fraction complexity, chromatographic resolution and reproducibility can also be affected by temperature fluctuations, contamination, solvent degradation, and pump calibration. For samples that are separated into many fractions, which can require days to weeks of instrument time for analysis, retention time fluctuations would be expected, and the consequences on peptide merging, and protein quantification, can be pronounced.

IntelliMerge strategy

In order to address the merging inaccuracies resulting from mergeFractions on real-world data, the aim of this work was to develop an approach that was independent of retention time to enable accuracy and experimental flexibility. The developed intelligent merge (intelliMerge) strategy bases the summed peptide ion volume on the ion volume of all precursor ions that were matched to the specified peptide in each of the fractions (Figure 2-2). While mergeFractions merges data prior to peptide identification, intelliMerge sums ion volumes after peptide identification, and uses peptide identity as the criterion for precursor ion matching and ion volume calculations.

Figure 2-2. The intelliMerge strategy uses peptide identification to match precursor ions for merging. Peptide X is identified in each of the fractions, and the ion volumes of that precursor ion are summed and applied to the Peptide X precursor ion found in the merged XML data.



The intelliMerge strategy is implemented in the Data Extraction and Correction (DECO) software package coded in Java. For a fractionated sample, each fraction is analyzed, and the RAW data files are searched and then merged. The merged file is searched, producing a collection of peptides encompassing those found in the individual fractions. In DECO, the PLGS project hierarchy can be accessed, and a merged file is selected for correction by the intelliMerge algorithm.

The intelliMerge algorithm parses the merged XML data file, identifying the source fractions that were merged, and storing all elements and attributes associated

with the peptides identified. Then, each of the source fraction XML files is parsed, and all elements and attributes associated with the peptides identified are stored. Next, the algorithm iterates through the array of peptides that were identified in the merged XML file, and for each peptide, it sums the ion volume for that peptide from each of the source fractions that it was identified, and stores this corrected volume in a separate array. Finally, DECO backs up the original merged XML file, and creates a new merged XML file, identical to the original, but switches in the corrected peptide ion volumes for each peptide. This allows for seamless integration into the PLGS pipeline allows for post-processing of the corrected file in Expression Analysis (Figure 2-3).

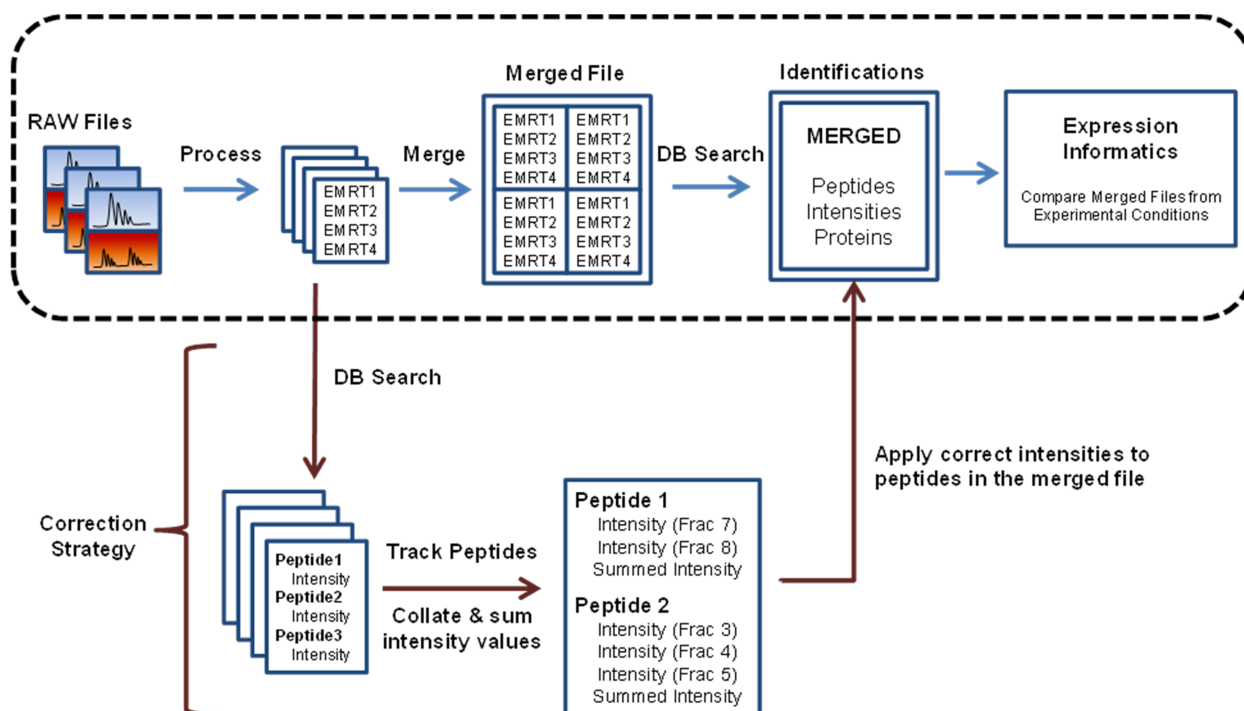


Figure 2-3. Correction of merged XML data in DECO with the intelliMerge algorithm. RAW files are processed in PLGS to produce processed XML files with precursor ion and fragment ion association data. These processed XML files can be merged and then searched, or they can be individually searched, resulting in peptide and protein identifications. intelliMerge collects all peptide data from the processed and searched fractions, sums ion volume data for any peptide that can be found in multiple fractions, and then applies these corrected values to the search results of the merged XML file. In this approach, PLGS can utilize the corrected merged XML file for downstream processing in Expression Informatics.

EXPERIMENTAL

To compare mergeFractions with intelliMerge we utilized a set of OFFGEL fractions acquired previously from the NAT8 lymphoblastoid cell line (See Chapter 6). Briefly, the non-irradiated human LCL lysate sample was processed, digested, and then separated into 12 fractions by OFFGEL isoelectric focusing. After fractionation, a specific quantity of MassPREP Enolase peptide digest was spiked into each fraction

prior to LC-MS^E analysis. Each fraction was analyzed in triplicate. The 12 RAW data files for each replicate were processed and searched in PLGS 2.5.

The processed files were merged with default parameters, and the merged file was searched and quantified. For intelliMerge corrected, the processed files were merged with strict tolerances (0.1 ppm mass, and 0.1 min retention time), searched, and processed in DECO by the intelliMerge algorithm.

RESULTS

The ion volumes of the top three ionizing peptides for enolase (TAGIQIVADDLTVTNPK, NVNDVIAPAFVK, and VNQIGTLSESIK) were extracted from the non-corrected (mergeFraction) merged file and the corrected (intelliMerge) merged file for each technical replicate (Figure 2-3). Results from the non-corrected file are problematic, with large inconsistencies in summed peptide ion volumes across the technical replicates that indicate improper precursor ion matching between fractions. After intelliMerge correction, the summed peptide ion volume data are consistent in all three replicates. In this example, even though enolase peptides were present at high levels in each fraction, mergeFractions was insufficient to properly merge the data, which highlights a major concern for protein quantification based on fractionated data when processed in the current PLGS pipeline.

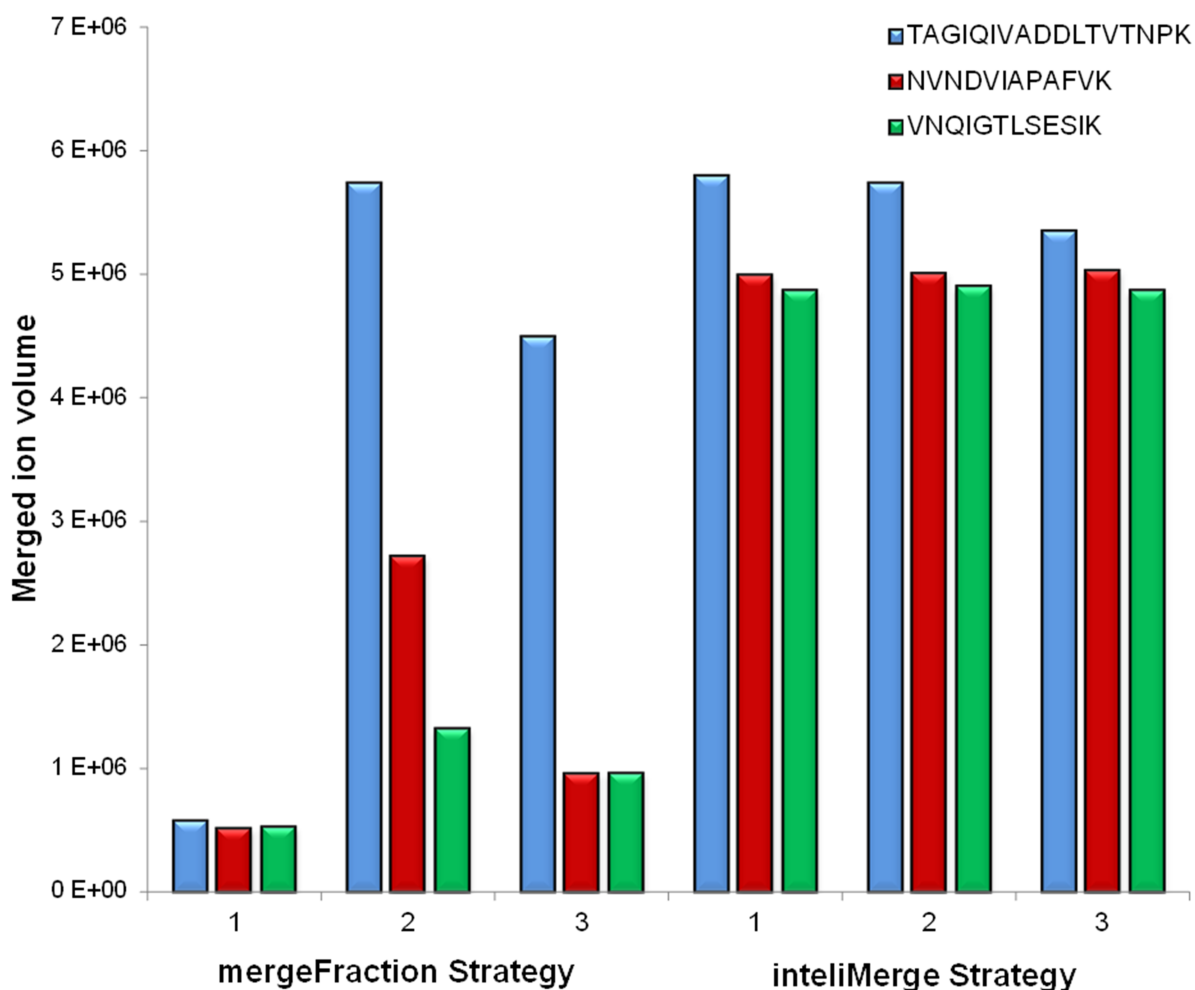


Figure 2-3. The intelliMerge strategy affords increased quantification accuracy in fractionation experiments. Merged file peptide ion volumes (Merged Ion Volume) for the top three ionizing peptides of enolase, in three technical replicate merged files that were not corrected (first three sets of columns), or were corrected by the intelliMerge algorithm (last three sets of columns).

Additional features of DECO

PLGS is protein-centric in its presentation of data to the user, effectively limiting the level of access that a user has to processed data, and provides rudimentary tools for data export. This poses two distinct challenges for data analysis. The first is the inability to access lower peptide or ion level data for comparative analysis. The second is the inability to perform batch operations for data retrieval, which is a major hindrance

when data must be retrieved from a large experiment. The DECO software package addresses these issues by providing tools for the extraction and collation of query, peptide, and protein data from data-dependent and data-independent experiments that are processed in PLGS. Batch features that can reduce data analysis time tremendously, and reduce human error involved with manual retrieval of data from PLGS, complement the increased access to experimental data. These features enable the analysis of experimental data in other software suites for statistical analysis and visualization, e.g., R, SPSS, Excel, and allow for filtering of large data sets by peptide attributes.

DECO also maintains quantitative tools that can be used in place of PLGS merging for protein-fractionated samples. In GeLC-MS^E experiments, merging of the data from each gel slice will complicate the search and quantification of the merged data file. Instead, DECO can be used to collate all proteins identified in all gel slices from a specific lane, and sum the abundances of any protein found in multiple gel slices. This data set is then exported into a comma separated value file that may be viewed and filtered in many software packages, e.g., Excel. This feature was used in the analysis of *Chlamydomonas reinhardtii*. (“Precious Metals: The Proteome of Copper, Iron, Zinc, and Manganese Micronutrient Deficiency in *Chlamydomonas reinhardtii*”, MCP, in review)

DECO can also calculate physiochemical properties of peptides and proteins from analyses. Grand Average of Hydropathicity index (GRAVY) value represents the sum of hydropathy values, based on the Kyte-Doolittle scale, for all amino acids in a protein, divided by the amino acid length ⁶. Isoelectric points are calculated by brute

force calculation of the net charge, based on the pKa values associated with the constituent amino acids, at pH values from 0 to 12.

REFERENCES

1. Manadas, B., Mendes, V. M., English, J. & Dunn, M. J. Peptide fractionation in proteomics approaches. *Expert Rev Proteomics* **7**, 655–663 (2010).
2. Fraterman, S., Zeiger, U., Khurana, T. S., Rubinstein, N. A. & Wilm, M. Combination of peptide OFFGEL fractionation and label-free quantitation facilitated proteomics profiling of extraocular muscle. *Proteomics* **7**, 3404–3416 (2007).
3. Cargile, B. J., Bundy, J. L., Freeman, T. W. & Stephenson, J. L. Gel Based Isoelectric Focusing of Peptides and the Utility of Isoelectric Point in Protein Identification. *J Proteome Res* **3**, 112–119 (2004).
4. Chenau, J., Michelland, S., Sidibe, J. & Seve, M. Peptides OFFGEL electrophoresis: a suitable pre-analytical step for complex eukaryotic samples fractionation compatible with quantitative iTRAQ labeling. *Proteome Sci* **6**, 9 (2008).
5. Hubner, N. C., Ren, S. & Mann, M. Peptide separation with immobilized pl strips is an attractive alternative to in-gel protein digestion for proteome analysis. *Proteomics* **8**, 4862–4872 (2008).
6. Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J Mol Biol* **157**, 105–132 (1982).

CHAPTER 3

Quantitative proteomic characterization of the para-crystalline sheath encapsulating *Methanospirillum hungatei* strain JF1: identity of the major structural components.

ABSTRACT

M. hungatei cells possess a protein sheath-like structure that encapsulates the linearly arrayed individual cells lined up end to end. The identity and properties of the protein(s) making up the sheath are unknown despite a wealth of ultra-structural information. Prior sheath disassembly studies have relied on extremely harsh dissolution techniques to characterize the polypeptide content of the proteinaceous paracrystalline sheath. Due to the nature of the techniques, the MW estimates by LC and SDS-PAGE are not consistent between studies. In this work we use quantitative mass spectrometry to identify protein components of the *M. hungatei* sheath based on tryptic peptides. Utilizing peptides to infer proteins obviates problems associated with partial protein degradation and allows for the determination of accurate MWs of sheath components, and, more importantly, sequence information and component abundance.

The sheath of *M. hungatei* JF1 was solubilized and digested by eFASP, a harsh mass-spectrometry compatible method, and analyzed in quadruplicate by LC-MS^E. Sixty-two proteins were identified in at least two replicate experiments based on over 1,600 peptide identifications. Quantitative proteomic analysis revealed a 40.7 kDa polypeptide (Uniprot Q2FRN9) encoded by Mhun_2271 as the predominant component

of the sheath by about 10 fold. Lower abundant sheath fraction components were also identified including several hypothetical proteins, proteins with energy generation roles, and several ribosomal proteins. Their structural and/or functional significance is discussed regarding the stability of the sheath and its function.

TABLE OF CONTENTS

INTRODUCTION.....	70
EXPERIMENTAL.....	73
Strains and cell growth	73
Enhanced filter aided sample preparation of <i>M. hungatei</i> sheath proteins	73
Analysis by nanoLC-MS ^E	75
Data processing and protein identification	76
RESULTS.....	77
The Mhun_2263 protein.....	78
The PKD protein (Q2FPM0)	78
Minor abundance polypeptides detected in the sheath fraction	78
Amino acid analysis of the <i>M. hungatei</i> JF1 and GP1 sheath proteins	79
DISCUSSION.....	81
REFERENCES.....	92

INTRODUCTION

Archaea can be found across all environments, from extreme hydrothermal vents and salt brines, to more temperate oceans and soils, and to the colonization of the mammalian gut ¹. Methanogenic archaea produce an estimated 1 billion tons of methane per year ².

Methanospirillum hungatei is an anaerobic methanogenic archaean important for the degradation of biomass and the recycling of global carbon ³. It can metabolize hydrogen, carbon dioxide, formate, and acetate for energy production, anabolic reactions, and methane formation ⁴. *M. hungatei* has the unique ability to metabolically interact with syntrophic anaerobic bacteria, enabling the thermodynamically unfavorable breakdown of fatty acids, alcohols, and aromatics, by consuming the bacterially produced hydrogen and formate byproducts. Interest in *M. hungatei* has been renewed following the recent discovery of energy storage structures located at opposing sides of the cell, which may prove to be a precursor to be an early form of polyphosphate energy storage ⁵.

M. hungatei cells are encased in a single layered S-layer cell wall, and grow end-to-end, about 10 cells long, within an open-ended cylindrical sheath. Adjacent cells are separated by terminal plug structures composed of many layers of carrying permeability. Together, these structures enable resistance to turgor pressure and osmotic lysis, restrict passage to nutrients and wastes, and maintain the rod shape of each cell ⁶⁻⁸.

The paracrystalline sheath is a hollow tube that exhibits stability in the presence of strong detergents and can withstand extreme pressures due to a cross-beta structure

and disulfide bond networks ^{9,10}. Electron diffraction and electron microscopy studies reveal that the subunits of the sheath contain polypeptides arranged in cross-beta conformations, which maximizes protein bonding, and that intra-hoop subunit bonds are stronger than inter-hoop bonds connecting adjacent hoops. These findings explain sheath strength in a circumferential direction, and flexibility for longitudinal elongation, parallel to the sheath axis, during growth ¹¹.

While the sheath maintains integrity in the presence of strong denaturing conditions, it is susceptible to disulfide bond breaking. When exposed to disulfide reducing agents, e.g., β -ME or DTT, the sheath decompiles perpendicular to its axis, yielding a series of rings or hoops, implying that disulfide bonds are responsible for connecting the series of hoops, and play a role in sheath stability. The involvement of cysteine in structural integrity of the sheath is interesting because the *M. hungatei* sheath has a very low proportion of cysteine residues, approximately 9% of total cellular cysteine, and organosulfur groups are not present in phenol-solubilized sheath proteins, and most S-layers are devoid of cysteine ^{2,9,12,13}. The cysteine components of the sheath are thought to be involved in two functions: the covalent bonding of sheath components, and the selective cleavage points to accommodate future growth or pressure buildup ^{3,14}.

The sheath can act as a pressure regulator during metabolism. The dense protein lattice that composes the sheath contains inter-subunit pores that are just large enough to allow the diffusion of H₂, CO₂, and CH₄, formate and acetate ^{4,10}. During metabolism, and the production of CH₄, the structural strength of the sheath can allow the development of internal pressures reaching 300-400 atm ^{5,10}. As the pressure on

the sheath increases, it is likely that the hoop-junctions expand, resulting in longitudinal stretching of the sheath, and the increase in pore diameter that would allow for easier entry of nutrients and exit of waste products ^{6-8,10}.

A few groups have proposed models for the remarkable stability and structure of the sheath, based on characterization of sheath components ^{9,10,12,14,15}. Despite structural and amino acid composition consistency between studies, there is a paucity of concordant data for the molecular weights of the sheath components. The discrepancy between studies has been ascribed to the harsh nature of sample preparation, which could cause protein degradation and artifacts in the SDS-PAGE gels used for determination of polypeptide components. To make the matter more complex, a later study presented a new group of phenol-soluble polypeptides in the sheath ¹⁵, which would not have been present in the SDS-PAGE gels. Therefore, the molecular weights of the major polypeptide components of the sheath are still contentious, and, more importantly, there is no study that identifies the sequences of those components.

The goal of this study was to query the predicted proteome for *M. hungatei* strain JF-1, based on the genome sequence published in 2006, to comprehensively analyze the proteinaceous components of the sheath. This was accomplished using a harsh mass spectrometry-compatible sample preparation method, eFASP, coupled with a label-free quantitative proteomic workflow to identify and quantify the proteins in the sheath. Bioinformatic analyses of the peptide and protein sequences present a view of the isoelectric points and hydrophobicities of sheath components. The amino acid composition of the sheath, and the molar ratios between amino acids, show a good deal of concordance with previous studies, and highlight the inaccuracy in calculations when

analysis is unable to differentiate between asparagine and aspartic acid, and glutamine and glutamic acid. We present a catalog of sheath components, with the major constituents being Mhun_2271 and Mhun_2263, and lesser abundant proteins involved in chemotaxis, nutrient transport, ribosome and hypothetical functions.

EXPERIMENTAL

Strains and cell growth

Methanospirillum hungatei strain JF1 (DSM 864) was cultured in 250 ml serum bottles containing 50 ml of medium as previously described ⁵. Cells were harvested by centrifugation at 7.5k rpm for 10 minutes. The spent cell growth medium was decanted and the cell pellet was re-suspended in 1mL TBS buffer (150 mM NaCl, 50 mM Tris, pH=8.0), and washed twice in TBS buffer with centrifugation at 13k rpm. Cell pellets were either frozen at -20°C or used immediately for cell sheath preparation. *The M. hungatei* strain JF1 sheath fraction was prepared by CHAPS extraction.

Enhanced filter aided sample preparation of M. hungatei sheath proteins

M. hungatei sheath (100 µg) was dried down in a SpeedVac and suspended in Solubilization Buffer A (4% ammonium dodecyl sulfate, 16 mM TCEP, 200 mM Tris-HCL, pH 8), and incubated on a heated block for 10 minutes at 90°C. The sample was sonicated for three 10-second intervals with a Sonic Dismembrator Model 100 (Fisher Scientific), and then spun at 16,000 rcf for 10 minutes in an Eppendorf centrifuge; this step was repeated once. The supernatant and any residual pellet, was sonicated and allowed to cool to 37°C. Proteins were alkylated by the addition of 4-vinylpyridine (4-

VP) at 16 mM final concentration, and incubated for 1 hour on a shaker at 37°C. DTT was added to a final concentration of 40 mM to deactivate residual 4-VP.

Eight volumes of Solubilization Buffer B (8 M urea, 0.2% deoxycholic acid, 100 mM Tris-HCL, pH 8) was added to the digest to begin buffer exchange of SDS. The diluted sample was applied to a Passivated Microcon Filter Unit (described below), and spun at 14,000rcf for 10 minutes. The eluate was discarded, and 200 µl of Solubilization Buffer B was added to the filter unit and then spun at 14,000 rcf for 10 minutes. This step was repeated twice. To remove urea, three buffer exchange steps were performed with Digestion Buffer (0.2% deoxycholic acid in 50 mM ABC). The Passivated Microcon Filter Unit (30,000 Da cutoff), and collection tube, were pre-incubated in 5% Tween-20 overnight, and then thoroughly rinsed with large volumes of MS-grade water.

Digestion Buffer (100 µl) was added to the Microcon unit, followed by the addition of 5 µg of trypsin (1 µg trypsin: 20 µg protein). Digestion proceeded for 12 hours on a shaker set to 37°C. The Microcon unit was transferred to a new Passivated Collection Tube, and spun at 14,000 rcf for 10 minutes. To assure complete removal of peptides, 50 µl of 50 mM ABC buffer was added to the Microcon unit, and collected by centrifugation, two times.

Removal of deoxycholic acid from the digest was accomplished by acidification and phase transfer of insoluble deoxycholate to an ethyl acetate layer ¹⁶. Instead of removing the aqueous layer as in Masuda et al. ¹⁶, the organic layer was removed, and then two more rounds of addition and removal of ethyl acetate facilitate the removal of deoxycholate from the solution, similar to the strategy Yeung et al. employ for the removal of n-octylglucoside ¹⁷.

Collected peptides were dried down in a SpeedVac and re-suspended in 50% methanol, three times. Dried peptides were suspended in modified Mobile Phase A (98.5% H₂O/1% ACN/0.5% formic acid), to which MassPREP Glycogen Phosphorylase B digestion standard was spiked in to a final concentration of 4 fmol/μl.

Analysis by nanoLC-MS^E

nanoUPLC of sample peptides was performed on an Waters nanoACQUITY UPLC system equipped with a nanoACQUITY UPLC Symmetry C18 180 μm x 20 mm trap column (Waters, 186003514), and a nanoACQUITY UPLC BEH C18 75 μm x 150 mm reversed phase analytical column (Waters, 186003543). The aqueous mobile phase (Mobile Phase A) contained 0.1% formic acid in water and the organic mobile phase (Mobile Phase B) contained 0.1% formic acid in acetonitrile. 5 μl of the sample was loaded onto the trap with and washed with 99% A and 1% B for 5 minutes, and then injected onto the column with 3% Mobile Phase B. Peptides were eluted from the column with a gradient of 3-40% B over 90 minutes at 300 nl/min, followed by a 5 minute rinse with 95% B. The column was re-equilibrated to 97% A for 20 minutes.

The lock mass compound, [Glu¹]fibrinopeptide (0.5 pmol/μl, 25% ACN, 0.1% formic acid), and the analyte were delivered at 300 nl/min to the NanoLockSpray dual electrospray ion source (Waters, UK) which was equipped with a microcapillary for the lockmass compound, and a PicoTip Emitter (New Objective) for the analyte. Two electrosprays are produced which may be selectively sampled by positioning of the baffle.

The ion source was interfaced with a Waters Synapt HDMS for mass spectrometric analysis. Acquisition was performed in positive nanoESI mode and the instrument was operated in V-mode with an average resolution of 9500. Accurate mass data for precursor and fragment ions were collected over the 100-2000 m/z range, with the mode of acquisition alternating between low and elevated energy at 1-second intervals, for the duration of analysis. The lock mass channel was sampled every 30 seconds for calibration of MS and MS^E data.

Data processing and protein identification

Raw data files were processed in ProteinLynx Global Server 2.5.2 (PLGS, Waters) to provide an inventory of precursor ions along with their respective fragment ions, to search against the Uniprot HAMAP *M. hungatei* JF1 database (strain JF-1 / DSM 864, August 2011) for peptide and protein identifications¹⁸. Database search settings included two missed cleavages, a 4% false positive rate, fixed S-pyridylethyl modification of cysteine residues, and variable oxidation modification of methionine residues. The Waters Expression Informatics component of the PLGS software was utilized to quantify data using glycogen phosphorylase B (Uniprot accession P00489) as the quantification reference (20 fmol)¹⁹⁻²¹. Data Extraction and Correction (DECO) software was used to calculate isoelectric points and grand average of hydropathicity index (GRAVY) values for peptides and proteins identified. GRAVY values were calculated based on the hydropathicity of constituent amino acids derived from Kyte and Doolittle²².

RESULTS

M. hungatei strain JF1 and strain GP1 sheath disassembly studies have previously relied on extremely harsh dissolution techniques to characterize the polypeptide content and the amino acid content of the *M. hungatei* sheath component(s). Due to the nature of these techniques, molecular weight estimates by LC and SDS-PAGE may have been skewed due to protein degradation. In this work we employed quantitative mass spectrometry to identify protein components of the *M. hungatei* JF1 sheath based on identification of tryptic peptides. This obviates the inherent problems associated with partial protein degradation and allows us to present an accurate assessment of sheath components based on their pre-processed sheath protein sequences. In addition to sequence information, protein abundance values were also generated.

To perform this analysis, *M. hungatei* JF1 sheath was solubilized and digested by our developed eFASP protocol (Experimental Procedures), yielding clean spectra and elution peaks (See Figure 3-1). Quadruplicate injections of the digested sheath sample, spiked with glycogen phosphorylase b for quantification reference, yielded approximately 70 protein identifications, based on 1,600 identified peptides. Quantitative analysis reveals one predominant polypeptide, indexed in Uniprot as accession Q2FRN9 encoded by Mhun_2271 (See Figure 3-2). This protein is annotated as a hypothetical protein in *M. hungatei* JF1. The Q2FRN9 protein had an on-column abundance of 51 fmol, a level nearly ten times greater than the second most abundant polypeptide, Q2FTS3 (7.8 fmol) encoded by Mhun_2263. Over 67% sequence coverage for Mhun_2271 protein was achieved in all replicate injections (See Figure 3-

3). With a predicted molecular weight of 40.7 kDa, and an isoelectric point 4.4, this protein is consistent with prior studies that identified a 40 kDa protein as the main constituent of the sheath ²³, that was present as a highly acidic amino acid composition for the sheath ^{12,24,25}. The Mhun_2271 gene product contains a predicted 23 amino acid signal peptide sequence for protein export and processing, consistent with its observed cellular location.

The Mhun_2263 protein

The second most abundant polypeptide detected, Q2FTS3, is encoded by Mhun_2263 and is annotated as a hypothetical protein with a predicted molecular weight of 93.2 kDa, and an isoelectric point of 4.5. Coverage was over 67% for Mhun_2263 across all replicate injections (See Figure 3-4). Like Mhun_2271, this second most abundant sheath component protein possesses an N-terminal SP sequence for cell export.

The PKD protein (Q2FPM0)

The third most abundantly detected protein was Mhun_0425 that comprised approximately 4.11% of the sample (Table 3-1).

Minor abundance polypeptides detected in the sheath fraction

A number of additional proteins were detected at low abundance relative to the major sheath component, Mhun_2271 (see Table 3-1). These included the following: Five protein subunits of the tetrahydromethanopterin S-methyltransferase enzyme

required for methyl transfer to HS-coenzyme M (subunits A through F that include mtrA, mtrE, Mhun_2169, Mhun_2170, Mhun_2171), one of the methyl-coenzyme M reductase subunits (i.e., the alpha subunit, Q2FSN0), the H⁺-transporting two-sector ATPase C subunit (Q2FNK4), and the formate/nitrite transporter (Q2FKY3) for formate uptake from the cell environment. Also detected was the ABC transporter, substrate-binding protein, aliphatic sulfonates (Q2FR15) and the PKD proteins (Q2FPM0, Q2FPL5, Q2FLP4, Q2FRU4) PKDs of unknown functions. The flagellin protein (Q2FUM4) and methyl-accepting chemotaxis sensory transducer (Q2FU40) were less abundant along with a number of ribosomal proteins with a relative abundance ranging from 1.1% (30S S8e) to 0.17% (50S L30P). Finally, a number of putative uncharacterized proteins were detected (Table 3-1).

The above observation of low abundance polypeptides in the sheath preparation suggests a minor contamination of the *M. hungatei* sheath by a relatively small amount of intact membrane fragments/vesicles due to the procedure used (Experimental Procedures).

Amino acid analysis of the M. hungatei JF1 and GP1 sheath proteins

Previous studies of the *M. hungatei* sheaths of strains JF1 and GP1 have described a high prevalence of acidic amino acids. We wanted to determine the closeness of the amino acid composition determined by quantitative mass spectrometry to previous studies employing chemical methods, e.g., acid hydrolysis, performic acid oxidation, and spectrophotometry^{12,24,25}. Amino acid composition for the identified proteins, weighted by abundance, was calculated as mole percent for comparison with

previous studies of *M. hungatei* JF1 and GP1 (Table 3-1). These studies highlight a prevalence of acidic amino acids, in strain JF-1, basic amino acids are outnumbered by about 4.37:1²⁴ and 3.61:1¹², and in strain GP-1, basic amino acids are outnumbered 2.87:1¹² and 4.37:1²⁵. Variations between the findings were ascribed to different methods for sheath preparation and cell spacer contamination, although slightly different methods were used for amino acid analysis. Amino acid determination by acid hydrolysis can yield problematic results due to the inability to differentiate between Asn/Asp and Gln/Glu, the oxidation of cysteine and tryptophan sidechains, and the overlap of cysteine and proline peaks.

Calculation of the ratio between acidic and basic amino acids based on protein abundance results in 1.75:1 when we mimic the previous calculations by grouping Asn with Asp, and Glu with Gln (Table 3-2). Since our proteomic analysis allows the differentiation between Glu and Gln, and Asp and Asn, the actual ratio of acidic to basic residues is 0.95:1. This calculation was not be ideal, since they are based on the theoretical protein sequences, rather than the mature sequence that is present.

In an attempt to address this, the same calculation was performed based on the peptide sequences identified, and their absolute ion volume. The result shifts both ratios in favor of acidic residues: 1.99:1 when Asp/Asn and Glu/Gln are grouped and 1.08:1 without grouping (Table 3-2). The near doubling of the ratio, by either method of calculation, when amino acids are grouped, indicates that previous studies overestimated the acidic composition of the sheath proteins significantly due to the inability to differentiate between Asp/Asn and Glu/Gln.

Amino acid ratios have also been used to differentiate methanogens from bacteria. The molar ratio of Lys:Ala:Glu in archaea was found to be 1:1.2:2, whereas in peptidoglycan containing bacteria it is 1:2:1²⁴. Our data compares favorably, with a Lys:Ala:Glu ratio of 1:1.2:1.6 (calculated based on protein data) or 1.0:1.5:2.0 (calculated based on peptide data) for the *M. hungatei* JF1 sheath (Table 3-2).

Additionally, we further characterize the physiochemical *M. hungatei* JF1 sheath proteins and identified peptides. All identified peptides were extracted using Data Extraction and Correction (DECO), and their isoelectric points and GRAVY values were calculated (See Figures 5 and 6). All identified protein sequences were extracted using DECO and their isoelectric points and GRAVY values were calculated (See Figures 7 and 8).

DISCUSSION

We used a gentle sheath extraction method, improved protein preparation methods and the newly available genome sequences for *M. hungatei*, to provide the first evidence for the identity and the quantity of the protein components of the sheath. Mhun_2271 is the primary sheath component, currently annotated as a hypothetical protein, which may now be analyzed for its role in cell structure and morphology. There are 8 paralogous Mhun_2271 proteins in *M. hungatei* JF1 plus a small group of orthologs (~15) in the bacteria (plus a few metagenomic methanogens). While there was little/no apparent relationship of Mhun_2271 to the bacterial orthologs, there were interestingly no major related proteins in any other archaea based on available genomic data (as of 07 29 12 IMG JGI searches).

If three detected proteins (Mhun_2271, 2762, 0425) are present, this gives a molar ratio of (12.4/1.89/1). If only two detected proteins were present (Mhun_2271, 2762) this would give a molar ratio of (6.571). The last clear possibility would be that Mhun_2271 is the sole sheath protein assuming that the methods used did not under-represent proteins detected.

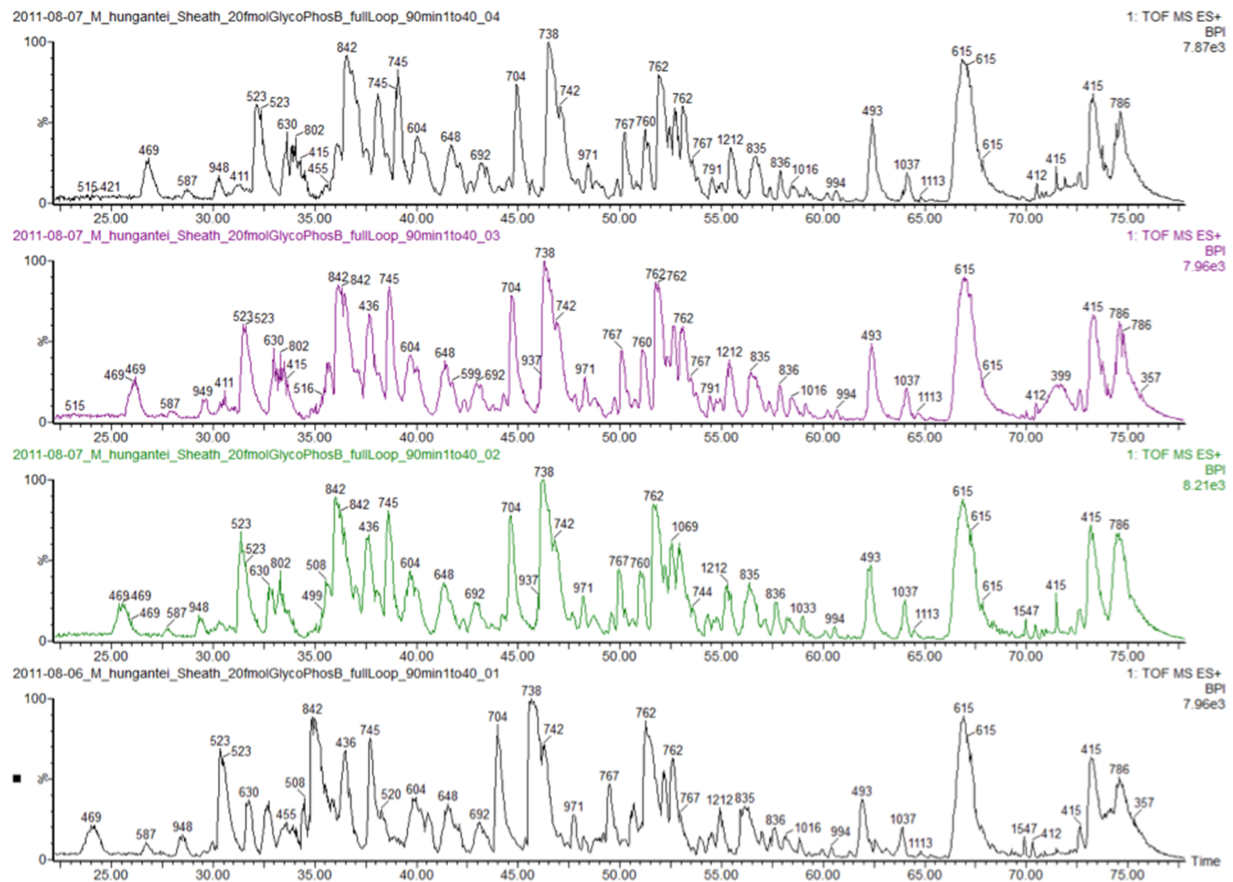


Figure 3-1. eFASP allows for the efficient solubilization and digestion of proteins from the sheath of *M. hungatei* JF-1. Base peak intensity chromatograms for all four technical replicates are displayed.

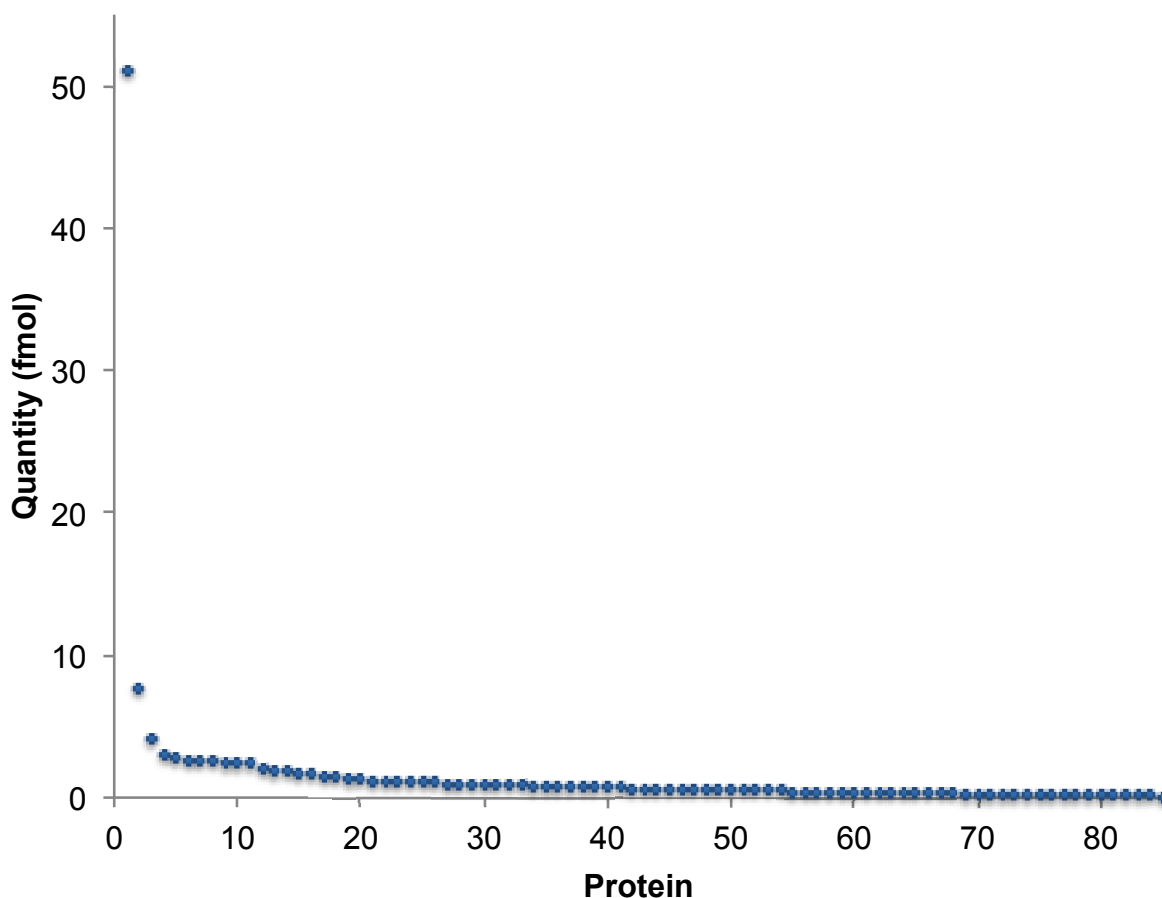


Figure 3-2. Absolute quantification of *M. hungatei* JF1 sheath components. The top 3 ionizing tryptic peptides (Hi3) for each protein was compared to the Hi3 for the spiked in standard to determine the molar quantity. Proteins are plotted in order of descending abundance.

1	MKFGYAIVLA	LVALVAVTGF	ASADRLPQQV	PENQIFTIDT	LIDVTGAVSQ
51	ESEMQWTIDS	QNSVAEGVIF	DNNNFDALAG	NAVVLTPAQL	NYLQTKATEW
101	ANLYGNAASD	YLSYKNIGGV	FYLTCLNVPV	DVNTNYMDPE	LFGSTTYANY
151	LAYIEANFNY	ADSTVKIPGA	IHDSKLYAGE	EIAILTWTDT	LRSNGGKLAL
201	NENIDFDSRN	KGKGLNNLEV	EKVLTYASTE	GAHLVGAEW	TLDVAGNWEK
251	TADSIRCVFA	AASTEYFPAF	CNVVKAKSEL	VNINSAQVST	KGAARAVAAT
301	GDIPAALNYQ	IAVTPDSNSG	SGFADGTVKT	LFGGSIMEAR	GKNQVPSATN
351	NWKDTASVTG	GIKNFQKTFN	YESGFTF		

Figure 3-3. Protein sequence coverage map of the most abundant detected sheath protein Mhun_2271 (Q2FRN9). The sequence coverage from each of the technical replicates was applied to the presented protein sequence for this 40.7 kDa protein. Residues in black were not identified in any of the replicate injections. The minimum sequence coverage was 67%.

1	MNAKFAVVM	VLVALALVAF	PASAAINKIP	AGGEVFLGEK	GLDVSAATGG
51	SSQIAWWQPG	TNTETECPAD	IQQVTNAQNF	YVSPDIFVGK	TGNWYQWNGS
101	VKGPLAFNIK	EPSLNLKIWD	GSVNEDVTGK	AIPVGNYGNF	VVETNMQSII
151	TRPGYQPADA	PFKIKVKSAD	GGVYTNLVGN	NGKEISLTNL	AVNQQLWYVW
201	SPDNKHTEPA	SNDGWNTAAE	DKYGNRMYKA	GTYYTVWAE	ANGMKDQYTA
251	PDGSDYTGKT	ISAVKSVQIA	TDQVKIEANK	DTVVRGNPFS	VTITGVPNAE
301	YYVWAKGTGS	MTGQPDDQPP	MILDAQVDVK	QDAAAGPYEI	GKYQYEGGAG
351	KSIKQDVPDD	PEFHGTFKYA	LVKLNSSGTR	TVEWKTSKDT	KDKKYTIRVE
401	RKSGSQYKSD	EVDVKVEKGD	VTIVASGDQS	YYLGEEVKLS	GTNSETDVTY
451	LYITGPNLPT	RGGMLKSPRK	EVNNDQPASF	DQADVQDDDT	YEFKWQTANL
501	EIDAGTYTVY	AVSAPRDKDH	LSDAQYDTVS	LVIKKPYIQA	TASASVAKG
551	DKMYLRGTAE	GDPSKGVAIW	ILGKNKVLYA	TESVNDDMSF	EHEIKSATT
601	SLYAGQYFVV	VQHMYNNEF	DVYPDNPLEP	QYVYGSYPTR	NSQVFKLGGS
651	GALQGTDAAE	ALIQAINSPM	VDDTYTKMQF	LVEEPKITIN	PIGEQAVGSK
701	FEIAGTTNLM	YDDNDLLVEV	TSSSFKPTDK	SQSGEFSGAT	GTVKVQQGTD
751	GLNKWAFNVD	ASTFKPDEYI	VRVSGVTTNT	VETALFNVVE	AGAIPTVAPT
801	VAPTEAPVVN	ETPTPEPTTV	PPTPEPTTVP	PTPEPTQQSP	GFGALVALAG
851	LGAVAFFVMR	RN			

Figure 3-4. Coverage map of the second most abundant protein, Mhun_2263 (Q2FTS3). The sequence coverage from each of the replicates was applied to the presented protein sequence for this 93.1 kDa protein. Residues in black were not identified in any of the replicate injections. The minimum sequence coverage was 63%.

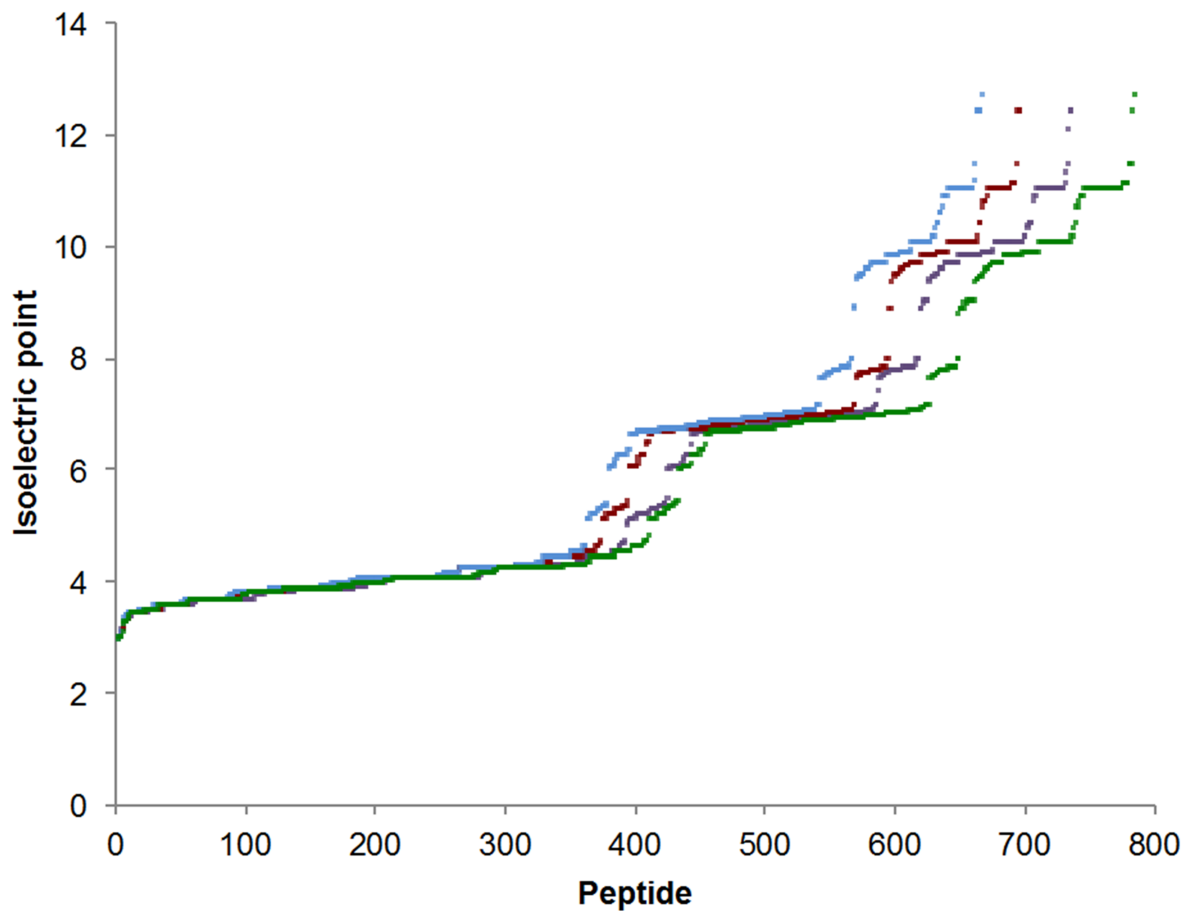


Figure 3-5. Peptide isoelectric point distribution of *M. hungatei* JF1 sheath proteins. The isoelectric point for all tryptic peptides identified in the four technical replicate analyses was calculated and plotted. Replicate 1 (blue) replicate 2 (red), replicate 3 (purple), and replicate 4 (green).

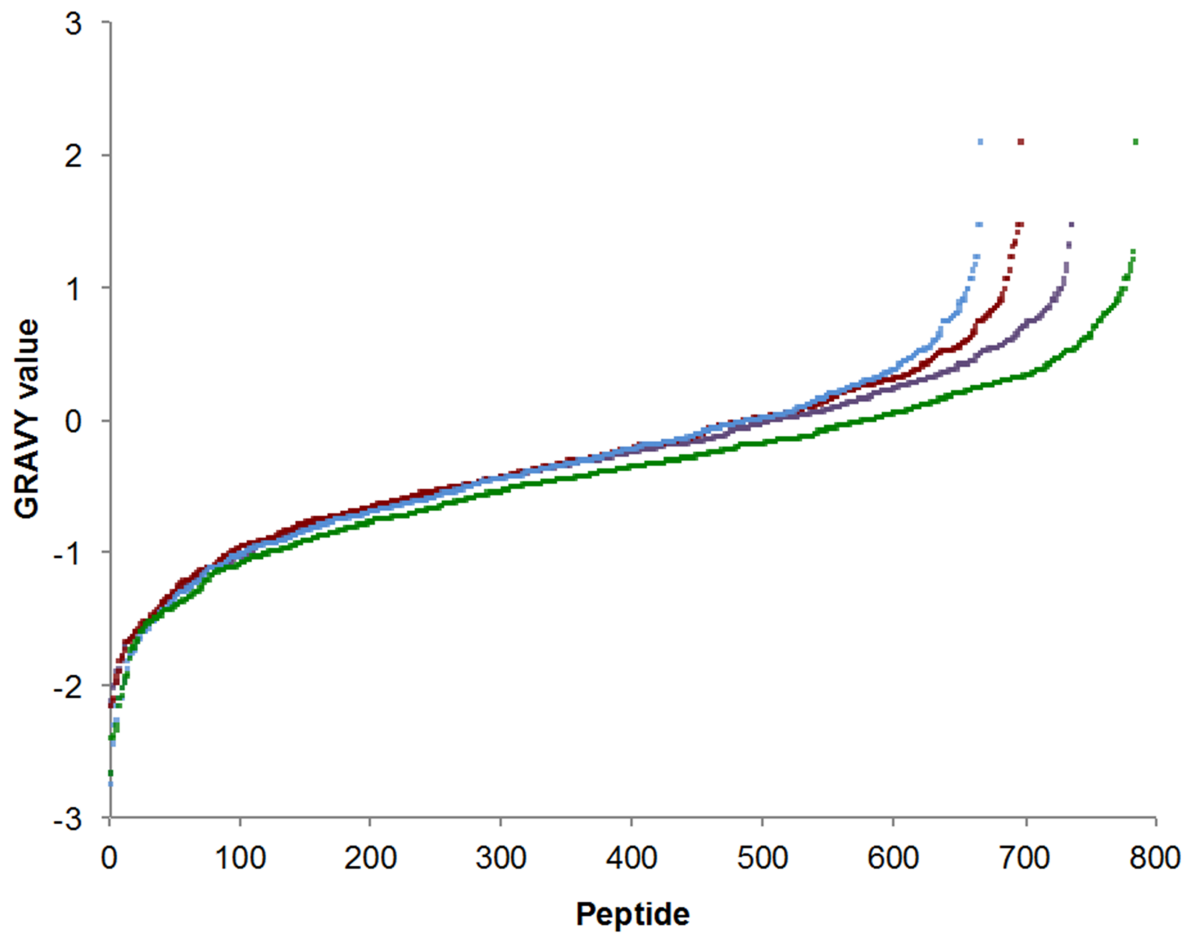


Figure 3-6. Peptide GRAVY value distribution of *M. hungatei* JF1 sheath proteins. The GRAVY values of all tryptic peptides identified in the four technical replicate analyses was calculated based on hydropathicity values for each amino acid, and plotted. Replicate 1 (blue) replicate 2 (red), replicate 3 (purple), and replicate 4 (green).

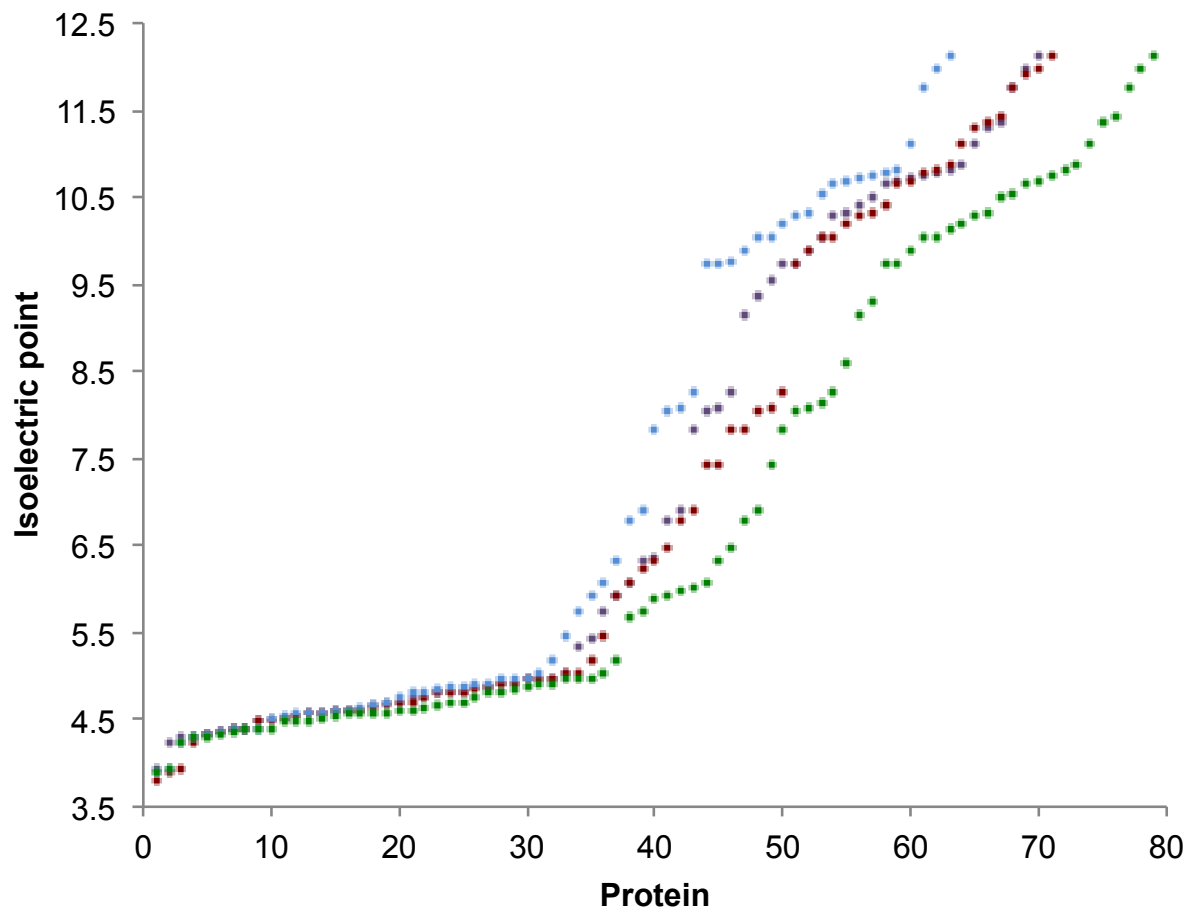


Figure 3-7. Protein isoelectric point distribution of *M. hungatei* JF1 sheath proteins. The isoelectric point of each of the protein sequences identified in the four technical replicate analyses was calculated and plotted. Replicate 1 (blue) replicate 2 (red), replicate 3 (purple), and replicate 4 (green).

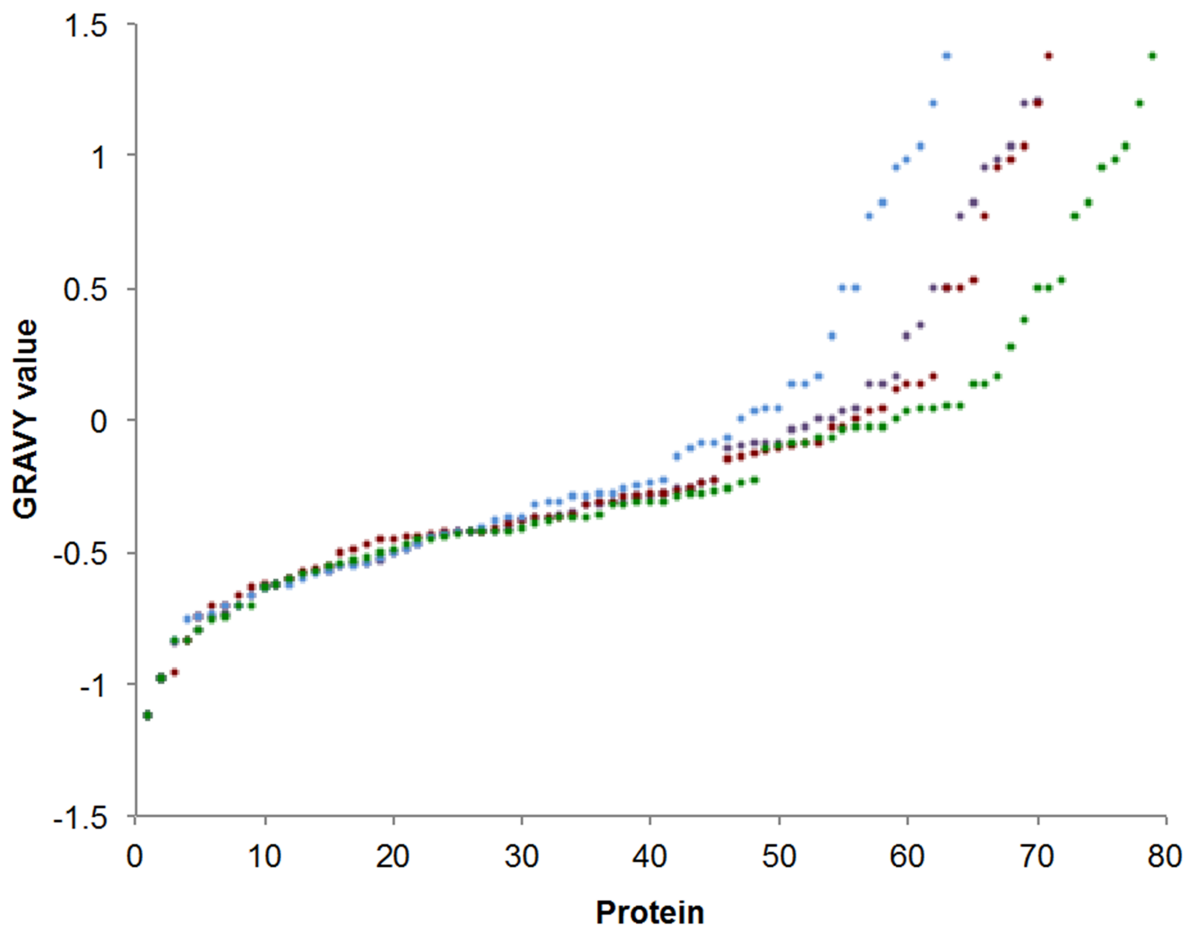


Figure 3-8. Protein GRAVY value distribution of *M. hungatei* JF1 sheath proteins. The GRAVY value of each of the protein sequences identified in the four technical replicate analyses was calculated and plotted. Replicate 1 (blue) replicate 2 (red), replicate 3 (purple), and replicate 4 (green).

ID	Description	Gene	Quantity, fmole		MW	GRAVY	pI
			Avg	SD			
Q2FRN9	Putative uncharacterized protein	Mhun_2271	51.02	8.24	47,253	-0.09	4.4
Q2FTS3	Putative uncharacterized protein	Mhun_2263	7.76	0.79	108,640	-0.41	4.5
Q2FPM0	PKD	Mhun_0425	4.11	0.51	157,217	-0.28	4.4
Q2FRN3	Tetrahydromethanopterin S-methyltransferase, subunit B	Mhun_2171	3.03	1.24	11,614	0.77	3.9
Q2FRN4	Tetrahydromethanopterin S-methyltransferase, subunit C	Mhun_2170	2.78	0.70	36,369	1.03	4.9
Q2FKY3	Formate/nitrite transporter	Mhun_1811	2.68	0.56	36,056	0.95	7.8
Q2FRN5	Tetrahydromethanopterin S-methyltransferase, subunit D	Mhun_2169	2.65	0.34	32,620	0.98	5.1
Q2FPL5	PKD	Mhun_0424	2.54	0.25	164,155	-0.24	4.3
Q2FRN6	Tetrahydromethanopterin S-methyltransferase subunit E	mtrE	2.49	0.19	37,126	0.50	6.1
Q2FLP4	PKD	Mhun_0597	2.40	0.19	233,108	-0.31	4.3
Q2FP79	Putative uncharacterized protein	Mhun_1218	2.35	0.59	15,484	0.50	6.3
Q2FUM4	Flagellin	Mhun_3140	1.79	0.32	21,248	0.13	9.7
Q2FSN0	Methyl-coenzyme M reductase, alpha subunit	Mhun_2148	1.78	0.17	72,947	-0.31	5.9
Q2FRN2	Tetrahydromethanopterin S-methyltransferase subunit A	mtrA	1.76	0.09	29,770	0.16	4.9
Q2FNK4	H ⁺ -transporting two-sector ATPase, C subunit	Mhun_1179	1.65	0.07	9,607	1.38	3.9
Q2FRU4	PKD	Mhun_2442	1.38	0.78	118,118	-0.29	4.6
Q2FR15	ABC transporter, substrate-binding protein, aliphatic sulfonates	Mhun_0085	1.10	0.18	43,801	0.03	4.7
Q2FU12	30S ribosomal protein S8e	rps8e	1.10	0.10	16,172	-0.84	11.8
Q2FU40	Methyl-accepting chemotaxis sensory transducer	Mhun_2944	1.09	1.03	101,574	-0.03	4.5
Q2FQ72	Putative uncharacterized protein	Mhun_1632	1.06	0.44	28,029	1.20	4.6
Q2FT98	30S ribosomal protein S11P	rps11p	1.06	0.17	16,210	-0.42	11.4
Q2FN87	Putative K(+)-stimulated pyrophosphate-energized sodium pump	hppA	1.02	0.05	80,584	0.82	5.7
Q2FS99	50S ribosomal protein L21e	rpl21e	0.97	0.08	12,998	-0.75	12.1
Q2FSG6	LSU ribosomal protein L32E	Mhun_2236	0.96	0.04	19,622	-1.12	10.7
Q2FTN5	50S ribosomal protein L13P	rpl13p	0.93	0.16	17,910	-0.36	10.9
Q2FU60	50S ribosomal protein L37Ae	rpl37ae	0.93	0.39	12,339	-0.45	11.4
Q2FSG7	LSU ribosomal protein L15P	Mhun_2231	0.88	0.24	17,730	-0.55	10.8
Q2FU92	50S ribosomal protein L4P	rpl4p	0.84	0.24	31,670	-0.37	11.1
Q2FRN1	Tetrahydromethanopterin S-methyltransferase, subunit A	Mhun_2174	0.80	0.13	30,256	0.13	4.8
Q2FRU7	PKD	Mhun_2441	0.78	0.42	128,125	-0.32	4.3
Q2FU91	50S ribosomal protein L3P	rpl3p	0.77	0.10	44,312	-0.60	10.3
Q2FT96	30S ribosomal protein S13P	rps13p	0.72	0.26	19,566	-0.49	10.4
Q2FSD5	Putative uncharacterized protein	Mhun_2609	0.65	0.09	34,300	-0.23	4.8
Q2FSG9	Ribosomal protein L19e	Mhun_2235	0.64	0.07	19,700	-0.98	12.0
Q2FT97	30S ribosomal protein S4P	rps4p	0.62	0.10	23,945	-0.80	10.5

Q2FTV5	Putative uncharacterized protein	Mhun_2617	0.60	0.03	29,858	0.04	4.9
Q2FRI1	30S ribosomal protein S10P	rps10p	0.58	0.36	13,521	-0.54	9.7
Q2FT35	30S ribosomal protein S19P	rps19p	0.57	0.10	18,423	-0.57	10.7
Q2FS30	30S ribosomal protein S15P/S13e	rps15p	0.55	0.07	20,364	-0.74	10.7
Q2FTN6	30S ribosomal protein S9P	Mhun_2896	0.54	0.17	17,022	-0.42	11.3
Q2FTA2	50S ribosomal protein L18e	rpl18e	0.51	0.02	15,528	-0.43	10.2
Q2FT37	30S ribosomal protein S4e	rps4e	0.48	0.04	31,234	-0.44	10.3
Q2FT30	50S ribosomal protein L29P	rpl29p	0.44	0.13	8,934	-0.66	10.8
Q2FRG5	SPFH domain, Band 7 family protein	Mhun_1614	0.44	0.09	46,868	-0.09	4.9
Q2FQX8	Putative uncharacterized protein	Mhun_1595	0.41	0.07	49,918	-0.28	5.2
Q2FMC0	Putative uncharacterized protein	Mhun_1947	0.38	0.04	45,955	-0.07	4.4
Q2FSH0	50S ribosomal protein L18P	rpl18p	0.38	0.10	22,224	-0.26	9.9
Q2FUL1	Protein translocase subunit SecD	secD	0.38	0.07	61,027	0.32	4.9
Q2FQ32	50S ribosomal protein L1P	rpl1p	0.36	0.09	27,753	-0.29	10.1
Q2FMZ6	Extracellular solute-binding protein, family 5	Mhun_0218	0.35	0.05	69,055	-0.32	4.8
Q2FS98	50S ribosomal protein L31e	rpl31e	0.32	0.01	11,677	-0.76	10.5
Q2FTS9	5,10-methylenetetrahydro - methanopterin reductase	mer	0.31	0.13	40,730	0.00	8.1
Q2FRV2	Tetratricopeptide TPR_2	Mhun_2391	0.30	0.31	26,982	-0.39	4.5
Q2FQ61	Methyl-accepting chemotaxis sensory transducer	Mhun_1844	0.30	0.07	89,395	0.04	4.3
Q2FS26	Putative uncharacterized protein	Mhun_2513	0.20	0.07	106,862	-0.42	4.4
Q2FTJ9	SSU ribosomal protein S19E	Mhun_3015	0.18	0.07	17,838	-0.49	10.1
Q2FS69	30S ribosomal protein S12P	rps12p	0.17	0.05	18,102	-0.42	10.8
Q2FSG4	50S ribosomal protein L30P	rpl30p	0.17	0.05	19,964	-0.37	10.0
Q2FRY1	Methyl-accepting chemotaxis sensory transducer	Mhun_2669	0.16	0.22	101,707	-0.03	4.6
Q2FLA6	RNA-binding region RNP-1 (RNA recognition motif)	Mhun_0356	0.15	0.02	13,151	-0.70	5.0
Q2FS15	TPR repeat	Mhun_2994	0.10	0.05	41,295	-0.14	5.5
Q2FP78	Putative uncharacterized protein	Mhun_1219	0.01	0.02	15,554	0.53	6.5

Table 3-1. Absolute quantification of *M. hungatei* JF1 sheath proteins quantified in at least two replicate injections.

Amino acids		Patel et al 1986		Kandler 1978	Sprott 1980	Current Study	
		JF1	GP1	JF-1	GP1	Protein	Peptide
Hydrophobic	Met	0.31	0.36	1.41	1.36	2.70	1.29
	Val	6.7	7.18	4.02	3.88	7.68	8.23
	Ile	4.31	5.17	3.42	3.3	6.85	6.08
	Leu	7.68	6.4	7.45	7.17	7.77	8.54
	Tyr	4.29	2.62	4.43	4.26	5.09	4.23
	Phe	5.12	3.9	4.43	4.26	5.42	4.95
total		28.41	25.63	25.14	24.23	35.51	33.32
Neutral	Cys	0.57	0.62	0.80	0.78	0.65	0.39
	Thr	7.95	10.82	4.43		6.59	7.74
	Ser	7.51	8.18	7.64	7.36	5.16	7.56
	Pro	3.93	3.17	3.02	2.91	3.53	4.94
	Gly	7.23	6.91	10.46	10.08	5.48	3.97
	Ala	12.77	9.46	13.89		7.67	7.18
total		39.96	39.16	40.25	21.13	29.09	33.16
Basic	Lys	5.31	6.72	4.43	4.26	6.18	4.69
	His	0.66	0.81	0.60	0.58	1.35	1.15
	Arg	1.63	1.56	1.41	1.36	4.54	5.36
total		7.6	9.09	6.44	6.2	12.08	11.20
Acidic	Asx	14.16	15.05	17.71	17.05	11.29	12.90
	Glx	9.87	11.07	10.46	10.08	9.85	9.42
total		24.03	26.12	28.17	27.13	21.14	22.32
Acidic:Basic		3.16	2.87	4.37	4.38	1.75	1.99
Lys:Ala:Glx		1:2.4:1.9	1:1.4:1.6	1:3.1:2.4	N/A	1:1.2:1.6	1:1.5:2.0

Table 3-2. Amino acid composition of identified *M. hungatei* sheath proteins. The data include *M. hungatei* sheath from JF1 and GP1 (Patel et al 1986), from *M. hungatei* JF1 (Kandler 1978); *M. hungatei* GP1 (Sprott 1980); and from *M. hungatei* JF1 protein and peptide, Mhun_2271 (this study).

REFERENCES

1. Jarrell, K. F. *et al.* Major players on the microbial stage: why archaea are important. *Microbiology* **157**, 919–936 (2011).
2. Thauer, R. K., Kaster, A. K., Seedorf, H., Buckel, W. & Hedderich, R. Methanogenic archaea: ecologically relevant differences in energy conservation. *Nat Rev Microbiol* **6**, 579–591 (2008).
3. Ferry, J. G., Smith, P. H. & Wolfe, R. S. *Methanospirillum*, a New Genus of Methanogenic Bacteria, and Characterization of *Methanospirillum hungatii* sp.nov. *Int J Syst Bacteriol* **24**, 465–469 (1974).
4. McInerney, M. J. *et al.* Physiology, Ecology, Phylogeny, and Genomics of Microorganisms Capable of Syntrophic Metabolism. *Ann N Y Acad Sci* **1125**, 58–72 (2008).
5. Toso, D. B., Henstra, A. M., Gunsalus, R. P. & Zhou, Z. H. Structural, mass and elemental analyses of storage granules in methanogenic archaeal cells. *Environ Microbiol* **13**, 2587–2599 (2011).
6. Firtel, M., Southam, G., Harauz, G. & Beveridge, T. J. Characterization of the cell wall of the sheathed methanogen *Methanospirillum hungatei* GP1 as an S layer. *J Bacteriol* **175**, 7550–7560 (1993).
7. Beveridge, T. J., Southam, G., Jericho, M. H. & Blackford, B. L. High-resolution topography of the S-layer sheath of the archaeobacterium *Methanospirillum hungatei* provided by scanning tunneling microscopy. *J Bacteriol* **172**, 6589–6595 (1990).
8. Beveridge, T. J., Sprott, G. D. & Whippey, P. Ultrastructure, inferred porosity, and gram-staining character of *Methanospirillum hungatei* filament termini describe a unique cell permeability for this archaeobacterium. *J Bacteriol* **173**, 130–140 (1991).
9. Beveridge, T. J., Stewart, M., Doyle, R. J. & Sprott, G. D. Unusual stability of the *Methanospirillum hungatei* sheath. *J Bacteriol* **162**, 728–737 (1985).
10. Xu, W. W. *et al.* Modeling and measuring the elastic properties of an archaeal surface, the sheath of *Methanospirillum hungatei*, and the implication of methane production. *J Bacteriol* **178**, 3106–3112 (1996).
11. Stewart, M. M., Beveridge, T. J. T. & Sprott, G. D. G. Crystalline order to high resolution in the sheath of *Methanospirillum hungatei*: a cross-beta structure. *J Mol Biol* **183**, 509–515 (1985).

12. Patel, G. B., Sprott, G. D., Humphrey, R. W. & Beveridge, T. J. Comparative analyses of the sheath structures of *Methanothrix concilii* GP6 and *Methanospirillum hungatei* strains GP1 and JF1. *Can J Microbiol* **32**, 623–631 (1986).
13. Sleytr, U. B. & Messner, P. Crystalline surface layers in procaryotes. *J Bacteriol* **170**, 2891–2897 (1988).
14. Southam, G. G. & Beveridge, T. J. T. Detection of growth sites in and protomer pools for the sheath of *Methanospirillum hungatei* GP1 by use of constituent organosulfur and immunogold labeling. *J Bacteriol* **174**, 6460–6470 (1992).
15. Southam, G. & Beveridge, T. J. Characterization of novel, phenol-soluble polypeptides which confer rigidity to the sheath of *Methanospirillum hungatei* GP1. *J Bacteriol* **174**, 935–946 (1992).
16. Masuda, T., Tomita, M. & Ishihama, Y. Phase Transfer Surfactant-Aided Trypsin Digestion for Membrane Proteome Analysis. *J Proteome Res* **7**, 731–740 (2008).
17. Yeung, Y. G., Nieves, E., Angeletti, R. H. & Stanley, E. R. Removal of detergents from protein digests for mass spectrometry analysis. *Anal Biochem* **382**, 135–137 (2008).
18. Li, G. Z. *et al.* Database searching and accounting of multiplexed precursor and product ion spectra from the data independent analysis of simple and complex peptide mixtures. *Proteomics* **9**, 1696–1719 (2009).
19. Silva, J. C. J., Gorenstein, M. V. M., Li, G.-Z. G., Vissers, J. P. C. J. & Geromanos, S. J. S. Absolute quantification of proteins by LCMS^E: a virtue of parallel MS acquisition. *Mol Cell Proteomics* **5**, 144–156 (2006).
20. Silva, J. C. J. *et al.* Simultaneous qualitative and quantitative analysis of the *Escherichia coli* proteome: a sweet tale. *Mol Cell Proteomics* **5**, 589–607 (2006).
21. Silva, J. C. *et al.* Quantitative Proteomic Analysis by Accurate Mass Retention Time Pairs. *Anal Chem* **77**, 2187–2200 (2005).
22. Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J Mol Biol* **157**, 105–132 (1982).
23. Southam, G. G. & Beveridge, T. J. T. Dissolution and immunochemical analysis of the sheath of the archaeobacterium *Methanospirillum hungatei* GP1. *J Bacteriol* **173**, 6213–6222 (1991).
24. Kandler, O. & König, H. Chemical composition of the peptidoglycan-free cell walls of methanogenic bacteria. *Arch Microbiol* **118**, 141–152 (1978).

25. Sprott, G. D. & McKellar, R. C. Composition and properties of the cell wall of *Methanospirillum hungatii*. *Can J Microbiol* **26**, 115–120 (1980).

CHAPTER 4

Comprehensive proteome characterization of the newly discovered *Anaerobaculum hydrogeniformans* strain OS-1, and the modulation of catabolic pathways to efficiently grow with glucose, serine, or histidine.

ABSTRACT

Anaerobaculum hydrogeniformans strain OS-1 is a novel anaerobic bacterium, isolated from oil production waters in Alaska, which can produce hydrogen at levels approaching the theoretical limit. This capability makes *A. hydrogeniformans* a potentially important organism for fossil fuel-free biohydrogen production, and for proteomic and genomic investigations that identify pathways and genes driving hydrogen production. We used label-free quantitative proteomics to catalog the identifiable proteins from the generated protein database for *A. hydrogeniformans* strain OS-1, and to analyze the proteins and pathways that are differentially regulated in the presence of different carbon sources.

A. hydrogeniformans mono-cultures were grown with glucose, histidine, or serine; cells grown with glucose were harvested at three timepoints that correspond to the oxidation of 1/3, 2/3, and 3/3 of total glucose. Harvested cells were processed by eFASP, and the produced peptides are analyzed by shotgun-LC-MS^E, or were fractionated by Hydrophilic Interaction Chromatography (HILIC) in TopTip cartridges (PolyLC Inc., Columbia, MD) prior to analysis by LC-MS^E.

Shotgun LC-MS^E analysis identified 1,002 proteins, with an average sequence coverage of 27.8%. HILIC-LC-MS^E analysis identified 1,223 proteins, with an average sequence coverage of 37.5%. HILIC fractionation enabled the identification of 26% more proteins with transmembrane helices (TMHs), and 32% more hydrophobic proteins with GRAVY values over 0.4. In this study, we identified 59% of non-TMH proteins, and 29% of TMH-containing proteins, contained in the hypothetical proteome of *A. hydrogeniformans* strain OS-1. FASP-HILIC allows for very good fractionation resolution, with over 70% of peptides being identified in only one fraction, which compares favorably to FASP coupled with Strong Anion Exchange (FASP-SAX), which identifies approximately 50% of peptides in only one fraction.

When grown with glucose, *A. hydrogeniformans* cells induce pathways involved in glucose utilization, energy storage and trehalose synthesis. Glycolytic enzymes are unregulated, as well as enzymes involved in the breakdown of pyruvate into formate and acetyl-CoA, or the funneling of pyruvate into starch or fatty acids. The enzymes responsible for the production of the disaccharide trehalose, from glucose, are upregulated under glucose conditions; trehalose has been associated desiccation endurance in many organisms. Many of these effects can be accounted for based on the up regulation of a variety of transcription factors, including, Rex, GntR, IclR, TrpR, and LuxR.

In the presence of serine, *A. hydrogeniformans* up regulates many pathways involved in utilization of this glycogenic amino acid as an energy source and as a carbon source, including those involved in the conversion of serine into pyruvate, into glycine, into 3-phosphoglycerate, and into threonine. Growth in the presence of

histidine results in the upregulation of enzymes required for histidine catabolism, including histidine ammonia lyase and urocanate hydratase.

TABLE OF CONTENTS

INTRODUCTION.....	98
EXPERIMENTAL.....	99
Strains and cell growth	99
Enhanced filter aided sample preparation of <i>A. hydrogeniformans</i>	99
Peptide fractionation by HILIC	101
Analysis by nanoLC-MS ^E	102
Data processing and protein identification	103
RESULTS AND DISCUSSION.....	104
HILIC as a simple and effective pre-fractionation step prior to LC-MS ^E	104
Proteome characterization.....	113
Analysis of <i>A. hydrogeniformans</i> grown on various carbon sources.....	118
Glucose: carbohydrate metabolism	121
Glucose: gene regulation.....	125
Serine: amino acid degradation and conversion	127
Histidine: amino acid degradation.....	129
REFERENCES.....	131

INTRODUCTION

Anaerobaculum hydrogeniformans strain OS-1 is a novel anaerobic bacterium that can produce hydrogen from glucose, at levels near the theoretical limit, with 3.68 mol hydrogen produced per mol glucose ¹. The development of sustainable, carbon neutral, energy sources, are increasingly important as stores of fossil fuels are limited, and limiting our contribution to global warming remains important. Hydrogen gas offers such an alternative.

Current methods for hydrogen production include electrolysis of water, steam reformation of natural gas; however, both depend on fossil fuels for electrical input. The fossil fuel-free alternatives include, biofuel production, like ethanol and diesel, and microbial produced hydrogen. The search for excellent hydrogen producing microbes has gained traction, yet has resulted in a few suitable organisms.

A. hydrogeniformans strain OS-1 was isolated from oil production waters in Alaska which were being assessed for corrosion causing methanogens and bacteria ². A hydrogen producer from this sample, *A. hydrogeniformans*, was identified and described in 2011, and may be important for biohydrogen production ¹. It is capable of producing hydrogen, acetate and carbon dioxide from a variety of organic substrates, including amino acids and sugars, and can exist independently or syntrophically with hydrogen-oxidizing methanogens ¹.

The goal of this work is to use quantitative proteomics to assess the metabolic pathways utilized by *A. hydrogeniformans* to produce hydrogen from various starting materials. We will analyze cells cultured with glucose, and harvested at three timepoints, and cells cultured with serine and histidine. The results of this work will be

three-fold: (i) to characterize the global proteome of this novel species qualitatively, (ii) to elucidate shared and substrate-specific pathways integral to hydrogen production, and (iii) to assess the utility of HILIC for pre-fractionation of eFASP produced peptides prior to LC-MS^E analysis.

EXPERIMENTAL

Strains and cell growth

A. hydrogeniformans strain OS-1 was cultured with glucose, serine, or histidine. Cells incubated with glucose were harvested at one of three points: when glucose was 1/3 oxidized, 2/3 oxidized, or fully oxidized. Collected cells were washed with PBS three times, and the pellets weighed. Sample pellets were processed by eFASP.

Enhanced filter aided sample preparation of A. hydrogeniformans

Washed cell pellets were suspended in Solubilization Buffer A (4% ammonium dodecyl sulfate, 50 mM TCEP, 0.1% deoxycholic acid, 0.1% n-octylglucoside, 100 mM Tris-HCL, pH 8), and incubated on a heated block for 10 minutes at 90°C. The sample was sonicated for three 10-second intervals with a Sonic Dismembrator Model 100 (Fisher Scientific), and then spun at 16,000 rcf for 10 minutes in an Eppendorf centrifuge; this step was repeated once. Glycogen Phosphorylase B (rabbit) was added to each sample at a 1:50 ratio. The supernatant and any residual pellet, was sonicated and allowed to cool to 37°C. Proteins were alkylated by the addition of 4-vinylpyridine (4-VP) at 35mM final concentration, and incubated for 1 hour on a shaker at 37°C. DTT was added to a final concentration of 50 mM to deactivate residual 4-VP. Based on a

20 mg:1 mg wet weight:protein estimate, 250 µg of each sample was transferred to a new tube for further processing.

Eight volumes of Solubilization Buffer B (8 M urea, 0.1% deoxycholic acid, 0.1% n-octylglucoside, 100 mM Tris-HCL, pH 8) was added to the digest to begin buffer exchange of SDS. The diluted sample was applied to a Passivated Microcon Filter Unit (described below), and spun at 14,000 rcf for 10 minutes. The eluate was discarded, and 200 µl of Solubilization Buffer B was added to the filter unit and then spun at 14,000 rcf for 10 minutes. This step was repeated twice. To remove urea, three buffer exchange steps were performed with Digestion Buffer (0.2% deoxycholic acid in 50 mM ABC). The Passivated Microcon Filter Unit (30,000 Da cutoff), and collection tube, were pre-incubated in 5% Tween-20 overnight, and then thoroughly rinsed with large volumes of MS-grade water.

Digestion Buffer (100 µl) was added to the Microcon unit, followed by the addition of 5 µg of trypsin (1 µg trypsin:50 µg protein). Digestion proceeded for 12 hours on a shaker set to 37°C. The Microcon unit was transferred to a new Passivated Collection Tube, and spun at 14,000 rcf for 10 minutes. To assure complete removal of peptides, 100 µl of 50 mM ABC buffer was added to the Microcon unit, and collected by centrifugation, two times.

Removal of deoxycholic acid from the digest was accomplished by acidification and phase transfer of insoluble deoxycholate to an ethyl acetate layer³. Instead of removing the aqueous layer as in Masuda et al.³, the organic layer was removed, and then two more rounds of addition and removal of ethyl acetate facilitate the removal of

deoxycholate from the solution, similar to the strategy Yeung et al. employ for the removal of n-octylglucoside ⁴.

Collected peptides were dried down in a SpeedVac and re-suspended in 50% methanol, three times. Dried peptides were suspended in modified Mobile Phase A (96.5% H₂O/3% ACN/0.5% formic acid). The peptide solutions were either used for shotgun LC-MS^E analysis, or fractionated by HILIC before LC-MS^E analysis.

Peptide fractionation by HILIC

Binding Buffer: 90% ACN, 15 mM ammonium formate, pH 3.

Release1 Buffer: 90% ACN, 15 mM ammonium formate, pH 3.

Release2 Buffer: 80% ACN, 15 mM ammonium formate, pH 3.

Release3 Buffer: 78% ACN, 15 mM ammonium formate, pH 3.

Release4 Buffer: 74% ACN, 15 mM ammonium formate, pH 3.

Release5 Buffer: 40% ACN, 15 mM ammonium formate, pH 3.

Condition Buffer: 10% ACN, 15 mM ammonium formate, pH 3.

TopTips (TT200HIL, PolyLC Inc., Columbia, MD) were conditioned by wetting the material with Condition Buffer, followed by incubation with Binding Buffer. Each sample (70 µg) was dried down to remove any acids or salts, and re-suspended in a small volume of 150 mM ammonium formate, pH 3. This solution was added to the TopTip, along with enough ACN to bring it to 90% organic. The peptide solution was passed through the TopTip by centrifugation at 500 rcf, and the eluate was passed through the TopTip again. All subsequent steps were done twice. The TopTip was washed with Binding Buffer. Peptides were eluted, into separate fractions, with Release1 Buffer

through Release5 Buffer. All fractions were dried down in a SpeedVac, twice. Peptides were re-suspended in modified Mobile Phase A (96.5% H₂O/3% ACN/0.5% formic acid), for LC-MS^E analysis.

Analysis by nanoLC-MS^E

nanoUPLC of sample peptides was performed on an Waters nanoACQUITY UPLC system equipped with a nanoACQUITY UPLC Symmetry C18 180 µm x 20 mm trap column (Waters, 186003514), and a nanoACQUITY UPLC BEH C18 75 µm x 150 mm reversed phase analytical column (Waters, 186003543). The aqueous mobile phase (Mobile Phase A) contained 0.1% formic acid in water and the organic mobile phase (Mobile Phase B) contained 0.1% formic acid in acetonitrile. 5 µl of the sample was loaded onto the trap with and washed with 99% A and 1% B for 5 minutes, and then injected onto the column with 3% Mobile Phase B. Peptides were eluted from the column with a gradient of 3-40% B over 90 minutes at 300 nl/min, followed by a 5 minute rinse with 95% B. The column was re-equilibrated to 97% A for 20 minutes.

The lock mass compound, [Glu¹]fibrinopeptide (0.5 pmol/µl, 25% ACN, 0.1% formic acid), and the analyte were delivered at 300 nl/min to the NanoLockSpray dual electrospray ion source (Waters, UK) which was equipped with a microcapillary for the lockmass compound, and a PicoTip Emitter (New Objective) for the analyte. Two electrosprays are produced which may be selectively sampled by positioning of the baffle.

The ion source was interfaced with a Waters Synapt HDMS for mass spectrometric analysis. Acquisition was performed in positive nanoESI mode and the

instrument was operated in V-mode with an average resolution of 9500. Accurate mass data for precursor and fragment ions were collected over the 100-2000 m/z range, with the mode of acquisition alternating between low and elevated energy at 1-second intervals, for the duration of analysis. The lock mass channel was sampled every 30 seconds for calibration of MS and MS^E data.

Data processing and protein identification

Raw data files were processed in ProteinLynx Global Server 2.5.2 (PLGS, Waters) to provide an inventory of precursor ions along with their respective fragment ions, to search against the Uniprot *Anaerobaculum hydrogeniformans* database (strain OS-1, March 2012) for peptide and protein identifications ⁵. Database search settings included two missed cleavages, a 4% false positive rate, fixed S-pyridylethyl modification of cysteine residues, and variable oxidation modification of methionine residues. The Waters Expression Informatics component of the PLGS software was utilized to quantify data using glycogen phosphorylase B (Uniprot AC P00489) and trypsin (Uniprot AC P00761) as the quantification references for normalization ⁶⁻⁸. All conditions were compared to Glucose T1 for ratio determination.

Data Extraction and Correction (DECO) software was used to calculate isoelectric points and grand average of hydropathicity index (GRAVY) values for peptides and proteins identified. GRAVY values were calculated based on the hydropathicity of constituent amino acids derived from Kyte and Doolittle ⁹.

Uniprot-GOA ¹⁰ was searched in the QuickGO! browser to obtain compartment and function annotations for all proteins ¹¹.

RESULTS AND DISCUSSION

HILIC as a simple and effective pre-fractionation step prior to LC-MS^E

The enhanced FASP protocol, eFASP, merges the best qualities of the FASP method with removable surfactants. The method produces pure peptide mixtures with good representation of proteins with various physiochemical characteristics. We have also shown its utility in the analysis of the paracrystalline sheath of *Methanospirillum hungatei*, which is notoriously difficult to analyze. In this study we used eFASP, with 0.2% deoxycholic acid, to solubilize and digest proteins from experimental samples. Peptides produced were then analyzed by LC-MS^E, or were separated into 5 fractions by Hydrophilic Interaction Chromatography (HILIC) in TopTip cartridges.

FASP has been coupled to SAX fractionation using StageTips (Thermo) ¹². This allowed for sample desalting, removal of urea used during digestion, and small-scale fractionation. FASP-SAX was used to comprehensively analyze membrane enriched samples and to globally map phosphorylation sites in *Caenorhabditis elegans* ¹²⁻¹⁴.

The primary benchmark for the quality of a separation method is the frequency with which a peptide is identified in the fractions. FASP-SAX of Hela cell lysate provided even an even spread of peptide identifications across the 6 fractions; however, the resolution of SAX is somewhat poor, with only 50% of peptides being identified in one fraction, and 21%, 13%, 9%, 8%, 5% being identified in 2, 3, 4, 5, or all 6 fractions ¹². eFASP-HILIC shows much better resolution, with over 70% of peptides being identified in only one fraction, and 17%, 5%, 2%, and 1% of peptides were found in 2, 3, 4, and 5 fractions (Figure 4-1). The peptide count per fraction was higher in the earlier fractions, but overall seemed comparable to SAX, in light up the increased resolution seen in this

method (Figure 4-2). Histidine was not included in the peptide count comparison because the sample had low protein and peptide counts, based on shotgun experiments and in all HILIC fraction analyses, relative to the other conditions

LC-MS^E experiments collect precursor ion information across the gradient, and are inherently quantitative. This feature allows for a quantitative comparison based on ion volume. Comparing samples based on protein quantification information will not give an idea as to how well the methods function for peptides. Using peptides for comparison allows for the assessment of a much greater feature set in the samples, including peptides from all parts of the dynamic range covered.

Data Extraction and Collation (DECO) collected the ion volume information for all Pass1 peptides identified in the Shotgun and HILIC analyses of all samples. The ion volumes of peptides identified in either all Shotgun analyses (65,815 peptides) or all HILIC analyses (128,147) were plotted in order of increasing intensity (Figure 4-3). From the plot one two major points are apparent. First, there is a large discrepancy of peptides with ion volumes more toward the mid-range, which widens the curve; thus, HILIC is allowing for a better sampling of mid-range peptides that may be important for the quantification of less abundant proteins. The second difference to note is the apparent similarity in the dynamic range of the methods, with HILIC increasing the ion intensity ceiling from under 7 million with Shotgun LC-MS^E, to over 7.5 million with HILIC separation. This indicates that the usage of fractionation is increasing the dynamic range of the method, even though the dynamic range of the instrument is not changing, i.e., highly abundant peptides are more likely to spread between two or more HILIC fractions, and the instrument will only sample partial abundance of the peptide in

each fraction, which allows for a summer total ion volume that exceeds the range of the instrument. This is more apparent in more complex samples.

To quantitatively compare the methods, the ion volumes for peptides from Glucose T1 that were quantified by shotgun LC-MS^E and by HILIC-MS^E were plotted against each other (Figure 4-4). Since the total amount of material injected is different, the correlation is expected to be poor, ($R^2 = 0.75$). While there are select peptides that have greater ion volume in the shotgun LC-MS^E analysis, there are far more with greater intensity in the HILIC LC-MS^E analyses. These data are comparable in the sense that normalization will yield reasonably similar peptide ion volume profiles. However, the number of peptides with ion volumes deviating from the trend line, in the mid- to high-volume range, supports the following two ideas. First, that the poorness of fractionation can be a feature with respect to quantification and instrumental dynamic range limitations. Second, that the co-elution of peptides can have a detrimental effect on the ion volume observed, whether this is due to algorithmic challenges or signal suppression. HILIC allows for the analysis of a greater amount of sample material, and deeper proteome coverage, particularly for lower abundance proteins.

Utilizing HILIC for small-scale pre-fractionation in TopTip cartridges, we identified proteins with much greater sequence coverage, which enables more reliable quantification. The average sequence coverage for proteins from shotgun analyses was 27.8% for 1,002 proteins, while pre-fractionation with HILIC enabled a large increase in sequence coverage to 37.5% for 1,223 proteins identified (Table 4-1 and Figure 4-5).

HILIC affords an 18% increase in protein identifications when compared to all proteins identified by shotgun analyses. To further characterize the difference, TMHMM server was utilized to predict the presence of TMHMMs in all proteins in the proteome of *A. hydrogeniformans*. TMH data were utilized to group and count proteins with TMHs in each analysis, comparing shotgun and HILIC to each other, and to the whole proteome (Table 4-1).

Proteins with TMHs account for about 20% of the potential proteins in the *A. hydrogeniformans* proteome. A combination of shotgun and HILIC data shows the identification of 29% of those TMH containing proteins, while identifying 59% of proteins without TMHs. Approximately 11% of the combined data set accounts for proteins with TMHs, which is very similar to that for the individual shotgun and HILIC data sets. However, the real gain with HILIC is broader and deeper proteome coverage, so while the proportion of proteins with TMHs to those without is similar, the actual number of proteins identified with TMHs increases 26% with HILIC analysis, while the increase in proteins with TMHs is only 18%. The higher proportion of proteins containing TMHs in the HILIC analysis is confirmed by plotting the GRAVY values of all proteins identified by the two methods. HILIC analysis enables a 32% gain in the representation of proteins with GRAVY values over 0.4 (these will generally not be seen in a 2D gel) ¹⁵; 44 proteins identified in the HILIC analyses compared to 30 identified in the shotgun analyses (Figure 4-5).

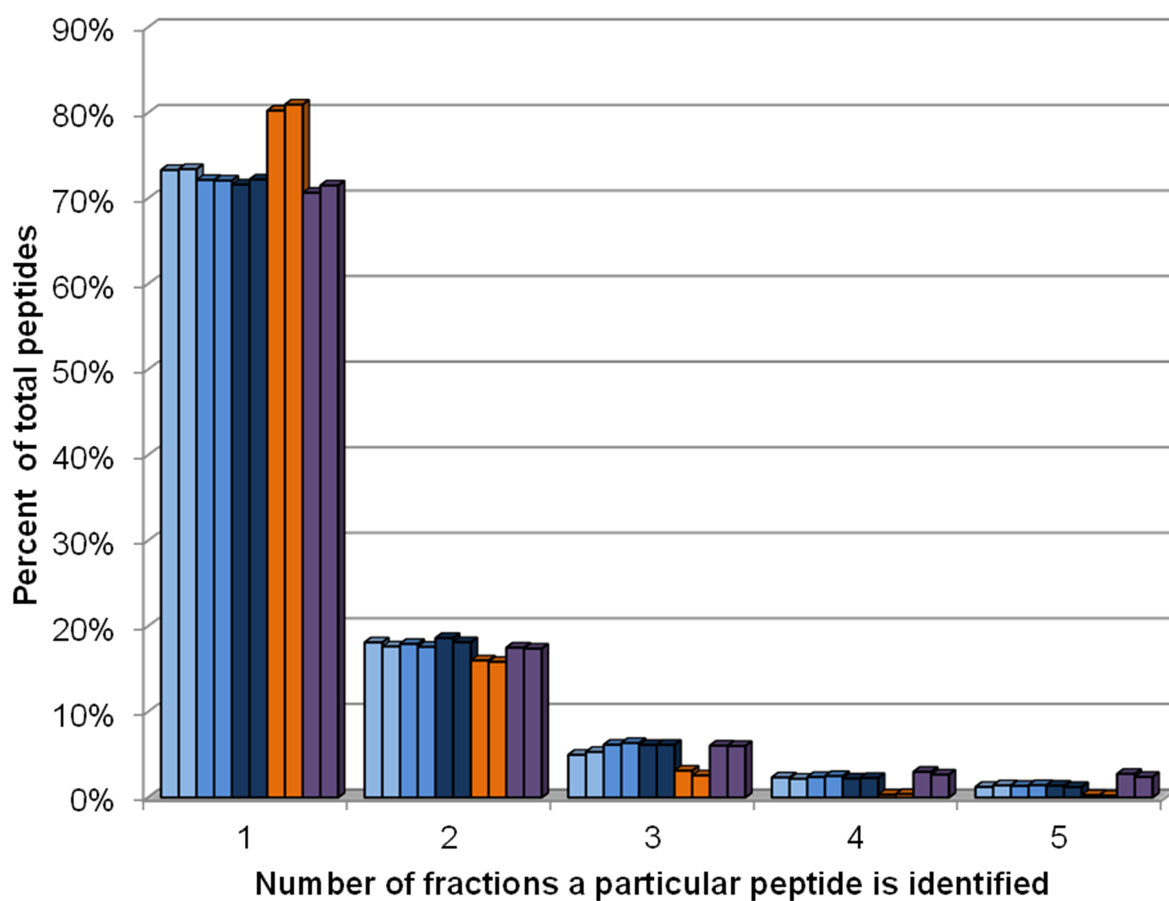


Figure 4-1. Resolution of separation. Glucose T1 (light blue), Glucose T2 (cornflower blue), Glucose T3 (blueberry), Histidine (gamboge) and Serine (aubergine). Both replicates of each HILIC fraction were compared and the number of fractions that each peptide was identified in was tabulated. An average of 74% of peptides were found in only one fraction, and 17%, 5%, 2%, and 1% of peptides were found in 2, 3, 4, and 5 fractions.

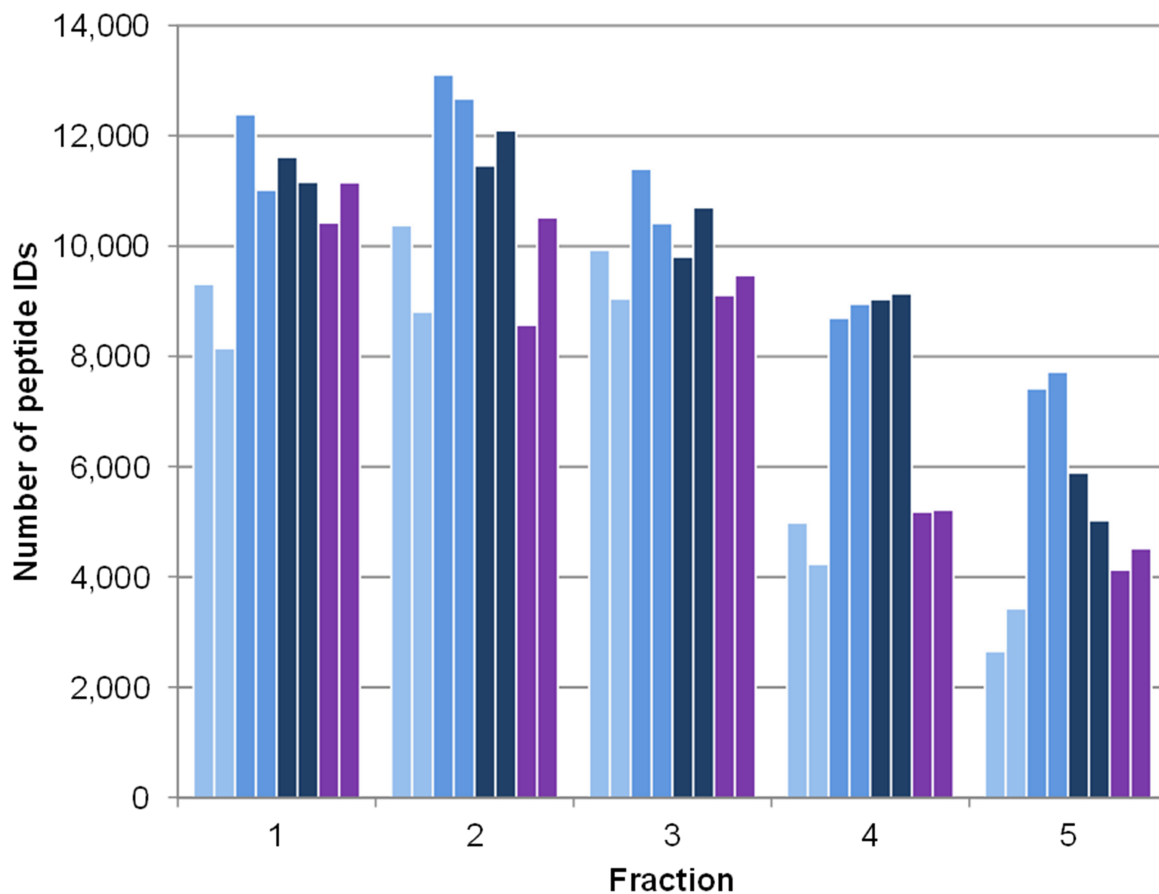


Figure 4-2. Peptide abundance in each eFASP-HILIC fraction. The number of peptides identified in technical replicate analyses of eFASP-HILIC fractions is presented for Glucose T1 (light blue), Glucose T2 (cornflower blue), Glucose T3 (blueberry), and Serine (aubergine).

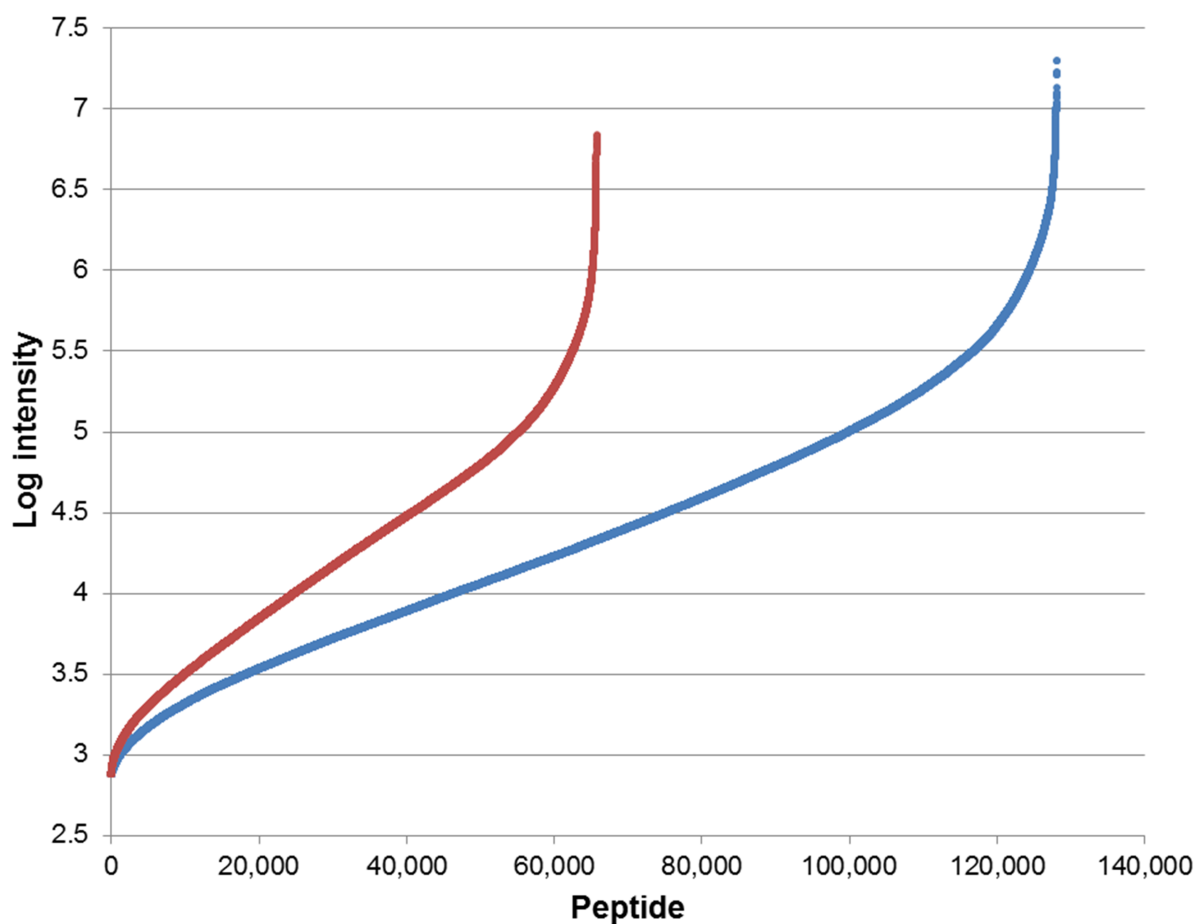


Figure 4-3. Comparison of the dynamic range in peptide abundance between Shotgun and HILIC analyses. The dynamic range of peptide intensities from HILIC and shotgun experiments is presented, based on 65,815 peptides from 15 shotgun analyses (red) and 128,147 peptides from 10 HILIC experiments (blue).

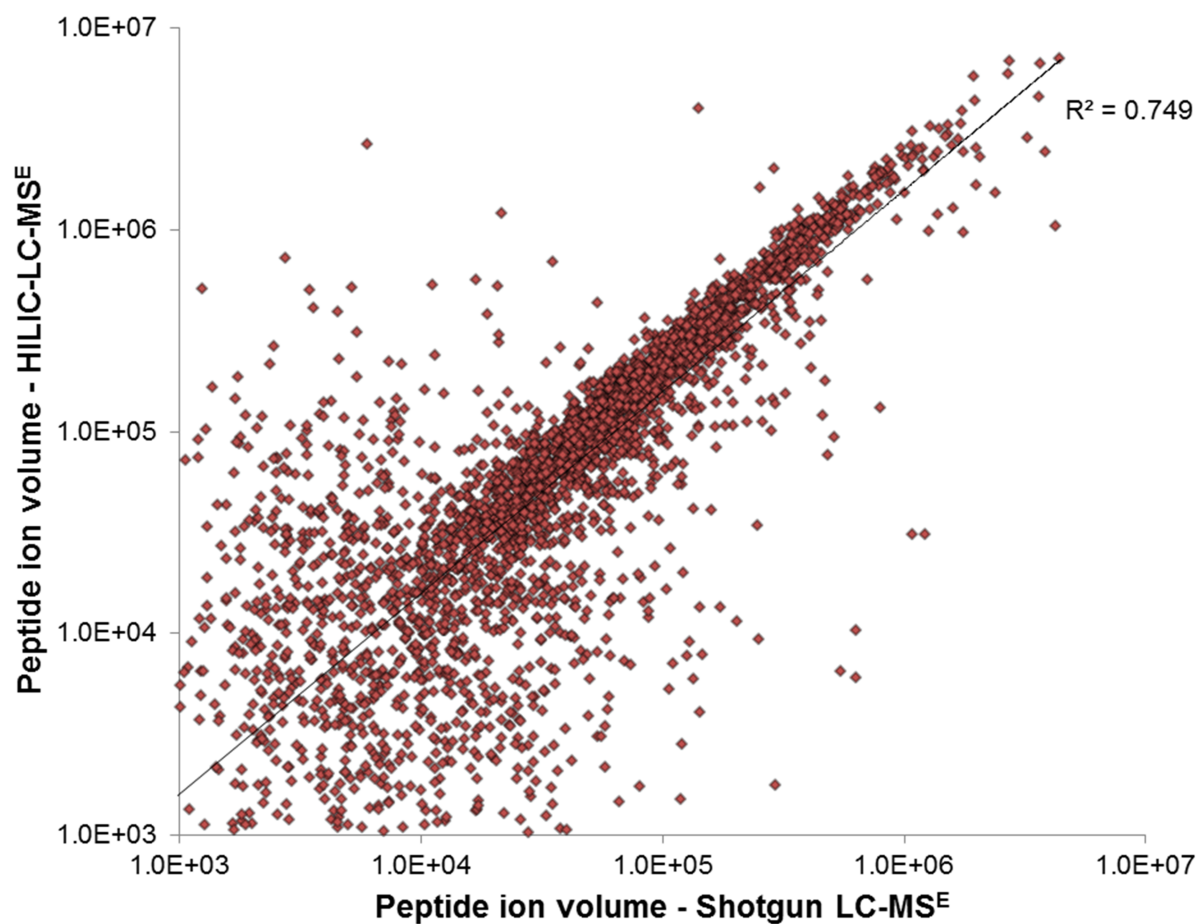


Figure 4-4. Minimal variation in peptide abundance between shotgun and HILIC analyses. Peptide ion volumes for the 3,167 peptides that were identified and quantified in both analyses (Shotgun-LC-MS^E (x-axis) and HILIC-MS^E (y-axis)). of Glucose T1 by shotgun LC-MS^E and HILIC-LC-MS^E.

	TMH Count			
	Shotgun + HILIC	Whole Proteome	Shotgun	HILIC
18	0	1	0	0
17	0	1	0	0
16	1	2	1	1
15	0	3	0	0
14	2	5	2	2
13	2	6	1	2
12	3	17	3	3
11	3	16	0	3
10	8	30	4	8
9	1	12	0	1
8	6	25	4	6
7	7	22	4	7
6	3	31	2	3
5	4	23	4	3
4	6	38	5	6
3	4	37	3	3
2	19	54	13	18
1	77	176	55	71
Proteins w/ TMHs	146	499	101	137
Proteins w/o TMHs	1,171	1,993	1,002	1,223
% proteins w/ TMHs	11%	20%	10%	11%

Table 4-1. Representation of proteins with TMHs in shotgun and HILIC analyses, and in the whole predicted proteome of *A. hydrogeniformans*. TMHMM software was used to predict the potential transmembrane helices present for proteins in analyzed samples, and in the proteome of *A. hydrogeniformans*. The range of trans-membrane helix counts per protein is shown (column 1) along with the corresponding number of proteins with specific TMH counts in each of the following data sets: both shotgun and HILIC experiments (column 2), all proteins in the proteome (column 3), only shotgun experiments (column 4), only HILIC experiments (column 5). The lower three rows represent the total number of counted proteins with and without TMHs, along with the percent with TMHs.

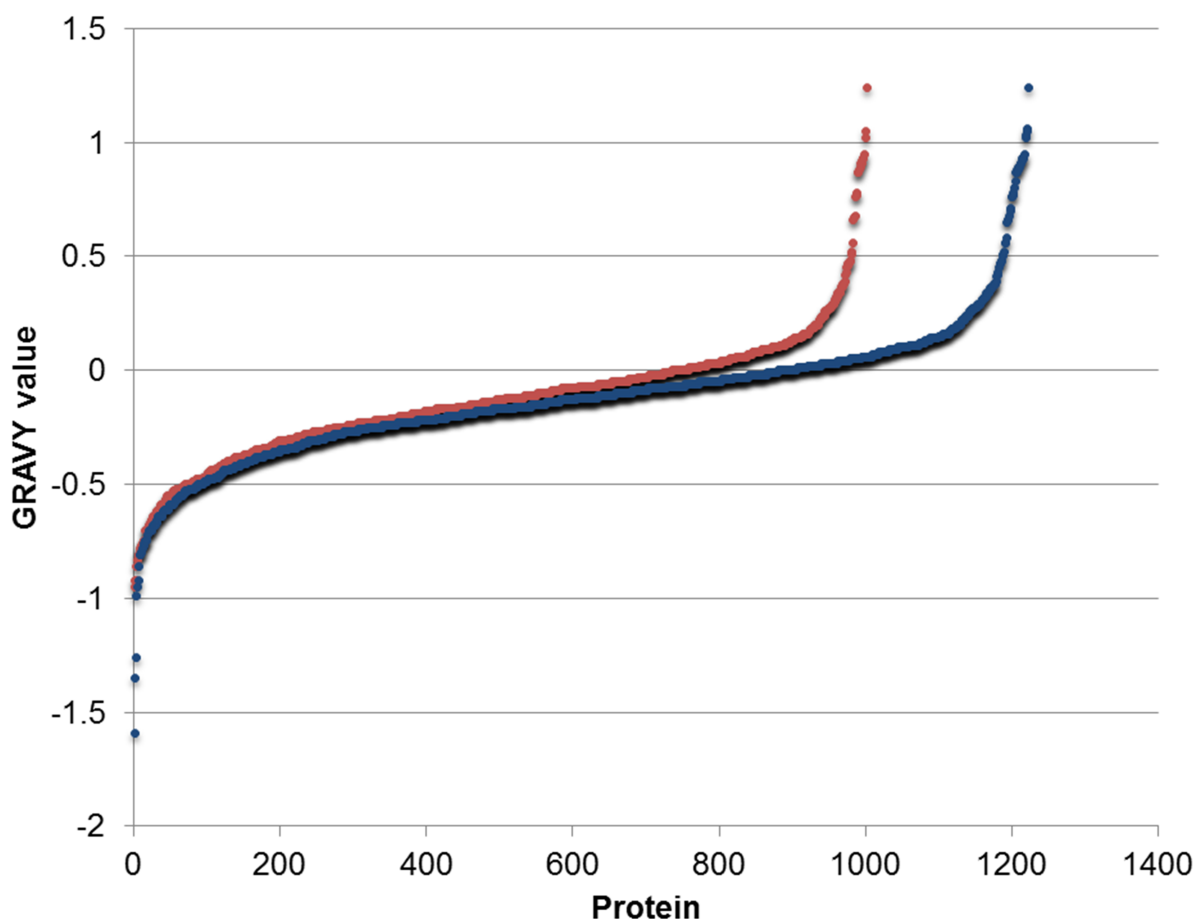


Figure 4-5. Comparison of the GRAVY values for proteins identified in Shotgun and HILIC analyses. GRAVY values calculated by DECO for 1,002 proteins identified in shotgun analyses (red) and 1,223 proteins identified in HILIC analyses (red).

Proteome characterization

With the sequencing of the genome and the generation of the predicted proteome, these data allow for a characterization of the proteins expressed over a variety of conditions, i.e. a general profile of the *A. hydrogeniformans* proteome. Proteins had an average molecular weight of 34,581 +/- 21,635 Da (Figure 4-7), with an isoelectric point around 6.4 +/- 1.8. The average GRAVY value is -0.12 +/- 0.28 (Figure 4-6). Previous studies have presented gravity values of 0.13 +/- 0.64 for *S. cerevisiae*, 0.13 +/- 0.65 for *C. elegans*, and -0.27 +/- 0.61 for a human embryonic kidney cell line¹⁶.

Annotation of all proteins identified in the study was performed with QuickGO!. All annotations were grouped with their associated protein for analysis. The frequency of keywords for each GO category (component, process, function) were tabulated (Figure 4-8). These associations indicate the major structural and functional features of *A. hydrogeniformans*.

For the component category, there were 50 unique keywords found in 553 annotations. The majority of annotations referred to subcellular location, cytoplasm, intracellular, membrane, or integral to membrane, with 35%, 12%, 7% and 6%, respectively. The remaining keywords referred to protein complex association, like ribosome, ribonucleoprotein complex, ATPase complex, and the glycine reductase complex, which represented 9%, 6%, 4% and 1%, respectively.

There were a much higher number of keywords in the Process category, with 321 unique keywords found in 2,990 annotations. A large portion of these relate to metabolism and oxidation-reduction processes for energy production. The second large group is composed of keywords involving transcription, translation, and amino acid biosynthesis.

The GO category for function represented 467 unique keywords found in 5,027 annotations. In line with the GO Process category, the most represented keywords were associated with ATP binding, oxidoreductase activity, and iron-sulfur cluster binding. The second most represented group involves nucleotide binding, DNA binding, and RNA binding. From these analyses with GO annotations, it is clear that the most important features of the *A. hydrogeniformans* species is the production of energy and growth.

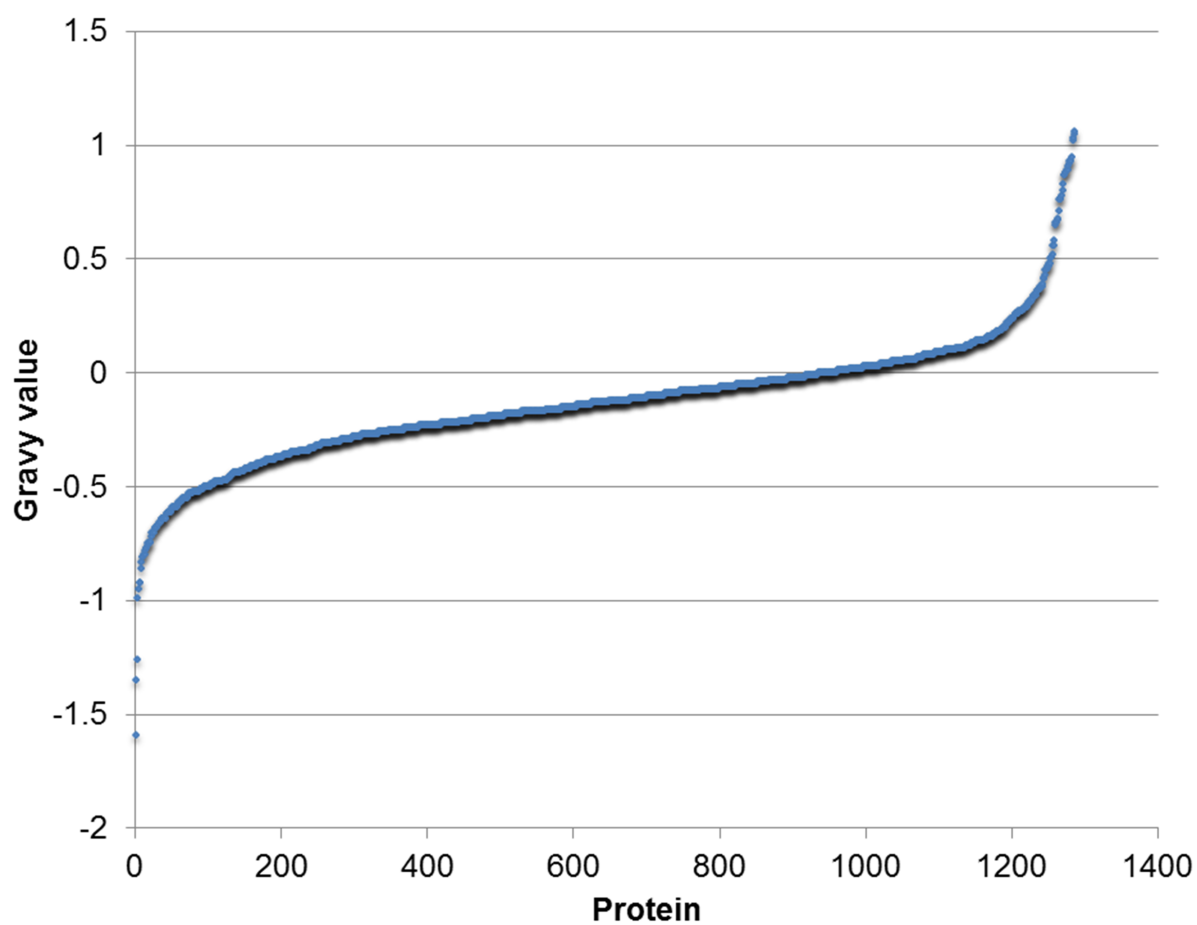


Figure 4-6. GRAVY values for all proteins identified. The GRAVY values for the 1,287 proteins that were identified, at least once, in shotgun or HILIC analyses, was calculated by DECO and is presented.

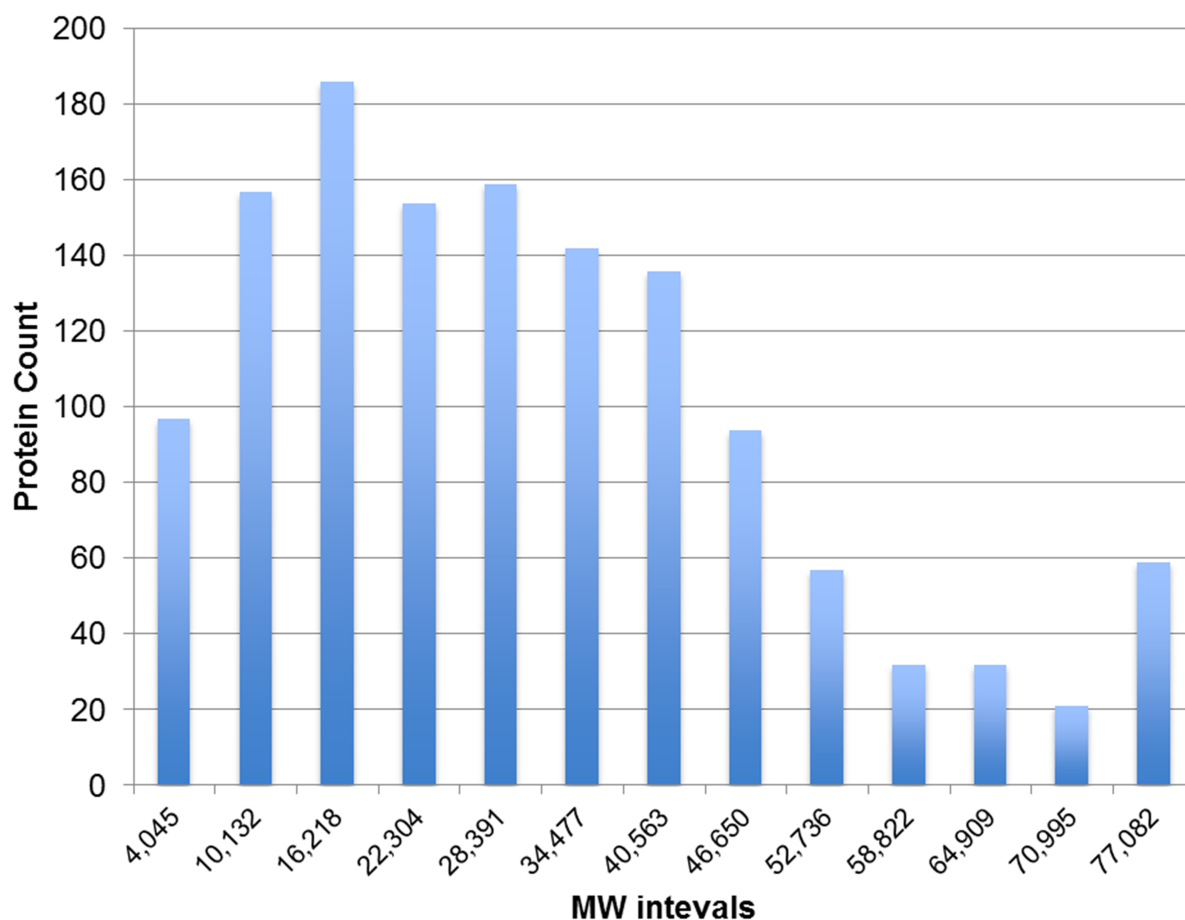
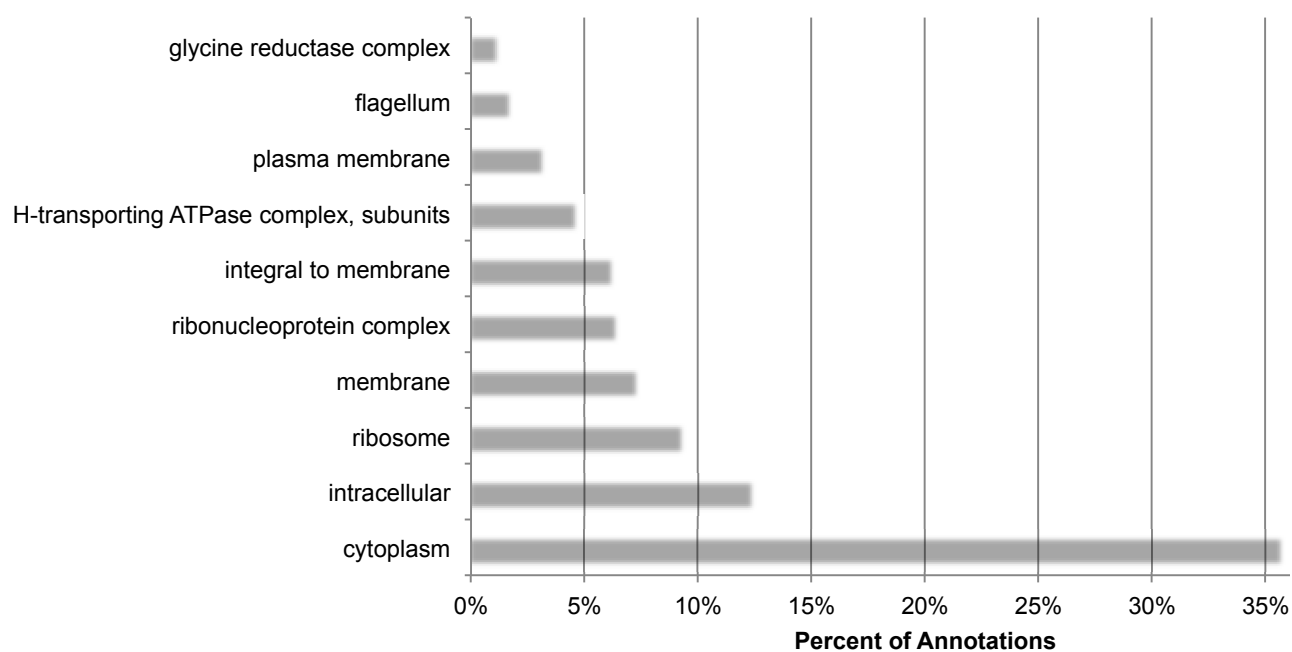
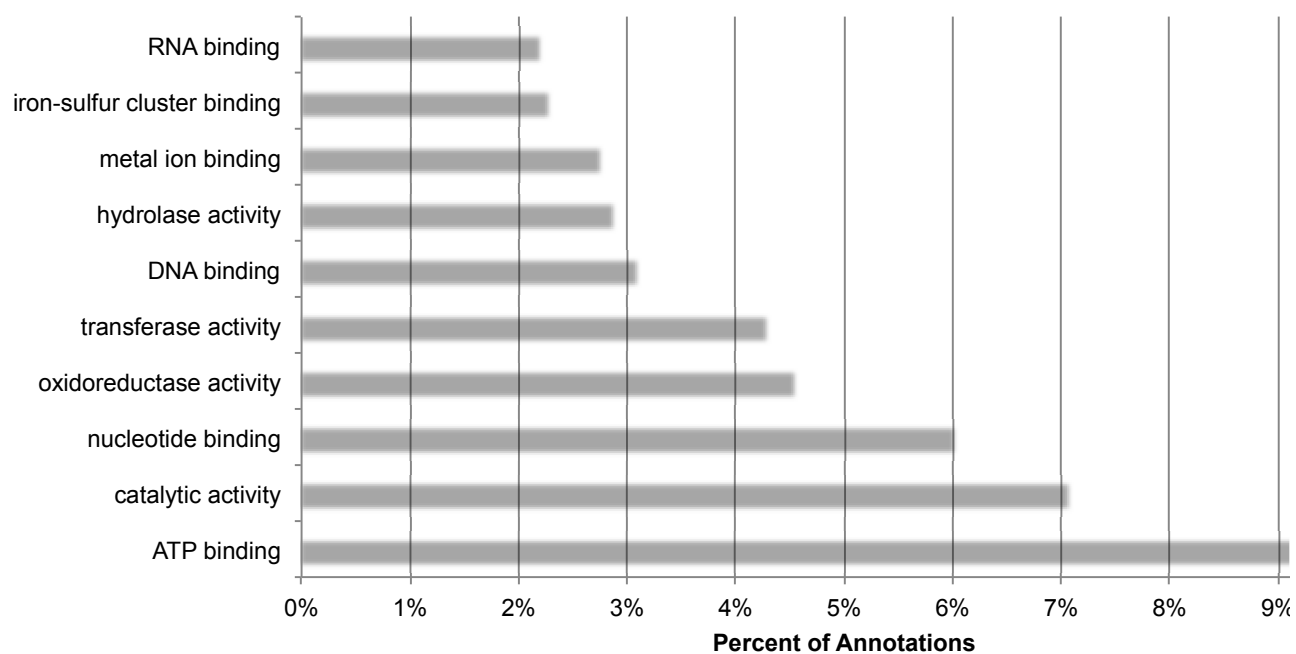


Figure 4-7. Molecular weight distribution for all proteins identified. The molecular weight for the 1,287 proteins identified, at least once, in shotgun or HILIC analyses was calculated and distributed in to the following thirteen bins: 0 to 4,045 Da, 4046 to 10,132 Da, 10,133 to 16,218 Da, 16,219 to 22,304 Da, 22,305 to 28,391 Da, 28,392 to 34,477 Da, 34,478 to 40,563 Da, 40,564 to 46,650 Da, 46,651 to 52,736 Da, 52,767 to 58,822 Da, 58,823 to 64,909 Da, 64,910 to 70,995 Da, 70,996 to 77,082 Da.

GO Component Annotations



GO Function Annotations



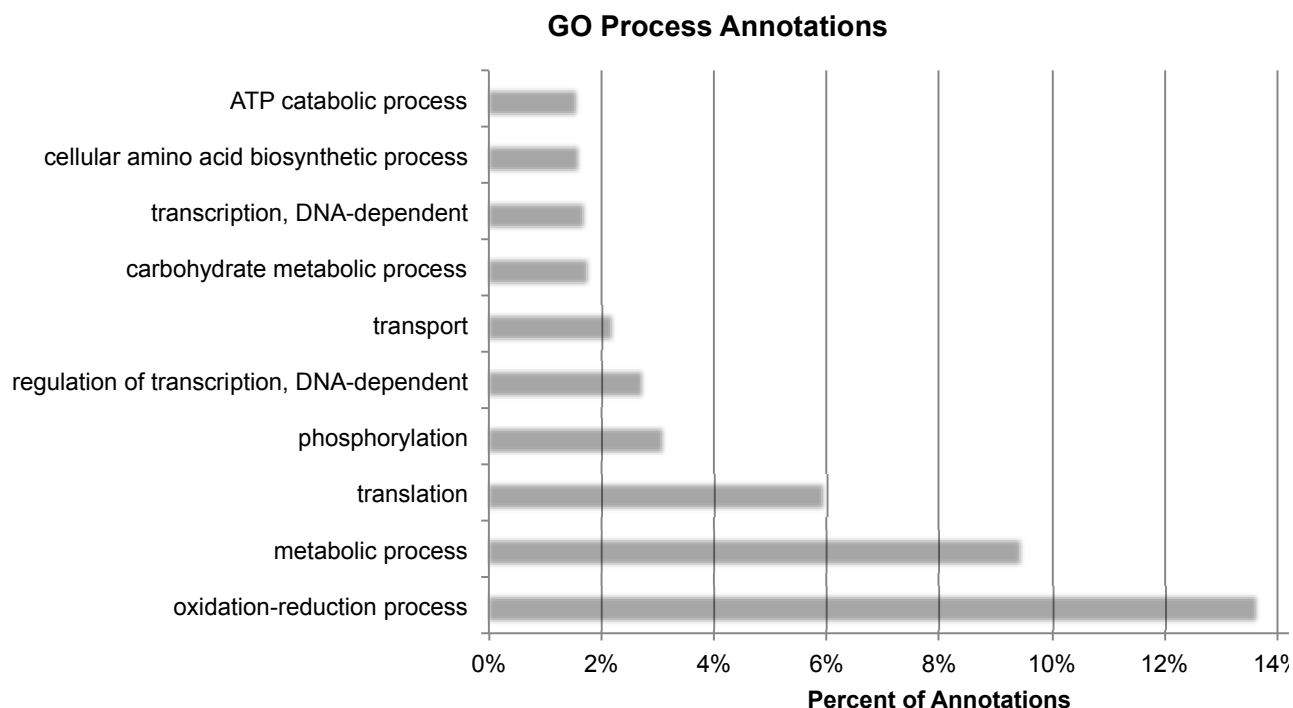


Figure 4-8. GO annotation analysis of the combined shotgun and HILIC data sets. The annotation keywords ascribed to each protein were collected and separated based on GO category: (a) cellular component, (b) process, and (c) Functions. Redundant annotations were removed. The most frequently occurring keywords in each category are presented as a percent of the total keyword count in the specified category.

Analysis of A. hydrogeniformans grown on various carbon sources

A. hydrogeniformans is a prolific producer of hydrogen, and may have utility in biotechnology and biofuel production. Using quantitative proteomics and quantitative mass spectrometry, the proteome of *A. hydrogeniformans* has been catalogued when grown with various carbon sources. Cells cultured with glucose, and harvested at three timepoints, as well as cells cultured with serine and histidine, were analyzed.

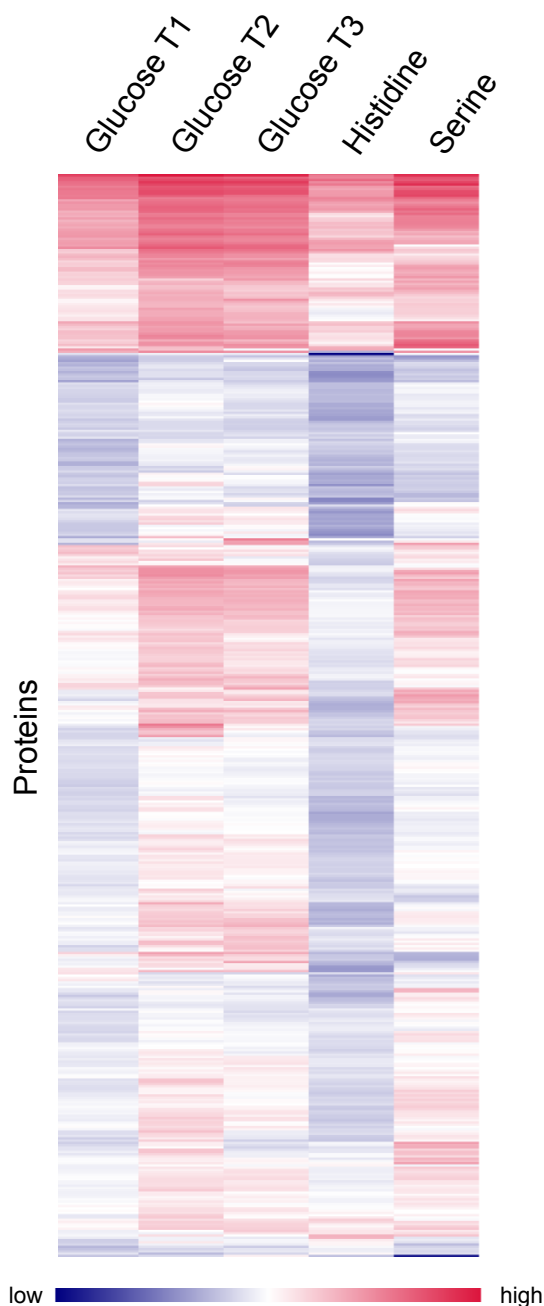


Figure 4-9. Carbon source dependent quantitative proteome profiles. The heatmap reflects the non-normalized abundance data for the 482 most well represented proteins across all samples. *A. hydrogeniformans* were grown with glucose and harvested at three timepoints (T1, T2, and T3), or grown with histidine or serine, and harvested. Samples were processed by eFASP-HILIC and analyzed by LC-MS^E. Quantitative information for each protein was determined by the average protein intensity in 2 technical replicates. Crimson indicates high protein abundance, while navy represents low protein abundance in a sample.

Anaerobaculum hydrogeniformans monocultures were grown with glucose, histidine, or serine. The cells from each condition were harvested, and washed with PBS. Cell pellets were processed by eFASP, in which proteins were extracted and solubilized in strong denaturing buffers. Glycogen phosphorylase b (rabbit) was spiked in as a quantification standard, and then the proteins were alkylated with 4-VP. SDS and urea were replaced with 0.2% DCA by buffer exchange, and then the proteins were digested with trypsin.

Peptides produced were either analyzed by LC-MS^E (shotgun), or separated into 5 fractions, by HILIC in TopTip cartridges, and then analyzed by LC-MSE. LC-MS^E RAW data files were processed in ProteinLynx Global Server for peptide and protein identification and quantification information. Fractionated samples required merging of processed data files to provide the search engine peptide information from all 5 fractions from which to base protein identifications. DECO was utilized to assure and to correct proper merged peptide ion volumes.

Unfractionated samples were analyzed by shotgun LC-MS^E in triplicate. An average of 380 proteins was identified in each sample, based on 7,990 peptide identifications. Each HILIC fraction of each fractionated sample was analyzed by LC-MS^E in duplicate. The merged raw file for each sample identified approximately 718 proteins based on 26,700 peptide identifications. 482 proteins that are identified and quantified across all replicate HILIC-LC-MS^E analyses are presented in a heatmap (Figure 4-9). Apparent are the significant changes in the latter two glucose timepoints, as well as smaller, but significant, up and down regulation of proteins that differentiate the glucose samples from the histidine and the serine samples.

Unfractionated samples were quantified in Expression Analysis, using trypsin and glycogen phosphorylase b as standards for normalization. All ratios, and statistics, were calculated for each sample using glucose condition at the first timepoint as a reference.

Analysis of the glucose timecourse revealed 101 proteins that were upregulated at later timepoints, with p-values above 0.95, and natural log ratios above 0.30, and 33 proteins downregulated, with ln ratio values below -0.30. There were 44 proteins that could only be identified in timepoints T2 and T3, and there were 101 proteins that could only be identified in timepoint T3 and not timepoints T1 or T2.

Analysis of the serine condition revealed 113 proteins upregulated in comparison to glucose T1, with p-values greater than 0.95 and ln ratio values greater than 0.30. There were 50 proteins unique to the serine condition. Analysis of the histidine condition revealed 56 proteins upregulated in comparison to glucose T1, with p-values greater than 0.95 and ln ratio values greater than 0.30. There were 30 proteins unique to the histidine condition.

Glucose: carbohydrate metabolism

Unusual pathways for carbohydrate utilization have been documented in extremophiles¹⁷. The degradation of glucose by *A. hydrogeniformans* strain OS-1 has been attributed to a modified Embden-Meyerhof-Parnas (EMP) pathway, utilizing ferredoxin-dependent metabolism to generate high levels of hydrogen^{18,19}. In this pathway, glyceraldehyde 3-phosphate dehydrogenase (NAD⁺ linked) is replaced by glyceraldehyde 3-phosphate ferredoxin oxidoreductase (GAPOR). GAPOR is an

oxygen-sensitive iron and tungsten-containing protein that catalyzes the reduction of ferredoxin with the oxidation of glyceraldehyde 3-phosphate^{20,21}. Pyruvate-ferredoxin oxidoreductase catalyzes the degradation of pyruvate formed by this pathway¹⁸.

The comparison of cells incubated with glucose for different periods of time, reveals significant up regulation in enzymes involved in carbohydrate degradation, as well as transcription factors that mediate long-term cellular effects (Table 4-2).

Key enzymes in glycolysis are upregulated with glucose as the main carbon and energy source, including the enzymes that follow. Glucose 6-phosphate isomerase (GPI) is responsible for the isomerization of glucose 6-phosphate to fructose 6-phosphate. 6-phosphofructokinase (6-PFK) is responsible for the phosphorylation of fructose 6-phosphate, using ATP, to fructose 1,6-bisphosphate. Triose phosphate isomerase (TIM) interconverts glyceraldehyde 3-phosphate and dihydroxyacetone phosphate. Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) oxidizes glyceraldehyde 3-phosphate to 1,3-diphosphoglycerate. Phosphoglycerate mutase (PGM) moves the phosphate group from carbon 3 to carbon 2. Enolase (ENO) removes water from 2-phosphoglycerate. These data are corroborated by metabolic, and enzymatic studies which show activity of GPI, 6-PFK, ENO in *A. hydrogeniformans* strain OS-1, in the presence of glucose¹⁸. However, in those studies, they did not detect activity for GAPDH when NAD⁺ or NADP⁺ were used as electron acceptors, which led them to conclude that GAPOR replaced GAPDH in this context¹⁸.

Pyruvate is the major intermediate whose fate must be decided. Multiple enzymes are upregulated during glucose utilization influence the fate of pyruvate, and its degradation to a wide variety of end products (Table 4-2). Pyruvate phosphate

dikinase (PPDK) phosphorylates pyruvate into phosphoenolpyruvate, utilizing ATP, and may be involved in the funneling of carbon into storage as starch or fatty acids, as it is upregulated in plants during stress, e.g., high salt, drought, or low-oxygen^{22,23}. Alcohol dehydrogenase (ADH), in yeast, converts acetaldehyde, produced from pyruvate cleavage, into ethanol, using of ethanol as electron donor and NADP⁺ as electron acceptor¹⁸. Formate acetyltransferase catalyzes the reversible conversion of pyruvate and co-A into acetyl-CoA and formate. Pyruvate ferredoxin oxidoreductase (PFOR) catalyzes the conversion of pyruvate, Coenzyme-A and oxidized ferredoxin, into acetyl-CoA, CO₂, reduced ferredoxin, and H⁺. These data are corroborated by metabolic, and enzymatic studies which show activity of ADH and PFOR in *A. hydrogeniformans* strain OS-1, in the presence of glucose¹⁸.

No glyceraldehyde 3-phosphate ferredoxin oxidoreductase (GAPOR) entry exists in the current protein database; however we identify a few aldehyde ferreredoxin oxidoreductases as being upregulated following incubation with glucose (Table 4-3). GO annotations for each protein include 'electron carrier activity', 'iron-sulfur cluster binding', and 'oxidoreductase activity, acting on the aldehyde or oxo group of donors, iron-sulfur protein as acceptor'. Alignment to known GAPOR in different phyla does not show clear sequence homology.

Energy storage

Trehalose is a disaccharide, synthesized from two molecules of glucose, which can be found in many organisms, including bacteria, insects, and plants. This non-reducing sugar has been associated with the ability to endure desiccation, to respond to

osmotic stress, and to serve as a stored energy source ²⁴. Two enzymes are responsible for trehalose biosynthesis: trehalose 6-phosphate synthase (TPS) and trehalose-6-phosphate phosphatase (TPP) ²⁵. *A. hydrogeniformans* upregulates a protein which has the activity of both enzymes, TPS and TPP (D3L045), following glucose incubation (ln ratio (GluT2/T1) = 0.53, p-value=0.99; ln ratio (GluT3/T1) = 0.85, p-value=1.00).

Glycolysis

Glycolysis		Quantification				Unique
		GluT2/T1		GluT3/T1		
		ln ratio	p-val	ln ratio	p-val	
Accession	Name					
D3L0E0	Glucose 6 phosphate isomerase	0.85	1.00	1.09	1.00	
D3L608	6 phosphofructokinase isozyme 1 Fragment					GluT3
D3L1F8	6 phosphofructokinase pyrophosphate dependent					GluT3
D3L1Q3	Triosephosphate isomerase	0.43	1.00	0.71	1.00	
D3L1Q5	Glyceraldehyde 3 phosphate dehydrogenase type I	0.35	1.00	0.58	1.00	
D3L326	Phosphoglycerate mutase					GluT3
D3L5C5	Phosphoglycerate mutase					GluT3
D3L489	Enolase	0.29	0.99	0.63	1.00	

Fate of Pyruvate

AccessionName		Quantification				Unique
		GluT2/T1		GluT3/T1		
		ln ratio	p-val	ln ratio	p-val	
D3L3U9	Pyruvate phosphate dikinase					GluT3
D3L5N0	Pyruvate ferredoxin oxidoreductase Fragment					GluT3
D3L5N2	Pyruvate ferredoxin oxidoreductase Fragment					GluT2/T3
D3L4G2	Formate acetyltransferase	0.49	1.00	0.84	1.00	
D3L0T9	Alcohol dehydrogenase IV	0.37	1.00	0.65	1.00	
D3L1A8	Alcohol dehydrogenase	0.67	1.00	0.91	1.00	

Table 4-2. Metabolic effects of glucose incubation. Expression analysis was used to compare the quantity of each protein across the glucose incubation timepoints (GluT1, GluT2, and GluT3), with data normalized to glycogen phosphorylase b and trypsin abundance. 101 proteins that were upregulated at later timepoints (GluT2 and GluT3), with p-values above 0.95, and natural log ratios above 0.30, and 33 proteins downregulated, with natural log ratios below -0.30. There were 44 proteins that could only be identified in timepoints T2 and T3, and there were 101 proteins that could only be identified in timepoint T3 and not timepoints T1 or T2. Presented in this table are enzymes found upregulated in the later glucose timepoints, and grouped by participation in glycolysis or in pyruvate metabolism.

Aldehyde ferredoxin oxidoreductases identified

In ratios

	T2/T1	T2/T1	MW	Notes
D3L4E6	0.29	0.51	22,179	fragment
D3L4E7	0.16	0.37	37,641	
D3L0G4	0.20	0.45	81,440	tungsten containing
D3L3V9		0.47	5,693	tungsten containing
D3L3W0	GluT2	GluT3	6,455	

Table 4-3. Aldehyde ferredoxin oxidoreductases identified as potential candidates as GAPOR. The *A. hydrogeniformans* protein database contains no GAPOR entries; however, GAPOR, rather than GAPDH, activity has been proposed as the enzyme responsible for G3P degradation. Assuming that GAPOR exists in the database, but is not designated properly, we present five aldehyde ferredoxin oxidoreductases that are upregulated with glucose incubation, and maintain relevant functional annotations, as candidates for GAPOR.

Glucose: gene regulation

Wide-scale control of gene expression and durable functional changes can be affected by transcription factor activity. In *A. hydrogeniformans*, the incubation with glucose elicits the expression of transcription factors with varying target genes and processes, including Rex, GntR, IclR, TrpR, and LuxR (Table 4-4).

The Rex transcription factor is a sensor for NADH/NAD⁺ ²⁶. In *Staphylococcus aureus*, Rex mediates transcriptional control of genes involved in anaerobic metabolism, allowing for adaptation to changing host physiology ²⁷. The overall function of Rex is to regenerate NAD⁺ and to synthesis ATP, which can be achieved through upregulating fermentation enzymes and transporters that remove eliminate the byproducts. A number of proteins that are found upregulated in *A. hydrogeniformans* with glucose incubation, overlap with Rex target genes, including: alcohol dehydrogenase, formate acetyltransferase, glucose 6-phosphate isomerase, transketolase, pyruvate ferreredoxin oxidoreductase, and NADH dehydrogenase ²⁷.

Gluconate utilization system GNT-I transcriptional repressor (GntR) is involved in gluconate metabolism²⁸. In *Corynebacterium glutamicum*, GntR controls gnt expression based on the availability of gluconate and an inhibitory carbohydrate like glucose²⁹. GntR also represses transcription of ptsG, which is involved in glucose uptake, so in the absence of gluconate and the presence of glucose, GntR effects on glucose transport are abrogated²⁹. The upregulation of GntR may occur in the anticipation of gluconate production, in the context of glucose abundance, which would eventually inhibit glucose import.

The glyoxylate pathway allows cells to utilize simple carbon compounds as a carbon source, when glucose is not available. Isocitrate lyase regulator (IclR) represses the expression of the genes required for the glyoxylate pathway in the presence of glucose³⁰. In *E. coli*, the deletion of the IclR gene prevents repression of the glyoxylate pathway, resulting in the redirection of isocitrate to succinate and malate without carbon dioxide generation³¹. In *A. hydrogeniformans*, grown with glucose, this upregulation of IclR is logical, since alternative carbon sources are not required for metabolism.

TrpR is a repressor of L-tryptophan synthesis. TrpR is activated when bound to L-tryptophan, and represses the trp operon as well as its own regulatory region. In *E. coli*, glucose can be used for L-tryptophan biosynthesis, and the TrpR functions as feedback inhibition³². This is likely the purpose in *A. hydrogeniformans*, which has plentiful glucose for biosynthesis, and is TrpR is functioning to shut off excessive L-tryptophan production.

One of the more curious transcription regulators is LuxR that is an activator of quorum sensing and sensory transduction³³. Bacterial populations can use signaling

molecules for the regulation of population density and the coordination of gene expression³⁴. Microfluidic studies of bacteria confined to various sized volumes show that bacteria in smaller volumes can quickly reach high cell densities, even when a single cell was used which indicates that the inability to culture bacteria, inoculated below a certain density, may be a result of sensory signaling^{35,36}. The upregulation of LuxR while energy sources are high may be related to future growth or population movement.

Transcription Regulation

Transcription Regulation		Quantification				Unique
		GluT2/T1		GluT3/T1		
		In ratio	p-val	In ratio	p-val	
D3L1X9	TrpR protein	0.33	1.00	0.60	1.00	
D3KZM8	Transcriptional regulator GntR family	0.80	1.00	1.20	1.00	
D3L4Q8	Transcriptional regulator lclR family					GluT2/T3
D3L215	DNA binding response regulator LuxR family					GluT3
D3L0X5	Toxin antitoxin system antitoxin cmpnt Xre family					GluT3
D3L488	Redox sensing transcriptional repressor rex					GluT2/T3

Table 4-4. Transcription factor upregulation in response glucose incubation. Expression analysis was used to compare the quantity of each protein across the glucose incubation timepoints (GluT1, GluT2, and GluT3), with data were normalized to glycogen phosphorylase b and trypsin abundance.

Serine: amino acid degradation and conversion

Serine is not a typical carbon source for cells, and a number of enzymes that catalyze its degradation are found expressed only in the absence of glucose. These proteins are constituents of multiple pathways for the degradation or conversion of serine for energy and as a carbon source (Table 4-5). Serine is one of three three-carbon glucogenic amino acids that can be converted into pyruvate. The most direct route to pyruvate is the deamidation of serine by L-serine ammonia-lyase³⁷; however, this is not the only path to pyruvate.

Serine can be converted into glycine by serine hydroxyl methyltransferase 1 (SHMT1) by hydroxymethyl removal. Glycine produced either be funneled for other catabolic processes, like purine synthesis, or it can be degraded in the glycine cleavage system, in which glycine decarboxylase (GDC) catalyzes the removal of the carboxyl group from glycine, to produce carbon dioxide and ammonia, and N5,N10-methylene-THF. The glycine cleavage system is stimulated by excess concentrations of glycine. This implies that in the presence of serine-rich media, Serine is being converted into glycine by SHMT1, and the increase in glycine upregulates the pathway responsible for handling excess glycine.

Serine to pyruvate

Accession Name	Ser/GluT1	
	In ratio	p-val
D3L128	Pyridoxal phosphate dependent serine hydroxymethyltransferase 1	0.86 1.00

Serine to glycine

Accession Name	Ser/GluT1	
	In ratio	p-val
D3L649	Glycine dehydrogenase Decarboxylating subunit 2 Fragment	0.53 0.99
D3L3M7	Glycine dehydrogenase Decarboxylating subunit 2	0.46 1.00
D3L3M8	Glycine dehydrogenase OS Anaerobaculum hydrogeniformans	0.77 1.00
D3L2W8	Glycine reductase complex component C beta subunit	0.56 1.00
D3L4I5	Probable glycine dehydrogenase decarboxylating	0.45 0.98
D3L4I6	Probable glycine dehydrogenase decarboxylating subunit 2	0.76 1.00
D3L6A6	Glycine reductase B Fragment	Unique to serine

Serine to 3-phosphoglycerate

Accession Name	Ser/GluT1	
	In ratio	p-val
D3L1Q4	Phosphoglycerate kinase	0.58 1.00

Serine to Threonine

Accession Name	Ser/GluT1	
	In ratio	p-val
D3L133	Threonine synthase	Unique to serine
D3L193	4 hydroxythreonine 4 phosphate dehydrogenase	Unique to serine

Table 4-5. Growth with Serine results in the upregulation of enzymes responsible serine metabolism. Expression Analysis was used to compare the quantity of each protein in the serine condition (SER) and in the glucose T1 condition (GluT1), with data normalized to glycogen phosphorylase b and trypsin abundance. 113 proteins upregulated in comparison to glucose T1, with p-values greater than 0.95 and In ratio values greater than 0.30. There were 50 proteins unique to the serine condition. Proteins relevant to serine metabolism are displayed and grouped based on pathway relationship.

Serine can be synthesized from 3-phosphoglycerate in three steps. The reversal of these steps, by different enzymes, can convert serine back into a glycolytic intermediate, and eventually into pyruvate³⁸. Serine-glyoxylate transaminase (SGT) converts serine into hydroxypyruvate through transamination. Hydroxypyruvate is then converted into glycerate by hydroxypyruvate reductase (HPR). Phosphorylation of glycerate by glycerate kinase (GK) produces 3-phosphoglycerate, which can then be processed in glycolysis by phosphoglycerate kinase (PGK).

Homoserine can be converted into threonine by threonine synthase (TS). In HILIC experiments, it was upregulated (ln ratio (SER/GluT1) = 0.74), and another Threonine synthase was identified, D3L2V7.

Histidine: amino acid degradation

A. hydrogeniformans grown with histidine as the carbon and energy source provided the least protein yield for analysis. This may indicate its inability to effectively survive with histidine. Histidine is a glucogenic amino acid, since the end product of catabolism is glutamate. Anaerobic bacteria can ferment glutamate, producing carbon dioxide, acetate, butyrate, ammonia, and small amounts of hydrogen^{39,40}.

There are four steps for the conversion of histidine to glutamate, and the first two are upregulated in *A. hydrogeniformans* grown with histidine (Table 4-6). Histidine ammonia-lyase converts histidine to urocanate through deamination of the α -amino group³⁸. Urocanate is hydrated into 4-imidazolone-5-propionate by urocanate hydratase³⁸. Imidazolone propionase hydrolyzes the carbon-nitrogen bond, and

formiminotransferase cyclodeaminase removes the formimino group to produce glutamate³⁸.

Accession Name		Hist/GluT1	
		ln ratio	p-val
D3L2B9	Histidine ammonia lyase	0.79	0.95
D3L1C8	Urocanate hydratase	1.13	1.00

Table 4-6. Growth with histidine results in the upregulation of enzymes responsible histidine catabolism. Expression analysis was used to compare the quantity of each protein in the histidine condition and in the glucose T1 condition, with data were normalized to glycogen phosphorylase b and trypsin abundance. Analysis of the histidine condition revealed 56 proteins upregulated in comparison to glucose T1, with p-values greater than 0.95 and ln ratio values greater than 0.30. There were 30 proteins unique to the histidine condition. Proteins relevant to histidine metabolism are displayed.

REFERENCES

1. Maune, M. W. & Tanner, R. S. Description of *Anaerobaculum hydrogeniformans* sp. nov., an anaerobe that produces hydrogen from glucose, and emended description of the genus *Anaerobaculum*. *Int J Syst Evol Microbiol* **62**, 832–838 (2012).
2. Duncan, K. E. *et al.* Biocorrosive thermophilic microbial communities in Alaskan North Slope oil facilities. *Environ Sci Technol* **43**, 7977–7984 (2009).
3. Masuda, T., Tomita, M. & Ishihama, Y. Phase transfer surfactant-aided trypsin digestion for membrane proteome analysis. *J Proteome Res* **7**, 731–740 (2008).
4. Yeung, Y. G., Nieves, E., Angeletti, R. H. & Stanley, E. R. Removal of detergents from protein digests for mass spectrometry analysis. *Anal Biochem* **382**, 135–137 (2008).
5. Li, G. Z. *et al.* Database searching and accounting of multiplexed precursor and product ion spectra from the data independent analysis of simple and complex peptide mixtures. *Proteomics* **9**, 1696–1719 (2009).
6. Silva, J. C., Gorenstein, M. V., Li, G. Z., Vissers, J. P. & Geromanos, S. J. Absolute quantification of proteins by LCMS^E: a virtue of parallel MS acquisition. *Mol Cell Proteomics* **5**, 144–156 (2006).
7. Silva, J. C. *et al.* Simultaneous qualitative and quantitative analysis of the *Escherichia coli* proteome: a sweet tale. *Mol Cell Proteomics* **5**, 589–607 (2006).
8. Silva, J. C. *et al.* Quantitative proteomic analysis by accurate mass retention time pairs. *Anal Chem* **77**, 2187–2200 (2005).
9. Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J Mol Biol* **157**, 105–132 (1982).
10. Dimmer, E. C. *et al.* The UniProt-GO Annotation database in 2011. *Nucleic Acids Res* **40**, D565–D570 (2012).
11. Binns, D. *et al.* QuickGO: a web-based tool for Gene Ontology searching. *Bioinformatics* **25**, 3045–3046 (2009).
12. Wisniewski, J. R., Zougman, A. & Mann, M. Combination of FASP and StageTip-based fractionation allows in-depth analysis of the hippocampal membrane proteome. *J Proteome Res* **8**, 5674–5678 (2009).
13. Wisniewski, J. R., Nagaraj, N., Zougman, A., Gnad, F. & Mann, M. Brain phosphoproteome obtained by a FASP-based method reveals plasma membrane protein topology. *J Proteome Res* **9**, 3280–3289 (2010).

14. Zielinska, D. F., Gnad, F., Jedrusik-Bode, M., Wisniewski, J. R. & Mann, M. *Caenorhabditis elegans* has a phosphoproteome atypical for metazoans that is enriched in developmental and sex determination proteins. *J Proteome Res* **8**, 4039–4049 (2012).
15. Kuntumalla, S. *et al.* Comparison of two label-free global quantitation methods, APEX and 2D gel electrophoresis, applied to the *Shigella dysenteriae* proteome. *Proteome Sci* **7**, 22 (2009).
16. Bereman, M. S., Egertson, J. D. & MacCoss, M. J. Comparison between procedures using SDS for shotgun proteomic analyses of complex samples. *Proteomics* **11**, 2931–2935 (2011).
17. Sakuraba, H., Goda, S. & Ohshima, T. Unique sugar metabolism and novel enzymes of hyperthermophilic archaea. *Chem Rec* **3**, 281–287 (2004).
18. Maune, M. *Anaerobaculum hydrogeniformans*: Insights into the novel biochemical and physiological ability of a thermophilic biocatalyst that can utilize both defined and undefined substrates for the production of renewable hydrogen energy. The University of Oklahoma, 1–190 (2011).
19. McInerney, M. J. *et al.* Novel hydrogen production systems operative at thermodynamic extremes. *Genomic Science Meeting 2011* 1–2 (2011).
20. Park, M. O., Mizutani, T. & Jones, P. R. Glyceraldehyde-3-phosphate ferredoxin oxidoreductase from *Methanococcus maripaludis*. *J Bacteriol* **189**, 7281–7289 (2007).
21. Mukund, S. & Adams, M. W. Glyceraldehyde-3-phosphate ferredoxin oxidoreductase, a novel tungsten-containing enzyme with a potential glycolytic role in the hyperthermophilic archaeon *Pyrococcus furiosus*. *J Biol Chem* **270**, 8389–8392 (1995).
22. Hatch, M. D. & Slack, C. R. A new enzyme for the interconversion of pyruvate and phosphopyruvate and its role in the C₄ dicarboxylic acid pathway of photosynthesis. *Biochem J* **106**, 141–146 (1968).
23. Liu, C. W. *et al.* Proteomic analysis of salt-responsive ubiquitin-related proteins in rice roots. *Rapid Commun Mass Spectrom* **26**, 1649–1660 (2012).
24. Iturriaga, G., Suárez, R. & Nova-Franco, B. trehalose metabolism: from osmoprotection to signaling. *Int J Mol Sci* **10**, 3793–3810 (2009).
25. Cabib, E. & Leloir, L. F. The biosynthesis of trehalose phosphate. *J Biol Chem* **231**, 259–275 (1958).
26. Brekasis, D. & Paget, M. S. A novel sensor of NADH/NAD⁺ redox poise in *Streptomyces coelicolor* A3(2). *EMBO J* **22**, 4856–4865 (2003).

27. Pagels, M. *et al.* Redox sensing by a Rex-family repressor is involved in the regulation of anaerobic gene expression in *Staphylococcus aureus*. *Mol Microbiol* **76**, 1142–1161 (2010).
28. Tong, S., Porco, A., Isturiz, T. & Conway, T. Cloning and molecular genetic characterization of the *Escherichia coli* gntR, gntK, and gntU genes of GntI, the main system for gluconate metabolism. *J Bacteriol* **178**, 3260–3269 (1996).
29. Frunzke, J., Engels, V., Hasenbein, S., Gatgens, C. & Bott, M. Co-ordinated regulation of gluconate catabolism and glucose uptake in *Corynebacterium glutamicum* by two functionally equivalent transcriptional regulators, GntR1 and GntR2. *Mol Microbiol* **67**, 305–322 (2008).
30. Lorca, G. L. *et al.* Glyoxylate and pyruvate are antagonistic effectors of the *Escherichia coli* IclR transcriptional regulator. *J Biol Chem* **282**, 16476–16491 (2007).
31. Waegeman, H. *et al.* Effect of iclR and arcA knockouts on biomass formation and metabolic fluxes in *Escherichia coli* K12 and its implications on understanding the metabolism of *Escherichia coli* BL21 (DE3). *BMC Microbiol* **11**, 70 (2011).
32. Zhao, Z. J. *et al.* Development of L-tryptophan production strains by defined genetic modification in *Escherichia coli*. *J Ind Microbiol Biotechnol* **38**, 1921–1929 (2011).
33. Tsai, C. S. & Winans, S. C. LuxR-type quorum-sensing regulators that are detached from common scents. *Mol Microbiol* **77**, 1072–1082 (2010).
34. Nasser, W. & Reverchon, S. New insights into the regulatory mechanisms of the LuxR family of quorum sensing regulators. *Anal Bioanal Chem* **387**, 381–390 (2007).
35. Boedicker, J. Q., Vincent, M. E. & Ismagilov, R. F. Microfluidic confinement of single cells of bacteria in small volumes initiates high-density behavior of quorum sensing and growth and reveals its variability. *Angew Chem Int Ed Engl* **48**, 5908–5911 (2009).
36. Stewart, E. J. Growing unculturable bacteria. *J Bacteriol* **194**, 4151–60 (2012).
37. Alfoldi, L., Rasko, I. & Kerekes, E. L-serine deaminase of *Escherichia coli*. *J Bacteriol* **96**, 1512–1518 (1968).
38. Moat, A. G., Foster, J. W. & Spector, M. P. *Microbial Physiology*. John Wiley & Sons: 2002.
39. Buckel, W. Unusual enzymes involved in five pathways of glutamate fermentation. *Appl Microbiol Biotechnol* **57**, 263–273 (2001).

40. Buckel, W. & Barker, H. A. Two pathways of glutamate fermentation by anaerobic bacteria. *J Bacteriol* **117**, 1248–1260 (1974).

CHAPTER 5

***Syntrophus aciditrophicus* adapts its metabolic machinery for growth with benzoate, crotonate, or cyclohexane carboxylate.**

ABSTRACT

The global cycling of carbon is dependent on the anaerobic degradation of biomaterials into carbon dioxide and methane. Highly specialized anaerobic bacteria and archaea effect the degradation of biomaterials syntrophically. *S. aciditrophicus* is an anaerobic bacterium that can degrade benzoate, and other organic matter, independently or in syntrophic association hydrogen/formate consuming methanogens, like *M. hungatei*. Studying the metabolic pathways utilized for the degradation of different materials will provide insight in alternative catabolic pathways and mechanisms employed for survival in such poor energy economies.

Co-cultures of *S. aciditrophicus* and *M. hungatei* were grown with 20 mM crotonate, 5 mM benzoate, or 5 mM cyclohexane carboxylate. *S. aciditrophicus* cells were isolated and processed by eFASP to completely solubilize and digest cytosolic and membrane associated proteins. Shotgun analysis of each digest identified an average of 242 proteins, based on 4,800 peptide identifications. Pre-fractionation by HILIC, enabled a significant increase in proteome coverage, with 513 protein identifications, based on 15,000 peptide identifications per analyzed digest. 149 proteins were quantified across all sample replicate analyses, and 995 unique proteins were quantified in at least one sample condition.

TABLE OF CONTENTS

INTRODUCTION.....	137
EXPERIMENTAL.....	138
Strains and cell growth	138
Enhanced filter aided sample preparation of <i>S. aciditrophicus</i>	138
Peptide fractionation by HILIC	140
Analysis by nanoLC-MS ^E	141
Data processing and protein identification	142
RESULTS AND DISCUSSION	143
REFERENCES.....	154

INTRODUCTION

S. aciditrophicus is a gram negative, rod-shaped, bacterium that is essential for the first step in the anaerobic conversion of organic matter in the environment to methane and carbon dioxide ¹. Found in sewage sludge, this anaerobic bacterium is capable of consuming a variety of organic substrates, like alcohols, fatty acids, sugars, and hydrocarbons, and producing acetate, bicarbonate, and hydrogen ¹. Degradation of these organic materials is dependent upon the presence of microbial species that can consume the degradation products, and drive the otherwise thermodynamically unfavorable reaction forward. It is thought that *S. aciditrophicus* is a missing link to understanding global carbon cycles and important in bioremediation of contaminants ^{2,3-5}.

S. aciditrophicus has limited ability to utilize external electron acceptors, like oxygen, nitrate, sulfate, and must exist in a state of syntrophy, in which fermentive and methanogenic organisms feed off of its metabolic products. Membrane proteins play a key role in the syntrophic metabolism, creating and utilizing ion gradients, and allowing for interspecies electron transfer ². *S. aciditrophicus* can utilize benzoate as an electron acceptor when grown in mono-culture.

S. aciditrophicus can grow in mono-culture conditions, without help from other organisms, when incubated with crotonate and benzoate. Crotonate fermentation yields acetate and cyclohexane carboxylate (with 99.4% carbon recovery and 95.8% hydrogen recovery), using benzoyl CoA as an intermediate ^{6,7}. When fed benzoate, cyclohexane carboxylate is a product, which presumably occurs through the entry of

benzoate, as benzoyl-CoA, into the crotonate degradation pathway^{3,6,8}. Although, this is based on genomic information⁶.

The aim of this work is to define the enzymes and pathways that *S. aciditrophicus* relies on in order to degrade crotonate, benzoate, and cyclohexane, in syntrophy with *M. hungatei*. To achieve this goal, we utilize a harsh mass-spectrometry compatible sample preparation method (eFASP), and label-free quantitative mass spectrometry, to collect a comprehensive inventory of proteins and the degree of differential regulation based on growth condition

EXPERIMENTAL

Strains and cell growth

Syntrophus aciditrophicus was co-cultured with *M. hungatei* JF-1. Cultures were grown with 20 mM crotonate, 5 mM benzoate, or 5 mM cyclohexane carboxylate. *S. aciditrophicus* cells were isolated from *M. hungatei* cells by Percol separation. Harvested cells from the different conditions were washed with PBS three times, and the pellets weighed. Sample pellets were processed by eFASP.

Enhanced filter aided sample preparation of S. aciditrophicus

Washed cell pellets were suspended in Solubilization Buffer A (4% ammonium dodecyl sulfate, 50 mM TCEP, 0.1% deoxycholic acid, 0.1% n-octylglucoside, 100 mM Tris-HCL, pH 8), and incubated on a heated block for 10 minutes at 90°C. The sample was sonicated for three 10-second intervals with a Sonic Dismembrator Model 100 (Fisher Scientific), and then spun at 16,000 rcf for 10 minutes in an Eppendorf

centrifuge; this step was repeated once. Glycogen phosphorylase B (rabbit) was added to each sample at a 1:50 ratio. The supernatant and any residual pellet, was sonicated and allowed to cool to 37°C. Proteins were alkylated by the addition of 4-vinylpyridine (4-VP) at 35 mM final concentration, and incubated for 1 hour on a shaker at 37°C. DTT was added to a final concentration of 50 mM to deactivate residual 4-VP. Based on a 20 mg:1 mg wet weight:protein estimate, 250 µg of each sample was transferred to a new tube for further processing.

Eight volumes of Solubilization Buffer B (8 M urea, 0.1% deoxycholic acid, 0.1% n-octylglucoside, 100 mM Tris-HCL, pH 8) was added to the sample to begin buffer exchange of SDS. The diluted sample was applied to a Passivated Microcon Filter Unit (described below), and spun at 14,000 rcf for 10 minutes. The eluate was discarded, and 200 µl of Solubilization Buffer B was added to the filter unit and then spun at 14,000 rcf for 10 minutes. This step was repeated twice. To remove urea, three buffer exchange steps were performed with Digestion Buffer (0.2% deoxycholic acid in 50 mM ABC). The Passivated Microcon Filter Unit (30,000 Da cutoff), and collection tube, were pre-incubated in 5% Tween-20 overnight, and then thoroughly rinsed with large volumes of MS-grade water.

Digestion Buffer (100 µl) was added to the Microcon unit, followed by the addition of 5 µg of trypsin (1 µg trypsin : 50 µg protein). Digestion proceeded for 12 hours on a shaker set to 37°C. The Microcon unit was transferred to a new Passivated Collection Tube, and spun at 14,000 rcf for 10 minutes. To assure complete removal of peptides, 100 µl of 50 mM ABC buffer was added to the Microcon unit, and collected by centrifugation, two times.

Removal of deoxycholic acid from the digest was accomplished by acidification and phase transfer of insoluble deoxycholate to an ethyl acetate layer ⁹. Instead of removing the aqueous layer as in Masuda et al. ⁹, the organic layer was removed, and then two more rounds of addition and removal of ethyl acetate facilitate the removal of deoxycholate from the solution, similar to the strategy Yeung et al. employ for the removal of n-octylglucoside ¹⁰.

Collected peptides were dried down in a SpeedVac and re-suspended in 50% methanol, three times. Dried peptides were suspended in modified Mobile Phase A (96.5% H₂O/3% ACN/0.5% formic acid). The peptide solutions were either used for shotgun LC-MS^E analysis, or fractionated by HILIC before LC-MS^E analysis.

Peptide fractionation by HILIC

Binding Buffer: 90% ACN, 15 mM ammonium formate, pH 3.

Release1 Buffer: 90% ACN, 15 mM ammonium formate, pH 3.

Release2 Buffer: 80% ACN, 15 mM ammonium formate, pH 3.

Release3 Buffer: 78% ACN, 15 mM ammonium formate, pH 3.

Release4 Buffer: 74% ACN, 15 mM ammonium formate, pH 3.

Release5 Buffer: 40% ACN, 15 mM ammonium formate, pH 3.

Condition Buffer: 10% ACN, 15 mM ammonium formate, pH 3.

TopTips (TT200HIL, PolyLC Inc., Columbia, MD) were conditioned by wetting the material with Condition Buffer, followed by incubation with Binding Buffer. Each sample (70 µg) was dried down to remove any acids or salts, and re-suspended in a small volume of 150 mM ammonium formate, pH 3. This solution was added to the TopTip,

along with enough ACN to bring it to 90% organic. The peptide solution was passed through the TopTip by centrifugation at 500 rcf, and the eluate was passed through the TopTip again. All subsequent steps were done twice. The TopTip was washed with Binding Buffer. Peptides were eluted, into separate fractions, with Release1 Buffer through Release5 Buffer. All fractions were dried down in a SpeedVac, twice. Peptides were re-suspended in 9 modified Mobile Phase A (96.5% H₂O/3% ACN/0.5% formic acid), for LC-MS^E analysis.

Analysis by nanoLC-MS^E

nanoUPLC of sample peptides was performed on an Waters nanoACQUITY UPLC system equipped with a nanoACQUITY UPLC Symmetry C18 180 µm x 20 mm trap column (Waters, 186003514), and a nanoACQUITY UPLC BEH C18 75 µm x 250 mm reversed phase analytical column (Waters, 186003545). The aqueous mobile phase (Mobile Phase A) contained 0.1% formic acid in water and the organic mobile phase (Mobile Phase B) contained 0.1% formic acid in acetonitrile. 5 µl of the sample was loaded onto the trap with and washed with 99% A and 1% B for 5 minutes, and then injected onto the column with 3% Mobile Phase B. Peptides were eluted from the column with a gradient of 3-40% B over 90 minutes at 300 nl/min, followed by a 5 minute rinse with 95% B. The column was re-equilibrated to 97% A for 20 minutes.

The lock mass compound, [Glu¹]fibrinopeptide (0.5 pmol/µl, 25% ACN, 0.1% formic acid), and the analyte were delivered at 300 nl/min to the NanoLockSpray dual electrospray ion source (Waters, UK) which was equipped with a microcapillary for the lockmass compound, and a PicoTip Emitter (New Objective) for the analyte. Two

electrosprays are produced which may be selectively sampled by positioning of the baffle.

The ion source was interfaced with a Waters Synapt HDMS for mass spectrometric analysis. Acquisition was performed in positive nanoESI mode and the instrument was operated in V-mode with an average resolution of 9500. Accurate mass data for precursor and fragment ions were collected over the 100-2000 m/z range, with the mode of acquisition alternating between low and elevated energy at 1-second intervals, for the duration of analysis. The lock mass channel was sampled every 30 seconds for calibration of MS and MS^E data.

Data processing and protein identification

Raw data files were processed in ProteinLynx Global Server 2.5.2 (PLGS, Waters) to provide an inventory of precursor ions along with their respective fragment ions, to search against a combined database composed of the Uniprot HAMAP complete proteome for *Syntrophus aciditrophicus* (strain SB, August 2011) and the Uniprot HAMAP complete proteome for *Methanospirillum hungatei* (strain JF-1 / DSM 864, August 2011), for peptide and protein identifications¹¹. Database search settings included two missed cleavages, a 4% false positive rate, fixed S-pyridylethyl modification of cysteine residues, and variable oxidation modification of methionine residues. The Waters Expression Informatics component of the PLGS software was utilized to quantify data using glycogen phosphorylase B (Uniprot accession P00489) as the quantification reference¹²⁻¹⁴. Data Extraction and Correction (DECO) software was used to calculate isoelectric points and grand average of hydropathicity index (GRAVY)

values for peptides and proteins identified. GRAVY values were calculated based on the hydrophobicity of constituent amino acids derived from Kyte and Doolittle¹⁵. Uniprot-GOA¹⁶ was searched in the QuickGO! browser to obtain compartment and function annotations for all proteins¹⁷.

RESULTS AND DISCUSSION

Co-cultures of *S. aciditrophicus* and *M. hungatei* were grown with 20 mM crotonate, 5 mM benzoate, or 5 mM cyclohexane carboxylate. Cells were harvested, and the *S. aciditrophicus* cells were separated from the *M. hungatei* cells by Percoll separation. Protein content was estimated from wet weight of each sample in order to process an equivalent amount of protein from each sample. Pellets were washed and then processed by eFASP.

A portion of the peptides resulting from eFASP were analyzed by shotgun LC-MS^E, and the remainder was fractionated by HILIC in TopTip spin columns prior to duplicate LC-MS^E analyses of each fraction. Fractions were merged with DECO to assure the accuracy of peptide ion volumes in the merged file. By shotgun LC-MS^E, we identify an average of 242 protein identifications, based on 4,800 peptide identifications per injection. HILIC pre-fractionation allows for the identification of approximately 513 protein identifications, based on 15,000 peptide identifications per sample.

Protein quantification was based on the average ion volume of the top three ionizing peptides for each protein. Each HILIC-LC-MS^E merged sample was processed by DECO, which calculated the Hi3 for each protein for export. 149 proteins were quantified across all sample replicate analyses, and 995 unique proteins were quantified

across the sample conditions (Figure 5-1 and Table 5-1). Comparative analysis was based on the average Hi3 in each technical replicate for a sample condition. There is a lot of redundancy in the proteome database, which makes the determination of functional conclusions difficult when only lists of up or down regulated proteins are analyzed. Out of the 995 unique protein accession numbers identified, there are 92 protein descriptions that are associated with multiple accession numbers (Table 5-2). Therefore, the analysis of the quantitative data needs to be performed based on protein descriptions, rather than by accession number, in such a way as to consider quantification data for all protein accession numbers assigned to a particular protein description. This is being performed by the McInerney lab (Department of Botany & Microbiology, University of Oklahoma).

Gene ontology analysis of proteins uniquely identified in cells incubated with crotonate (151 unique proteins), with benzoate (131 unique proteins), or with cyclohexane carboxylate (113 unique proteins), was performed with QuickGO. The number of annotations associated with each GO identifier in the data, and the number of proteins associated with each GO identifier was calculated (Table 5-3). These data indicate the major pathways and functional components that are modulated in the three conditions.

Uniprot ID	Protein Description	Average intensity of 2 replicates		
		Crotonate	Benzoate	Cyclohexane
Q2LTG3	10 kDa chaperonin	2,348,685	285,030	2,045,462
Q2LU43	10 kDa chaperonin	1,662,828	528,263	2,918,890
Q2LPJ7	10 kDa chaperonin	1,300,516	340,387	1,916,724
Q2LUD2	3 dehydroquinase synthase	240,027	47,429	129,982
Q2LUN4	3 hydroxyacyl CoA dehydrogenase	1,916,424	809,542	1,364,811
Q2LWJ2	3 isopropylmalate dehydratase large subunit	553,731	37,099	348,494
Q2LWI5	3 isopropylmalate dehydrogenase	471,831	184,083	353,559
Q2LTJ5	3 methyl 2 oxobutanoate hydroxymethyltransferase Syntrophus aciditrophicus	156,306	57,561	154,082
Q2LXS6	3 oxoacyl reductase	2,776,120	1,208,043	3,057,200
Q2LXQ4	3 oxoacyl reductase	215,936	89,235	110,062
Q2LQA2	30S ribosomal protein S10	890,003	212,396	270,262
Q2LVU9	30S ribosomal protein S16	528,342	122,558	138,939
Q2LUJ7	30S ribosomal protein S18	509,469	152	146,747
Q2LTQ7	30S ribosomal protein S2	622,596	98,147	386,164
Q2LQK7	30S ribosomal protein S20	281,681	37,756	72,103
Q2LQ95	30S ribosomal protein S3	561,090	102,080	694,810
Q2LQD0	30S ribosomal protein S4	509,270	38,681	259,113
Q2LQA6	30S ribosomal protein S8	469,312	78,103	216,214
Q2LRH0	4 hydroxybenzoate CoA ligase benzoate CoA Syntrophus aciditrophicus	8,861,637	5,780,930	7,902,778
Q2LRH7	4 hydroxybenzoate CoA ligase benzoate CoA Syntrophus aciditrophicus	2,300,209	713,395	1,550,225
Q2FTS9	5 10 methylenetetrahydromethanopterin reductase Methanospirillum hungatei	434,538	359,013	3,639,273
Q2LWJ5	5 methyltetrahydrofolate homocysteine methyltransferase Syntrophus aciditrophicus	217,334	63,820	92,529
Q2LQ90	50S ribosomal protein L1	636,070	124,759	140,724
Q2LQ89	50S ribosomal protein L10	386,432	81,651	956,675
Q2LQ91	50S ribosomal protein L11	383,415	81,937	194,934
Q2LPM2	50S ribosomal protein L13	511,066	48,970	99,173
Q2LQ96	50S ribosomal protein L22	309,485	76,886	92,960
Q2LUK6	50S ribosomal protein L25	805,951	141,441	144,924
Q2LQC2	50S ribosomal protein L6	585,001	113,391	255,442
Q2LQ88	50S ribosomal protein L7 L12	2,658,199	322,346	476,577
Q2LUJ5	50S ribosomal protein L9	478,031	70,100	120,545
Q2LQN1	6 7 dimethyl 8 ribityllumazine synthase	369,646	51,394	50,335
Q2LXU3	6 hydroxycyclohex 1 ene 1 carboxyl CoA dehydrogenase Syntrophus aciditrophicus	8,611,235	2,929,017	5,772,117
Q2LPJ8	60 kDa chaperonin 1	2,771,247	1,128,452	1,984,971
Q2LTG7	60 kDa chaperonin 2	5,787,703	1,907,801	4,347,649
Q2LU42	60 kDa chaperonin 3	2,862,586	1,379,937	4,019,843
Q2FPN5	60 kDa chaperonin Methanospirillum hungatei	40,252	24,142	1,566,345
Q2LQ52	ABC type transport system involved in resistance Syntrophus aciditrophicus	67,643	23,029	155,305

Q2LPG9	ABC type tungstate transport system solute bi Syntrophus aciditrophicus	201,900	15,863	394,136
Q2LXP0	Acetolactate synthase small subunit	80,919	35,040	197,613
Q2LXR7	Acetyl CoA acetyltransferase	3,485,057	1,345,636	3,954,269
Q2LWQ5	Acetyl CoA acetyltransferase	3,346,653	1,273,657	1,922,862
Q2LWR8	Acetyl CoA synthetase	3,761,649	1,750,763	8,646,857
Q2LW77	Acetyl coenzyme A synthetase	320,553	194,764	119,394
Q2LY62	Acriflavin resistance periplasmic protein	405,708	127,759	471,248
Q2LXR0	Acyl carrier protein	274,122	89,355	456,017
Q2LQP0	Acyl CoA dehydrogenase medium chain specific	129,578	208,166	5,561,128
Q2LUY8	Acyl CoA dehydrogenase short chain specific	6,214,506	2,150,175	2,875,669
Q2LSQ9	Adenosylhomocysteinase	648,487	122,261	471,690
Q2LTN2	Adenylate kinase	208,932	44,020	802,278
Q2LTE3	Adenylate kinase	71,879	11,680	116,997
Q2LUB5	Aldolase	63,548	58,085	859,285
Q2LUE2	Anthranilate phosphoribosyltransferase	565,928	218,416	396,217
Q2FN15	Archaeal histone Methanospirillum hungatei	265,694	85,412	1,399,469
Q2FLE1	Archaeal histone Methanospirillum hungatei	34,775	24,774	528,246
Q2FMX3	Archaeal histone Methanospirillum hungatei	27,270	27,687	1,137,513
Q2LT97	Argininosuccinate synthase	816,226	125,331	147,880
Q2LR84	Aromatic ring hydroxylase	1,735,557	362,819	534,484
Q2LWK1	Aspartate aminotransferase	399,294	64,280	333,493
Q2LTN0	Aspartate aminotransferase	93,280	27,614	82,306
Q2LPY7	Aspartyl glutamyl tRNA Asn Gln amidotransfera Syntrophus aciditrophicus	74,301	46,641	179,870
Q2LTK5	ATP dependent Clp protease proteolytic subunit Syntrophus aciditrophicus	453,835	76,346	599,803
Q2LT90	ATP dependent endopeptidase hsl ATP binding su Syntrophus aciditrophicus	112,328	120,759	1,243,264
Q2LPQ4	ATP dependent protease La	231,601	94,464	139,667
Q2LR00	ATP synthase B chain sodium ion specific	416,364	172,619	552,935
Q2LR04	ATP synthase epsilon chain	850,902	295,572	239,692
Q2LQZ6	ATP synthase gamma chain	1,003,322	352,349	808,962
Q2LQZ7	ATP synthase subunit alpha 1	2,926,468	715,001	1,955,997
Q2LQZ9	ATP synthase subunit b 1	339,357	62,856	238,695
Q2LR05	ATP synthase subunit beta	3,216,933	1,065,689	2,716,290
Q2LQZ8	ATP synthase subunit delta	971,555	342,326	788,406
Q2LRN6	Bacterial HflC protein	475,550	144,684	244,867
Q2LVV7	CAMP dependent transcriptional regulator	93,799	47,060	52,686
Q2LUP7	Carbonic anhydrase	32,647	8,989	122,032
Q2LUT7	CarD like transcriptional regulator	56,318	28,513	499,047
Q2LUH6	Chaperone protein dnaK	2,281,127	593,781	1,196,397
Q2LPL3	CinA like protein	152,634	73,297	270,662
Q2FKZ1	CoB CoM heterodisulfide reductase subunit A Methanospirillum hungatei	162,836	56,980	574,373
Q2LQ45	Cold shock protein	975,558	68,638	605,399
Q2LSJ2	D 3 phosphoglycerate dehydrogenase	1,542,206	423,633	718,775

Q2LY87	DegT dnrJ eryC1 strS aminotransferase family p Syntrophus aciditrophicus	85,182	136,644	95,422
Q2LUA3	Deoxyhypusine synthase	170,033	60,953	87,415
Q2LTA0	Dihydrodipicolinate reductase	663,508	276,135	900,342
Q2LTA1	Dihydrodipicolinate synthase	1,494,174	131,087	216,178
Q2LXP6	Dihydroxy acid dehydratase	258,347	60,295	190,132
Q2LSX6	DNA binding protein HU	1,377,493	618,284	1,023,839
Q2LW81	DNA binding protein HU	1,103,817	288,537	686,525
Q2LR25	DNA binding protein HU	346,103	83,732	159,152
Q2LQC9	DNA directed RNA polymerase subunit alpha	650,428	70,194	307,640
Q2LYA2	DNA polymerase III beta chain	126,432	52,791	152,331
Q2LSS1	DnaK suppressor protein	64,292	74,518	164,488
Q2LWR0	Electron transfer flavoprotein alpha subunit	6,211,407	3,388,833	7,550,031
Q2LWQ9	Electron transfer flavoprotein beta subunit	3,700,027	1,063,712	6,857,103
Q2FRI3	Elongation factor 1 alpha Methanospirillum hungatei	53,638	17,196	1,519,233
Q2LTQ6	Elongation factor Ts	758,252	151,082	265,710
Q2LQA3	Elongation factor Tu	2,354,981	360,438	2,389,478
Q2LR33	Enolase	637,340	183,217	432,383
Q2LUN3	Enoyl CoA hydratase	932,145	338,044	680,947
Q2LXU2	Enoyl CoA hydratase isomerase family protein	9,072,882	5,723,711	9,201,213
Q2FU89	F420 dependent methylenetetrahydromethanopteri Methanospirillum hungatei	273,277	175,035	1,864,579
Q2LRW1	Flavin nucleotide binding protein structurally Syntrophus aciditrophicus	279,009	166,367	461,389
Q2LVZ2	Flavodoxin oxidoreductase	846,198	119,719	356,485
Q2LS98	Formate dehydrogenase	147,370	115,248	1,494,557
Q2FKY1	Formate dehydrogenase alpha subunit Methanospirillum hungatei	98,106	54,943	744,450
Q2FSV2	Formate dehydrogenase alpha subunit Methanospirillum hungatei	61,084	32,490	409,193
Q2FKY0	Formate dehydrogenase beta subunit Methanospirillum hungatei	223,710	108,641	671,691
Q2LS99	Formate dehydrogenase iron sulfur subunit	201,466	119,545	1,182,712
Q2LVY8	Formate dehydrogenase major subunit	933,094	169,216	479,210
Q2LVY7	Formate dehydrogenase major subunit	292,880	205,075	996,171
Q2LUC0	Fructose 1 6 biphosphatase class 1	212,992	59,078	434,431
Q2LU85	Gamma glutamyl phosphate reductase	184,271	106,913	201,715
Q2LVW5	GDP mannose 4 6 dehydratase	283,168	75,318	197,290
Q2LUY7	Glutaconyl CoA decarboxylase A subunit	5,423,660	1,340,860	2,120,902
Q2LQ26	Glutaconyl CoA decarboxylase beta subunit	487,347	195,868	205,695
Q2LQG2	Glutamate dehydrogenase	938,578	325,180	424,019
Q2LW88	Glutamate dehydrogenase	907,736	258,540	301,469
Q2LVY4	Glutamate synthase Syntrophus aciditrophicus	1,155,540	281,370	442,531
Q2LWW2	Glyceraldehyde 3 phosphate dehydrogenase	1,726,694	644,981	920,790
Q2LVI0	Glycyl tRNA synthetase alpha subunit	989,320	306,180	1,878,559
Q2LQN0	GTP cyclohydrolase II 3 4 dihydroxy 2 butano Syntrophus aciditrophicus	435,123	78,591	92,635

Q2LUP4	Hydantoinase oxoprolinase family	222,472	60,353	97,228
Q2LXR8	Hypothetical cytosolic protein	3,254,385	861,913	3,486,833
Q2LS73	Hypothetical cytosolic protein	614,351	159,139	458,736
Q2LS49	Hypothetical cytosolic protein	533,567	123,381	95,735
Q2LY43	Hypothetical cytosolic protein	483,677	101,115	359,316
Q2LY42	Hypothetical cytosolic protein	461,174	51,271	384,645
Q2LVQ7	Hypothetical cytosolic protein	279,136	57,143	82,522
Q2LSC6	Hypothetical cytosolic protein	236,770	75,791	153,895
Q2LPN6	Hypothetical cytosolic protein	210,160	106,281	304,693
Q2LW31	Hypothetical cytosolic protein	205,602	79,860	71,810
Q2LWA1	Hypothetical cytosolic protein	99,555	17,286	68,442
Q2LXL6	Hypothetical cytosolic protein	96,652	19,378	484,309
Q2LPN1	Hypothetical cytosolic protein	66,459	23,724	1,655,137
Q2LWJ7	Hypothetical cytosolic protein	53,128	25,598	382,426
Q2LU51	Hypothetical cytosolic protein	43,762	20,745	134,920
Q2LUL0	Hypothetical cytosolic protein	43,103	6,457	110,346
Q2LU87	Hypothetical cytosolic protein	39,940	12,617	118,529
Q2LWR5	Hypothetical exported protein	10,060,902	3,181,209	8,651,967
Q2LTM6	Hypothetical exported protein	789,016	390,436	284,936
Q2LW10	Hypothetical exported protein	456,446	147,680	269,664
Q2LW16	Hypothetical exported protein	265,007	13,792	127,881
Q2LRF0	Hypothetical exported protein	211,002	70,037	182,775
Q2LWW5	Hypothetical exported protein	9,794	7,435	72,216
Q2LUP1	Hypothetical membrane protein	485,355	340,269	474,880
Q2LWD5	Indolepyruvate oxidoreductase subunit	182,691	96,601	235,997
Q2LTT4	Inosine 5 monophosphate dehydrogenase	395,993	107,780	225,184
Q2LSB7	Iron only hydrogenase large subunit	635,534	225,860	1,317,342
Q2LWE4	Iron sulfur cluster and nucleotide binding pro Syntrophus aciditrophicus	345,751	21,885	385,779
Q2LWF2	Iron sulfur cluster assembly repair protein	261,459	46,427	153,059
Q2LWR1	Iron sulphur binding reductase	1,289,906	637,538	1,600,612
Q2LX27	Isochorismatase	85,879	26,690	58,937
Q2LXP7	Ketol acid reductoisomerase 2 dehydropantoat Syntrophus aciditrophicus	251,911	62,359	821,571
Q2LUI9	Large conductance mechanosensitive channel	20,056	53,430	205,497
Q2LQK5	Leucyl tRNA synthetase	314,733	97,987	349,622
Q2LR69	Long chain fatty acid CoA ligase	131,356	168,313	3,996,816
Q2FQ36	LSU ribosomal protein L12AE Methanospirillum hungatei	22,438	23,884	456,641
Q2LTM0	Membrane protease subunit stomatin prohibitin Syntrophus aciditrophicus	303,204	66,658	59,312
Q2LYC2	Metallo beta lactamase protein	252,489	60,914	65,673
Q2FSN0	Methyl coenzyme M reductase alpha subunit Methanospirillum hungatei	447,037	377,469	5,393,304
Q2FSD8	Methyl coenzyme M reductase beta subunit Methanospirillum hungatei	335,726	337,645	3,726,675
Q2FSN1	Methyl coenzyme M reductase gamma subunit Methanospirillum hungatei	436,619	257,438	4,104,035

Q2LPJ6	Molecular chaperone	717,792	194,626	515,807
Q2LPJ5	Molecular chaperone	524,409	190,901	384,345
Q2LYH4	Molecular chaperone	374,793	158,245	354,995
Q2LQF2	MoxR like ATPase with AAA domain	260,934	58,789	476,552
Q2LWH9	Multimeric flavodoxin	101	15,900	37,269
Q2LVL7	NAD dependent oxidoreductase	88,797	42,990	148,929
Q2LSB6	NADH quinone oxidoreductase chain F	768,929	190,974	692,621
Q2LS97	NADH quinone oxidoreductase chain F	136,723	148,479	2,607,593
Q2LYE8	NADH ubiquinone oxidoreductase 24 kD subunit N Syntrophus aciditrophicus	372,386	65,511	297,122
Q2LUB3	NifU like domain protein	43,811	30,156	168,609
Q2LWJ4	Orotate phosphoribosyltransferase	23,394	16,472	157,688
Q2LQF8	Outer membrane OmpA family protein	1,022,314	508,122	860,218
Q2LS67	Outer membrane OmpA family protein	220,010	69,779	82,629
Q2LRQ5	Outer membrane porin protein	2,065,423	1,201,874	986,226
Q2LRP6	Peptidoglycan associated lipopeptide	809,026	307,926	524,522
Q2LR26	Phenylalanyl tRNA synthetase beta chain	303,070	73,291	304,697
Q2LRB5	Phosphoenolpyruvate synthase	1,687,816	348,650	595,239
Q2LWU0	Phosphoesterase DHH family protein	48,285	15,154	28,227
Q2LR90	Phosphoribosylaminoimidazolecarboxamide formyl Syntrophus aciditrophicus	421,920	145,759	782,582
Q2LSJ6	Phosphoserine aminotransferase	361,586	44,312	367,409
Q2LWT4	Polyribonucleotide nucleotidyltransferase	433,401	125,617	786,815
Q2LUH5	Protein grpE	432,729	135,961	182,304
Q2LTP2	Protein translation elongation factor G	306,699	78,128	949,424
Q2LUL1	Proton translocating pyrophosphatase	2,313,055	840,176	1,126,096
Q2LUK4	Proton translocating pyrophosphatase	2,292,783	655,505	1,562,058
Q2LXU6	Putative enoyl CoA hydratase	6,767,392	2,767,725	6,976,392
Q2FN87	Putative K stimulated pyrophosphate energiz Methanospirillum hungatei	514,800	89,184	198,682
Q2LTH8	Putative peptidoglycan binding protein	144,537	115,870	357,245
Q2FS26	Putative uncharacterized protein Methanospirillum hungatei	348,976	299,216	3,602,534
Q2FTS3	Putative uncharacterized protein Methanospirillum hungatei	341,205	290,386	3,217,502
Q2FUG9	Putative uncharacterized protein Methanospirillum hungatei	170,573	131,095	1,927,570
Q2FLG2	Putative uncharacterized protein Methanospirillum hungatei	32,326	19,293	615,488
Q2LYH6	Pyruvate dehydrogenase Syntrophus aciditrophicus	525,090	192,103	408,685
Q2LYE9	Pyruvate ferredoxin oxidoreductase beta subunit Syntrophus aciditrophicus	510,616	102,914	181,500
Q2LUK7	Ribose phosphate pyrophosphokinase	342,314	82,878	242,121
Q2LSN2	Ribosome associated factor Y	661,948	268,281	319,808
Q2LVE8	RNA polymerase sigma factor	298,321	60,658	342,519
Q2LQZ5	Rod shape determining protein	219,049	32,374	57,156
Q2LW56	Rubrerhythrin	102,088	26,759	141,089
Q2LUY1	Rubrerhythrin	88,626	50,029	361,655

Q2LQN2	Rubrerethrin and thioesterase superfamily prot Syntrophus aciditrophicus	196,605	96,504	274,935
Q2LSR0	S adenosylmethionine synthetase	612,663	244,805	264,710
Q2LVL9	Single stranded DNA binding protein	112,097	95,360	83,592
Q2LU35	Small heat shock protein	345,665	134,460	422,154
Q2LXU0	Sodium translocating NADH quinone reductase su Syntrophus aciditrophicus	242,041	143,427	97,553
Q2LSR9	SSU ribosomal protein S1P	645,968	259,859	434,671
Q2LUJ8	SSU ribosomal protein S6P	482,033	106,903	147,527
Q2LPK8	Succinyl CoA synthetase alpha subunit	249,551	18,302	100,105
Q2LPK9	Succinyl CoA synthetase beta chain	490,710	76,494	204,370
Q2FRN0	Tetrahydromethanopterin S methyltransferase s Methanospirillum hungatei	166,237	55,502	721,565
Q2FRN1	Tetrahydromethanopterin S methyltransferase s Methanospirillum hungatei	83,439	68,459	809,332
Q2FRN3	Tetrahydromethanopterin S methyltransferase s Methanospirillum hungatei	15,526	16,987	271,488
Q2FRN2	Tetrahydromethanopterin S methyltransferase su Methanospirillum hungatei	89,719	48,573	678,561
Q2LVP3	Tfp pilus assembly protein major pilin	1,136,269	486,910	214,532
Q2FTL9	Thermosome subunit Methanospirillum hungatei	61,617	13,456	687,834
Q2LVG3	Thioredoxin	185,800	37,503	1,030,299
Q2LR22	Threonyl tRNA synthetase	262,354	56,305	146,320
Q2LQ92	Transcription antitermination protein nusG	145,858	25,152	283,699
Q2LWU5	Transcription elongation factor	286,142	84,689	266,531
Q2LTP9	Transcription elongation factor	46,475	8,385	54,557
Q2LPL0	Transcriptional regulator GntR family	544,103	85,631	278,217
Q2LX56	Transporter	68,330	30,322	42,322
Q2LTK6	Trigger factor	584,795	124,344	259,942
Q2LUE0	Tryptophan synthase alpha chain	134,085	28,111	128,797
Q2LUH1	Tryptophan synthase beta chain	433,392	148,975	154,526
Q2LY69	Type I secretion outer membrane protein	307,452	52,066	222,832
Q2LRX9	Tyrosine phenol lyase	937,424	109,513	493,012
Q2LR13	UDP glucose 6 dehydrogenase	117,757	50,475	297,334
Q2LWH5	Uncharacterized solute binding protein SYNAS 2 Syntrophus aciditrophicus	163,851	50,466	486,328
Q2LVI7	Universal stress protein family	79,855	46,552	154,681
Q2LRF2	Xaa pro aminopeptidase	345,615	164,986	315,138

Table 5-1. Quantification of 233 proteins across all the sample conditions. Intensity values represent the average ion intensity of the top three ionizing peptides.

Protein description	Count
Hypothetical cytosolic protein	79
Putative uncharacterized protein <i>Methanospirillum hungatei</i>	42
Hypothetical membrane protein	23
Hypothetical exported protein	21
Tetrahydromethanopterin S methyltransferase s <i>Methanospirillum hungatei</i>	6
DNA binding protein HU	4
Archaeal histone <i>Methanospirillum hungatei</i>	4
Protein translocase subunit	4
Formate dehydrogenase beta subunit <i>Methanospirillum hungatei</i>	4
Aspartyl glutamyl tRNA Asn Gln amidotransferase	4

Table 5-2. Protein description redundancy challenges analysis. Uniprot protein descriptions associated with multiple identified protein accession numbers were selected for frequency analysis. The number of unique protein accession numbers identified in these data sets, which had identical protein descriptions, was counted. Displayed are the protein descriptions with the largest number of associated accession numbers, along with the count of unique accessions for that protein description.

Crotonate

Count of Annotations per GO Identifier		
GO Identifier Description	%	Count
cytoplasm	4.2	41
ATP binding	3.8	37
transferase activity	3.4	33
nucleotide binding	3.1	30
metabolic process	2.9	28
catalytic activity	2.9	28
oxidation-reduction process	2.2	22
regulation of transcription, DNA-dependent	2.1	21
GTP binding	1.8	18
metal ion binding	1.6	16

Count of Proteins per GO Identifier		
GO Identifier Description	%	Count
transferase activity	30.4	31
nucleotide binding	22.6	23
catalytic activity	19.6	20
metabolic process	17.7	18
ATP binding	15.7	16
metal ion binding	14.7	15
cytoplasm	13.7	14
hydrolase activity	12.8	13
membrane	10.8	11
intracellular	10.8	11

Benzoate

Count of Annotations per GO Identifier		
GO Identifier Description	%	Count
ATP binding	4.9	29
oxidation-reduction process	4.1	24
catalytic activity	3.4	20
oxidoreductase activity	3.2	19
metabolic process	2.7	16
iron-sulfur cluster binding	2.6	15
cytoplasm	2.6	15
nucleotide binding	2.6	15
transferase activity	2.2	13
membrane	2.2	13

Count of Proteins per GO Identifier		
GO Identifier Description	%	Count
ATP binding	18.2	14
transferase activity	16.9	13
oxidation-reduction process	15.6	12
metabolic process	15.6	12
catalytic activity	15.6	12
oxidoreductase activity	14.3	11
nucleotide binding	14.3	11
hydrolase activity	13.0	10
membrane	11.7	9
iron-sulfur cluster binding	10.4	8

Cyclohexane

Count of Annotations per GO Identifier		
GO Identifier Description	%	Count
oxidation-reduction process	6.1	43
ribosome	4.7	33
translation	3.9	27
intracellular	3.7	26
oxidoreductase activity	3.6	25
structural constituent of ribosome	3.6	25
ATP binding	2.9	20
cytoplasm	2.7	19
regulation of transcription, DNA-dependent	2.1	15
transferase activity	1.9	13

Count of Proteins per GO Identifier		
GO Identifier Description	%	Count
oxidation-reduction process	20.0	16
oxidoreductase activity	18.8	15
intracellular	17.5	14
transferase activity	16.3	13
translation	13.8	11
ribonucleoprotein complex	12.5	10
ribosome	12.5	10
structural constituent of ribosome	12.5	10
ATP binding	11.3	9
nucleotide binding	11.3	9

Table 5-3. Analysis of unique proteins in each condition. QuickGO associations between GO identifiers and proteins in the data set are determined. The number of annotations associated with each Go identifier in the data was calculated, and the top 10 are listed to the left. The number of proteins associated with each Go identifier was calculated, and the top 10 are listed to the right.

REFERENCES

1. Jackson, B. E., Bhupathiraju, V. K., Tanner, R. S., Woese, C. R. & McInerney, M. J. *Syntrophus aciditrophicus* sp. nov., a new anaerobic bacterium that degrades fatty acids and benzoate in syntrophic association with hydrogen-using microorganisms. *Arch Microbiol* **171**, 107–114 (1999).
2. McInerney, M. J. *et al.* The genome of *Syntrophus aciditrophicus*: life at the thermodynamic limit of microbial growth. *Proc Natl Acad Sci U S A* **104**, 7600–7605 (2007).
3. Mouttaki, H., Nanny, M. A. & McInerney, M. J. Metabolism of hydroxylated and fluorinated benzoates by *Syntrophus aciditrophicus* and detection of a fluorodiene metabolite. *Appl Environ Microbiol* **75**, 998–1004 (2009).
4. McInerney, M. J. *et al.* Physiology, ecology, phylogeny, and genomics of microorganisms capable of syntrophic metabolism. *Ann N Y Acad Sci* **1125**, 58–72 (2008).
5. McInerney, M. J., Sieber, J. R. & Gunsalus, R. P. Syntrophy in anaerobic global carbon cycles. *Curr Opin Biotechnol* **20**, 623–632 (2009).
6. Mouttaki, H., Nanny, M. A. & McInerney, M. J. Cyclohexane carboxylate and benzoate formation from crotonate in *Syntrophus aciditrophicus*. *Appl Environ Microbiol* **73**, 930–938 (2007).
7. Elshahed, M. S., Bhupathiraju, V. K., Wofford, N. Q., Nanny, M. A. & McInerney, M. J. Metabolism of benzoate, cyclohex-1-ene carboxylate, and cyclohexane carboxylate by ‘*Syntrophus aciditrophicus*’ strain SB in syntrophic association with H₂-using microorganisms. *Appl Environ Microbiol* **67**, 1728–1738 (2001).
8. Mouttaki, H., Nanny, M. A. & McInerney, M. J. Use of benzoate as an electron acceptor by *Syntrophus aciditrophicus* grown in pure culture with crotonate. *Environ Microbiol* **10**, 3265–3274 (2008).
9. Masuda, T., Tomita, M. & Ishihama, Y. Phase transfer surfactant-aided trypsin digestion for membrane proteome analysis. *J Proteome Res* **7**, 731–740 (2008).
10. Yeung, Y. G., Nieves, E., Angeletti, R. H. & Stanley, E. R. Removal of detergents from protein digests for mass spectrometry analysis. *Anal Biochem* **382**, 135–137 (2008).
11. Li, G. Z. *et al.* Database searching and accounting of multiplexed precursor and product ion spectra from the data independent analysis of simple and complex peptide mixtures. *Proteomics* **9**, 1696–1719 (2009).

12. Silva, J. C., Gorenstein, M. V., Li, G. Z., Vissers, J. P. & Geromanos, S. J. Absolute quantification of proteins by LCMS^E: a virtue of parallel MS acquisition. *Mol Cell Proteomics* **5**, 144–156 (2006).
13. Silva, J. C. *et al.* Simultaneous qualitative and quantitative analysis of the *Escherichia coli* proteome: a sweet tale. *Mol Cell Proteomics* **5**, 589–607 (2006).
14. Silva, J. C. *et al.* Quantitative proteomic analysis by accurate mass retention time pairs. *Anal Chem* **77**, 2187–2200 (2005).
15. Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J Mol Biol* **157**, 105–132 (1982).
16. Dimmer, E. C. *et al.* The UniProt-GO Annotation database in 2011. *Nucleic Acids Res* **40**, D565–D570 (2012).
17. Binns, D. *et al.* QuickGO: a web-based tool for Gene Ontology searching. *Bioinformatics* **25**, 3045–3046 (2009).

CHAPTER 6

Differential exosome packaging by melanoma cells treated with Plexxikon (PLX-4720): probing a potentially rich source of protein and nucleic acid biomarkers for diagnostics and therapeutic response monitoring.

ABSTRACT

Tumor-derived exosomes are quickly becoming a major field of research in cancer biology and diagnostics. These extracellular organelles can transmit signals, proteins, microRNAs, and mRNAs, to target cells. Exosomes found in circulation reflect their cell of origin, and its status, which makes them a prime candidate for biomarker discovery.

A large percentage of patients with melanoma have the V600E mutation in B-raf kinase. Inhibition of B-Raf(V600E) with vemurafenib demonstrates remarkable response rates in patients; however, drug resistance is frequent. Vemurafenib was found to stimulate the MAPK pathway in some normal cells, which can have widespread effects. We hypothesize that melanoma cell lines treated with a vemurafenib analogue, PLX-4720, will display measureable RNA and protein profile changes in the cells and in the secreted exosomes.

We employed a label-free quantitative proteomic approach to produce a comprehensive collection of exosomal proteins derived from M249 melanoma cells that were treated with vehicle or PLX-4720. Using eFASP and LC-MS^E, 476 proteins were quantified in the conditions. PLX-4720 treatment induced the upregulation of 82

proteins, and the downregulation of 40 proteins, in the collected exosomes. The most compelling finding is the appearance of two tyrosine kinases, VEGFR and PDGFR- β , in exosomes following drug treatment. Upregulation of these receptor tyrosine kinases has been found to be a result of the release of feedback inhibition following treatment with MAPK targeted therapeutics.

PLX-4720 treatment of melanoma cells alters the protein composition of the derived exosomes *in vitro*. Proteins affected may be functionally relevant or early indicators of resistance response. These data suggest that exosomal proteins can be used to monitor cellular responses to B-Raf inhibition.

TABLE OF CONTENTS

INTRODUCTION.....	158
EXPERIMENTAL.....	161
Strains and cell growth	161
Enhanced filter aided sample preparation of exosome proteins	161
Peptide fractionation by HILIC	163
Analysis by nanoLC-MS ^E	164
Data processing and protein identification	165
RESULTS AND DISCUSSION	166
PLX-induced down regulation	167
PLX-induced up regulation	168
Melanoma exosome data for ExoCarta	169
REFERENCES.....	176

INTRODUCTION

In addition to intracellular organelles, eukaryotic cells also have extracellular organelles that are released or shed into the microenvironment. These include exosomes, shedding microvesicles, and apoptotic blebs. Exosomes in contrast have well defined cup-like shape with sizes ranging from 50-100nm, and population homogeneity. All exosomes maintain a common set of membrane and cytosolic proteins that are involved in structure and trafficking, as well as distinct lipid composition ¹. Exosomes from different cell types can be differentiated based on distinct sets of proteins that presumably reflect specific functions or the status the cell of origin. In addition to protein material, exosomes carry mRNAs and miRNAs ². Exosomes have been shown to have distinct roles in intercellular communication.

Most intercellular communication is presumed to be a result of cell surface molecules and soluble ligands that activate receptors. Studies with exosomes, reveal potentially new mechanisms of communication ³. The first mechanism is based on membrane interactions, whereby components of the exosome membrane interact with those on a recipient cell. Second, the cleavage of ectodomain proteins produces ligands appropriate for activating a cell surface receptor on a target cell. Third, fusion or receptor mediated endocytosis, which results in the nonselective transfer of proteins, nucleic acids, and lipids to the recipient cell. This is functionally relevant since receptors may be added and new proteins may be expressed or suppressed in target cells by mRNA and miRNA transmission. An interesting example is the transfer of oncogenic EGFR to a normal recipient cell via tumor-derived exosomes.

The physiological roles of exosomes have been largely studied with respect to immune function and cancer biology. Exosomes have three major functions in immunomodulation. First, in cases of tumors and intracellular pathogens, exosomes may transmit antigens from infected cells, inducing response amplification by transfer of antigenic materials to DCs that have not previously encountered them ⁴. Second, exosomes maintain peptide MHCs on the surface which can directly activate T cells, making them act as APCs. Finally, they have been shown to present various ligands that are capable of inducing T cell apoptosis, enhance regulatory T (T_{reg}) cell function, or induce cytokine secretion ⁴.

Exosomes carry a lot of biologically active mRNA and microRNAs. This shuttling of RNA from one cell to another followed by translation of that RNA has been demonstrated in mast cells. Viruses like HIV, as well as pathogenic Prions, can traffic to MVBs and release from the infected cell in exosomes, and potentially move on to neighboring cells ⁵⁻⁷.

With respect to tumor biology, secreted exosomes can be involved in metastasis and immune escape ⁸. They can reprogram recipient cells through the transfer of mutant proteins, or miRNAs that can reprogram recipient cells. High circulating levels of exosomes are found in the later stages of cancer although the reason for this is not known, but it may represent communication between tumor sites or establishment of new sites ³.

While exosomes have been shown to be a source of tumor antigens and of tumor antigen presentation to T cells to promote an anti-tumor response, they maintain abilities to suppress immune responses. Exosomes from melanoma and ovarian

cancer cell lines can suppress the anti-tumor immune response with a ligand (FasL) that induces T cell death ⁴. While the exact role of exosomes in cancer biology is not clear, they may be involved in communication between tumor sites, or establishment of new tumor sites.

Exosomes can be found in, urine, milk, amniotic fluid, plasma, bronchiolar lavage, and synovial fluid ^{2,9-12}. Coupled with the large overlap of exosomal proteins with the, exosomes present a potentially rich source of biomarkers that may be easily sampled to provide information on the cell of origin, its functional status, or disease progression. One can envision multiple uses for exosomes in diagnostics, monitoring of therapeutic response and pharmacodynamics, based on protein and nucleic acid content.

Tumor-derived exosomes are quickly becoming a major field of research in cancer biology and diagnostics. These extracellular organelles can transmit signals, proteins, microRNAs, and mRNAs, to target cells. Exosomes found in circulation reflect their cell of origin, and its status, which makes them a prime candidate for biomarker discovery.

A large percentage of patients with melanoma have the V600E mutation in B-raf kinase. Inhibition of B-Raf(V600E) with vemurafenib demonstrates remarkable response rates in patients; however, drug resistance is frequent ¹³⁻¹⁵. Vemurafenib was found to stimulate the MAPK pathway in some normal cells, which can have widespread effects ¹³. We hypothesize that melanoma cell lines treated with a vemurafenib analogue, PLX-4720, will display measureable RNA and protein profile changes in the cells and in the secreted exosomes.

The aim of this work is to define the nucleic acid and protein components of *in vitro* tumor-derived exosomes, and to quantify the changes in composition following drug treatment. This will provide a set of markers that may be validated in *in vivo* models with the same tumors to assess whether *in vitro* markers are predictive of *in vivo* response. A label-free quantitative proteomic approach is utilized to produce a comprehensive collection of exosomal proteins derived from M249 melanoma cells that were treated with vehicle or PLX-4720.

EXPERIMENTAL

Strains and cell growth

M249 melanoma cells were cultured with DMSO or PLX-4720, in duplicate. After 72 hours, exosomes were isolated from flasks by ultracentrifugation, and prepared by eFASP (Figure 6-1).

Enhanced filter aided sample preparation of exosome proteins

Collected exosomes were suspended in Solubilization Buffer A (4% ammonium dodecyl sulfate, 50 mM TCEP, 0.1% deoxycholic acid, 0.1% n-octylglucoside, 100 mM Tris-HCL, pH 8), and incubated on a heated block for 10 minutes at 90°C. The sample was sonicated for three 10-second intervals with a Sonic Dismembrator Model 100 (Fisher Scientific), and then spun at 16,000 rcf for 10 minutes in an Eppendorf centrifuge; this step was repeated once. Glycogen phosphorylase B (rabbit) was added to each sample at a 1:50 ratio. The supernatant and any residual pellet was sonicated and allowed to cool to 37°C.

Eight volumes of Solubilization Buffer B (8 M urea, 0.1% deoxycholic acid, 0.1% n-octylglucoside, 100 mM Tris-HCL, pH 8) was added to the sample to begin buffer exchange of SDS. The diluted sample was applied to a Passivated Microcon Filter Unit (described below), and spun at 14,000 rcf for 10 minutes. The eluate was discarded, and 200 µl of Solubilization Buffer B was added to the filter unit and then spun at 14,000 rcf for 10 minutes. This step was repeated twice. The Passivated Microcon Filter Unit (30,000 Da cutoff), and collection tube, were pre-incubated in 5% Tween-20 overnight, and then thoroughly rinsed with large volumes of MS-grade water.

Proteins were alkylated by the addition of Alkylation Buffer (50 mM iodoacetamide, 8M urea) to the Microcon filter unit, and incubated for 1 hour on a shaker at 37°C. DTT was added to a final concentration of 50 mM to deactivate residual IAN. To remove IAN, DTT and urea, three buffer exchange steps were performed with Digestion Buffer (0.2% n-octylglucoside in 50 mM ABC).

Digestion Buffer (100 µl) was added to the Microcon unit, followed by the addition of 2 µg of trypsin (1 µg trypsin:50 µg protein). Digestion proceeded for 12 hours on a shaker set to 37°C. The Microcon unit was transferred to a new Passivated Collection Tube, and spun at 14,000 rcf for 10 minutes. To assure complete removal of peptides, 100 µl of 50 mM ABC buffer was added to the Microcon unit, and collected by centrifugation, two times. Removal of n-octylglucoside from the digest was accomplished by phase transfer to an ethyl acetate layer^{16,17}, with multiple rounds of EA extraction (addition, phase mixing, and EA removal).

Collected peptides were dried down in a SpeedVac and re-suspended in 50% methanol, three times. Dried peptides were suspended in modified Mobile Phase A

(96.5% H₂O/3% ACN/0.5% formic acid). The peptide solutions were either used for shotgun LC-MS^E analysis, or fractionated by HILIC before LC-MS^E analysis.

Peptide fractionation by HILIC

Binding Buffer: 90% ACN, 15 mM ammonium formate, pH 3.

Release1 Buffer: 76% ACN, 15 mM ammonium formate, pH 3.

Release2 Buffer: 72% ACN, 15 mM ammonium formate, pH 3.

Release3 Buffer: 65% ACN, 15 mM ammonium formate, pH 3.

Release4 Buffer: 60% ACN, 15 mM ammonium formate, pH 3.

Release5 Buffer: 55% ACN, 15 mM ammonium formate, pH 3.

Release6 Buffer: 50% ACN, 15 mM ammonium formate, pH 3.

Condition Buffer: 10% ACN, 15 mM ammonium formate, pH 3.

TopTips (TT200HIL, PolyLC Inc., Columbia, MD) were conditioned by wetting the material with Condition Buffer, followed by incubation with Binding Buffer. Each sample (70 µg) was dried down to remove any acids or salts, and re-suspended in a small volume of 150 mM ammonium formate, pH 3. This solution was added to the TopTip, along with enough ACN to bring it to 90% organic. The peptide solution was passed through the TopTip by centrifugation at 500 rcf, and the eluate was passed through the TopTip again. All subsequent steps were done twice. The TopTip was washed with Binding Buffer. Peptides were eluted, into separate fractions, with Release1 Buffer through Release5 Buffer. All fractions were dried down in a SpeedVac, twice. Peptides were re-suspended in 9 modified Mobile Phase A (96.5% H₂O/3% ACN/0.5% formic acid), for LC-MS^E analysis.

Analysis by nanoLC-MS^E

nanoUPLC of sample peptides was performed on an Waters nanoACQUITY UPLC system equipped with a nanoACQUITY UPLC Symmetry C18 180 μm x 20 mm trap column (Waters, 186003514), and a nanoACQUITY UPLC BEH C18 75 μm x 250 mm reversed phase analytical column (Waters, 186003545). The aqueous mobile phase (mobile phase A) contained 0.1% formic acid in water and the organic mobile phase (mobile phase B) contained 0.1% formic acid in acetonitrile. 5 μl of the sample was loaded onto the trap with and washed with 99% A and 1% B for 5 minutes, and then injected onto the column with 3% mobile phase B. Peptides were eluted from the column with a gradient of 3-40% B over 90 minutes at 300 nl/min, followed by a 5 minute rinse with 95% B. The column was re-equilibrated to 97% A for 20 minutes.

The lock mass compound, [Glu¹]fibrinopeptide (0.5 pmol/ μl , 25% ACN, 0.1% formic acid), and the analyte were delivered at 300 nl/min to the NanoLockSpray dual electrospray ion source (Waters, UK) which was equipped with a microcapillary for the lockmass compound, and a PicoTip Emitter (New Objective) for the analyte. Two electrosprays are produced which may be selectively sampled by positioning of the baffle.

The ion source was interfaced with a Waters Synapt HDMS for mass spectrometric analysis. Acquisition was performed in positive nanoESI mode and the instrument was operated in V-mode with an average resolution of 9500. Accurate mass data for precursor and fragment ions were collected over the 100-2000 m/z range, with the mode of acquisition alternating between low and elevated energy at 1-second

intervals, for the duration of analysis. The lock mass channel was sampled every 30 seconds for calibration of MS and MS^E data.

Data processing and protein identification

Raw data files were processed in ProteinLynx Global Server 2.5 (PLGS, Waters) to provide an inventory of precursor ions along with their respective fragment ions, to search against the Uniprot Swissprot Human database (April 2011) for peptide and protein identifications ¹⁸. Database search settings included two missed cleavages, a 4% false positive rate, fixed carbamidomethyl modification of cysteine residues, and variable oxidation modification of methionine residues. The Waters Expression Informatics component of the PLGS software was utilized to quantify data using glycogen phosphorylase B (Uniprot accession P00489) as the quantification reference ¹⁹⁻²¹. Data Extraction and Correction (DECO) software was used to calculate isoelectric points and grand average of hydropathicity index (GRAVY) values for peptides and proteins identified. GRAVY values were calculated based on the hydropathicity of constituent amino acids derived from Kyte and Doolittle ²². Uniprot-GOA ²³ was searched in the QuickGO! browser to obtain compartment and function annotations for all proteins ²⁴. DECO was also used to collate protein abundance and identification information from all exosome experiments performed in the lab for submission into ExoCarta.

RESULTS AND DISCUSSION

Biological duplicates of M249 melanoma cells are grown and treated with control or a B-Raf kinase inhibitor, PLX-4720^{13,14}. Exosomes are isolated from the cell media by ultracentrifugation, and prepared for RNA analysis or protein analysis. The RNA work began with microarrays and RT PCR, and later deep sequencing, in order to identify and quantify RNA components of the exosomes.

The exosomes for protein analysis and prepared in an SDS buffer, and processed by eFASP. Resulting peptides are then separated into 5 fractions by HILIC in TopTip spin columns prior to five replicate LC-MS^E analyses of each fraction.

Quantitative comparisons are made in Expression Analysis. We are looking for proteins unique to treated or control exosomes, as well as those changing in abundance. In order to gauge the quality of quantification, DECO was utilized to extract the top 3 ionizing peptides for each protein, average them, and then these averages were plotted for a protein found in multiple samples (Figure 6-2).

Proteins that were identified and quantified in at least 2 technical replicates of a sample were used for comparative analysis. Of the 476 unique proteins in the analysis, 82 proteins were upregulated, and 40 proteins were downregulated, following PLX-treatment (Table 6-1 and Table 6-2). Proteins unique to either condition must have been identified in at least 2 technical replicates. There were 83 proteins unique to the exosomes from PLX-treated cells and 52 proteins unique to the control DMSO-treated exosomes.

PLX-induced down regulation

A number of proteins were found downregulated in exosomes following PLX treatment (Table 6-2). Thrombospondin 2 (TSP-2; In ratio (PLX/DMSO) = -1.6 +/- 0.1) mediates cell-to-cell and cell-to-matrix interactions. It was found downregulated in PLX -treatment (In ratio (PLX/DMSO) = -1.6 +/- 0.1). TSP-1 has been found to induce the generation of T_{reg} cells²⁵. T_{reg} cells represent an important immune escape mechanism in advanced melanoma. Essential factors required for T_{reg} cell induction by melanoma remain unidentified; however, thrombospondin-1 (TSP-1) has been found to generate T_{reg} cells through its receptor, CD47²⁵. Therefore, TSP-2 may mediate attachment or spreading of melanoma cells, it may be actively suppressing immune response via induction of T_{reg} cells.

A disintegrin and metalloproteinase domain-containing protein 10 (ADAM 10); In ratio (PLX/DMSO) = -0.9 +/- 0.2) is a sheddase that catalyzes peptide hydrolysis. It can cleave and release cell-surface proteins, e.g., TNF- α or ephrin. Elevated expression of ADAM10 has been found in melanoma metastasis²⁶. In HER2-overexpressing breast cancer cell lines, the usage of ADAM10 inhibitors can decrease cell proliferation²⁷.

EGF-like repeats and discoidin I-like domains 3 (DEL1; In ratio (PLX/DMSO) = -1.1 +/- 0.1) is a secreted β -3 integrin ligand that promotes adhesion of endothelial cells. It has been found to mediate angiogenesis in the context of cancer²⁸. In HCC, high expression of DEL1 correlates to poor prognosis²⁹.

Two proto oncogene tyrosine protein kinases, Src and Yes, were only identified in the DMSO-derived exosomes. These may be involved in the transmission of function

PLX-induced up regulation

The most interesting proteins that were found only in the exosomes derived from PLX treated melanoma cells, were receptor tyrosine kinases, VEGFR 1/2/3, and PDGFR- α/β .

Vascular endothelial growth factor receptor (VEGFR 1/2/3) is involved in vasculogenesis and angiogenesis. Many tumors exploit VEGF signaling for vascularization and migration³⁰. Knockdown of VEGFR-1 has been found to inhibit tumor growth in human hepatocellular carcinoma cells and clear cell renal cell carcinoma^{30,31}.

Platelet-derived growth factor receptor (PDGFR) is involved in regulating cell growth and division. There are two forms of which are only identified after PLX-treatment, PDGFR- α and PDGFR- β . PDGF can stimulates the B-Raf pathway, and is involved in a number of cancers, including melanoma and breast cancer³². PDGF signaling is the target of a imatinib³³. What is interesting here is that upregulation of PDGF- β is a mechanism of PLX resistance.

Upregulation of these receptor tyrosine kinases has been demonstrated to result from the release of feedback inhibition following treatment with MAPK targeted therapeutics. Increased levels of these proteins in exosomes could be useful for validating target inhibition.

Treatment of melanoma cells with PLX-4720 alters the protein composition of the derived exosomes *in vitro*. Proteins affected may be functionally relevant or early indicators of resistance response. These data suggest that exosomal proteins can be used to monitor cellular responses to B-Raf inhibition.

Melanoma exosome data for ExoCarta

The only melanoma data in the ExoCarta database is from a 2D gel electrophoresis study, in which 48 proteins were identified³⁴. The quantity of identifications is lacking for a comprehensive reference, and this is partially due to the generic sample preparation scheme. The solubilization and digestion schemes employed by many experiments that serve as data sources for ExoCarta are not optimal for membrane proteins. Inclusion of membrane protein targeted proteomic sample preparation methods, like eFASP, will significantly increase the quality of the reference information in ExoCarta. We will be submitting all of our melanoma and breast cancer exosome protein data to ExoCarta.

Over the past few years, a number of exosome samples, derived from different cells lines, were profiled for protein content. A collection of these data will be prepared and submitted to the ExoCarta DB, to provide a comprehensive inventory of proteins associated with exosomes from melanoma cells (M202 and M249), as well as breast cancer cells (BT474). Data for M202 and M249 cell lines are derived from shotgun-LC-MS^E, reversed phase-reverse phase 2D-LC-MS^E, and HILIC-LC-MS^E analyses. The number of shotgun data sets and merged data sets was 24, and 44, for M202 and M249, respectively. BT474 derived exosomes were only analyzed by shotgun-LC-MS^E, totaling 14 data sets.

A total of 33,451 proteins were identified, with 1,097 being identified in at least 2 data sets. 338 proteins were matched to entries already in ExoCarta.

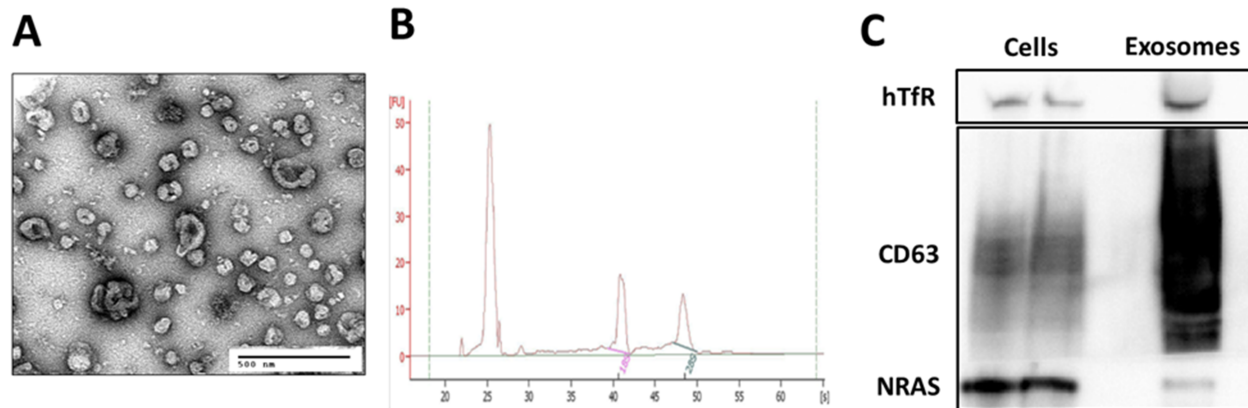


Figure 6-1. Isolation and characterization of exosomes secreted from M202 melanoma cells. (A) Transmission electron microscopy analysis of purified exosomes. (B) Bioanalyzer trace of exosomal RNA. The peak at 25 seconds is low molecular weight RNA. (C) Western blot analysis of exosomal marker CD63 in cells and exosomes.

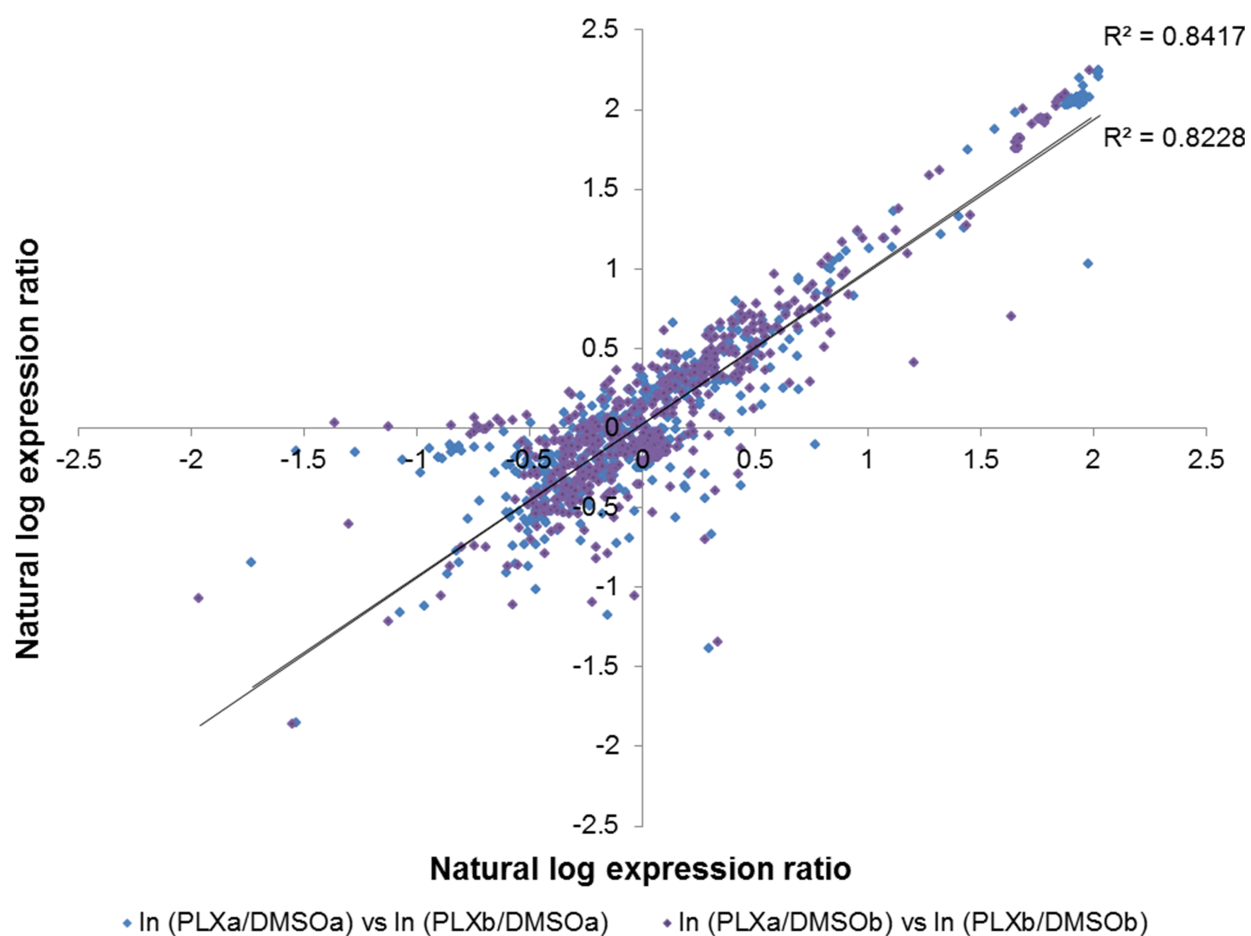


Figure 6-2. Concordance of biological replicates analyzed by HILIC-LC-MS^E. Each series represents the comparison of protein ratios determined for each PLX-treated replicate (PLXa or PLXb) relative to only one DMSO-treated replicate (DMSOa or DMSOb). Ln protein expression ratio comparisons calculated relative to DMSOa are in light blue and relative to DMSOb are in purple.

Enriched in Exosomes following PLX4720 treatment

Uniprot Accession	Gene Symbol	In Ratio (PXL4720/DMSO)
P62241	RS8	1.97
Q71DI3	H32	1.96
P84243	H33	1.95
Q6NWP9	HIST1H3A	1.94
Q16695	H31T	1.94
P62805	H4	1.90
Q6NXT2	H3C	1.88
Q8IUE6	H2A2B	1.86
Q96KK5	H2A1H	1.85
Q93077	H2A1C	1.85
P16104	H2AX	1.85
P04908	H2A1B	1.84
P20671	H2A1D	1.84
Q6FI13	H2A2A	1.84
Q16777	H2A2C	1.84
Q96QV6	H2A1A	1.84
Q99878	H2A1J	1.84
Q7L7L0	H2A3	1.84
P0C0S8	H2A1	1.84
Q9BTM1	H2AJ	1.83
P0C0S5	H2AZ	1.83
Q8N257	H2B3B	1.83
Q71UI9	H2AV	1.82
P58876	H2B1D	1.82
Q5QNW6	H2B2F	1.82
P23527	H2B1O	1.82
Q16778	H2B2E	1.82
Q93079	H2B1H	1.81
P06899	H2B1J	1.81
Q96A08	H2B1A	1.81
O60814	H2B1K	1.81
P57053	H2BFS	1.81
Q99880	H2B1L	1.80
Q99879	H2B1M	1.80
P62807	H2B1C	1.80
Q99877	H2B1N	1.80
P33778	H2B1B	1.80
Q07020	RL18	1.68
Q14764	MVP	1.44
P13010	XRCC5	1.40
P18124	RL7	1.13
Q00610	CLH1	1.11
P53675	CLH2	1.04
P47914	RL29	0.96
P62888	RL30	0.95
P35580	MYH10	0.88
Q03135	CAV1	0.86
Q14204	DYHC1	0.84
Q86VU1	PSMD1	0.81
P17987	TCPA	0.77
P09496	CLCA	0.75
P49368	TCPG	0.73

Uniquely identified in Exosomes following PLX4720 treatment

Uniprot Accession	ID	Frequency of identification in PLX4720 analyses
P10412	H14	10
P16403	H12	10
P16402	H13	10
Q9BW65	RPL14	10
P12956	XRCC6	10
Q6PIM9	HP1BP3	10
O75531	BAF	10
P17643	TYRP1	10
Q08211	DHX9	10
P27635	RL10	9
Q6FH47	GNB2L1	9
O75367	H2AY	9
P11717	MPRI	9
P61313	RL15	8
P61247	RS3A	8
Q05519	SRS11	7
P32004	L1CAM	7
Q92841	DDX17	7
Q12840	KIF5A	7
P26640	SYVC	6
P49841	GSK3B	6
P46777	RL5	6
P62917	RL8	6
P62263	RS14	6
Q04837	SSBP	6
Q00839	HNRPU	6
P35968	VGFR2	5
Q86X73	EPRS	5
P09619	PGFRB	5
P84103	SRSF3	5
O00487	PSDE	5
P62249	RS16	5
Q9UNM6	PSD13	4
P17948	VGFR1	4
P16234	PGFRA	4
P35916	VGFR3	4
P08237	K6PF	4
P42677	RS27	4
Q71UM5	RS27L	4
P62906	RL10A	4
P27361	MK03	4
Q6S381	PLEC	4
P62847	RS24	4
O15127	SCAM2	4
O00571	DDX3X	4
P23396	RS3	4
P18621	RL17	4
Q9Y678	COPG	4
Q86U45	HNRNPC	3
P53618	COPB	3
P62333	PRS10	3
Q9UBE8	NLK	3

O00468	AGRIN	0.73
Q7KZF4	SND1	0.68
Q9NR45	SIAS	0.68
P50991	TCPD	0.65
P61160	ARP2	0.65
P35749	MYH11	0.65
Q86VC0	RPSA	0.63
Q9NRY6	PLS3	0.62
P60201	MYPR	0.60
Q9CWI7	RAN	0.60
P50990	TCPQ	0.57
P35579	MYH9	0.57
P31939	PUR9	0.56
P15586	GNS	0.53
Q9Y4I1	MYO5A	0.52
Q96IT0	FASN	0.52
P48643	TCPE	0.49
Q08431	MFGM	0.49
O75054	IGSF3	0.47
P22314	UBA1	0.45
Q9UC36	HSPB1	0.44
Q99867	TBB4Q	0.43
O75083	WDR1	0.42
O43175	SERA	0.42
Q3ZCM7	TBB8	0.40
P23526	SAHH	0.39
P68371	TBB4B	0.35
P68363	TBA1B	0.33
Q71U36	TBA1A	0.33
P68366	TBA4A	0.32

P36888	FLT3	3
P61916	NPC2	3
P62266	RS23	3
P14868	SYDC	3
P62318	SMD3	3
Q14344	GNA13	3
P62910	RL32	3
P27816	MAP4	3
P46779	RL28	3
P06703	S10A6	3
P22087	FBRL	3
Q16658	FSCN1	3
Q9UCS3	TPM4	3
P07951	TPM2	3
Q9ULV0	MYO5B	3
P26373	RL13	3
P02545	LMNA	2
Q96NG7	DYNC1I2	2
Q8IWA5	CTL2	2
P63151	2ABA	2
Q96NS1	YPEL4	2
Q16555	DPYL2	2
O60911	CATL2	2
Q00013	EM55	2
O60282	KIF5C	2
Q14914	PTGR1	2
Q99798	ACON	2
Q13263	TIF1B	2
Q9NRW1	RAB6B	2
Q86W92	LIPB1	2
P21964	COMT	2
Q13310	PABP4	2

Table 6-1. Proteins enriched in PLX-4720 treated cells. Proteins upregulated in PLX-4720 treated cells, compared to control, are in the left table, with their Uniprot accession number, their gene ID, and the natural log of the abundance ratio. Proteins unique to PLX-4720 treated cells are in the right table, with their Uniprot accession number, their gene ID, and the frequency of identification across the technical replicate analyses.

Depleted from Exosomes following PLX4720 treatment

Uniprot Accession	Gene Symbol	In Ratio (PXL4720/DMSO)
Q9HAP3	FN1	-1.41
P35442	TSP2	-1.70
O43854	EDIL3	-1.15
Q15758	AAAT	-1.01
Q9UQV9	CEACAM1	-0.81
P31689	DNJA1	-0.79
O14672	ADA10	-0.85
O15162	PLS1	-0.74
P61225	RAP2B	-0.53
P51149	RAB7A	-0.53
Q969P0	IGSF8	-0.72
P54709	AT1B3	-0.52
Q9UEG1	UBA52/RPS27A/UBB/UBC	-0.52
P54707	AT12A	-0.62
Q13683	ITA7	-0.57
P11021	GRP78	-0.44
Q13733	AT1A4	-0.56
P13637	AT1A3	-0.51
P05023	AT1A1	-0.51
P20648	ATP4A	-0.47
A6NIZ1	RP1BL	-0.46
Q9UPK7	UBA52/RPS27A/UBB/UBC	-0.48
Q9UEK8	UBA52/RPS27A/UBB/UBC	-0.48
O15431	COPT1	-0.80
P50993	AT1A2	-0.49
P61224	RAP1B	-0.42
P05362	ICAM1	-0.44
P84077	ARF1	-0.41
Q9P2B2	FPRP	-0.49
P20020	AT2B1	-0.38
O00161	SNP23	-0.39
P35241	RADI	-0.36
Q8WTV0	SCRIB1	-0.41
Q8WUM4	PDC6I	-0.46
P15311	EZRI	-0.35
P54764	EPHA4	-0.51
Q6UVK1	CSPG4	-0.41
P11142	HSP7C	-0.39
P54652	HSP72	-0.36
P13987	CD59	-0.35

Not identified in Exosomes following PLX4720 treatment

Uniprot Accession	ID	Frequency in DMSO
Q96K70	ITIH5	10
Q9NPH3	IL1AP	9
Q9H0H5	RGAP1	9
O95819	M4K4	7
Q92626	PXDN	7
P24821	TENA	7
P61604	CH10	6
Q9H9H4	VP37B	6
P05026	AT1B1	6
P63241	IF5A1	6
Q8NG11	TSN14	6
O94887	FARP2	6
Q14126	DSG2	5
Q14160	SCRIB	5
P53990	IST1	4
Q5JXB2	UE2NL	4
P07947	YES	4
P12931	SRC	4
O95573	ACSL3	4
Q13884	SNTB1	4
O95164	UBL3	3
O75954	TSN9	3
Q2TAP0	GOG7B	3
P04083	ANXA1	3
O43852	CALU	3
Q9UN37	VPS4A	3
P13645	K1C10	3
Q9ULC3	RAB23	3
O60462	NRP2	3
Q96CS7	PKHB2	3
Q9H2W4	NEDD4L	3
P54756	EPHA5	3
Q9UC78	SERPINC1	3
Q99816	TS101	3
Q14847	LASP1	3
P61088	UBE2N	3
O15403	MOT7	3
Q99988	GDF15	3
Q4KMQ2	ANO6	2
Q9P0V0	SPTAN1	2
P21108	PRPS3	2
P20742	PZP	2
P48029	SC6A8	2
P50570	DYN2	2
P19012	K1C15	2
O75347	TBCA	2
O43491	E41L2	2
P28065	PSB9	2
Q9NP97	DLRB1	2
O15498	YKT6	2
P60900	PSA6	2
P29320	EPHA3	2

Table 6-2. Proteins depleted in PLX-4720 treated cells. Proteins downregulated in

PLX-4720 treated cells, compared to control, are in the left table, with their Uniprot accession number, their gene ID, and the natural log of the abundance ratio. Proteins unique to DMSO treated cells are in the right table, with their Uniprot accession number, their gene ID, and the frequency of identification across the technical replicate analyses.

REFERENCES

1. Théry, C., Ostrowski, M. & Segura, E. Membrane vesicles as conveyors of immune responses. *Nat Rev Immunol* **9**, 581–593 (2009).
2. Simpson, R. J., Jensen, S. S. & Lim, J. W. E. Proteomic profiling of exosomes: Current perspectives. *Proteomics* **8**, 4083–4099 (2008).
3. Mathivanan, S., Ji, H. & Simpson, R. J. Exosomes: extracellular organelles important in intercellular communication. *J Proteomics* **73**, 1907–1920 (2010).
4. Schorey, J. S. & Bhatnagar, S. Exosome function: from tumor immunology to pathogen biology. *Traffic* **9**, 871–881 (2008).
5. Gan, X. & Gould, S. J. Identification of an inhibitory budding signal that blocks the release of HIV particles and exosome/microvesicle proteins. *Mol Biol Cell* **22**, 817–830 (2011).
6. Fang, Y. *et al.* Higher-order oligomerization targets plasma membrane proteins and HIV gag to exosomes. *PLoS Biol* **5**, e158 (2007).
7. Coleman, B. M., Hanssen, E., Lawson, V. A. & Hill, A. F. Prion-infected cells regulate the release of exosomes with distinct ultrastructural features. *FASEB J* (2012).doi:10.1096/fj.11-202077
8. Hood, J. L., San, R. S. & Wickline, S. A. Exosomes released by melanoma cells prepare sentinel lymph nodes for tumor metastasis. *Cancer Res* **71**, 3792–3801 (2011).
9. Gonzalez-Begne, M. *et al.* Proteomic analysis of human parotid gland exosomes by multidimensional protein identification technology (MudPIT). *J Proteome Res* **8**, 1304–1314 (2009).
10. Gonzales, P. A. *et al.* Large-scale proteomics and phosphoproteomics of urinary exosomes. *J Am Soc Nephrol* **20**, 363–379 (2009).
11. Bard, M. P. Proteomic analysis of exosomes isolated from human malignant pleural effusions. *Am J Respir Cell Mol Biol* **31**, 114–121 (2004).
12. Hegmans, J. P. *et al.* Proteomic analysis of exosomes secreted by human mesothelioma cells. *Am J Pathol* **164**, 1807–1815 (2004).
13. Sondergaard, J. *et al.* Differential sensitivity of melanoma cell lines with BRAFV600E mutation to the specific Raf inhibitor PLX4032. *Exosomes_2012-07-07* **8**, 39 (2010).

14. Yang, H. *et al.* RG7204 (PLX4032), a selective BRAFV600E inhibitor, displays potent antitumor activity in preclinical melanoma models. *Cancer Res* **70**, 5518–5527 (2010).
15. Sambade, M. J. *et al.* Melanoma cells show a heterogeneous range of sensitivity to ionizing radiation and are radiosensitized by inhibition of B-RAF with PLX-4032. *Radiother Oncol* **98**, 394–399 (2011).
16. Yeung, Y.-G., Nieves, E., Angeletti, R. H. & Stanley, E. R. Removal of detergents from protein digests for mass spectrometry analysis. *Anal Biochem* **382**, 135–137 (2008).
17. Masuda, T., Tomita, M. & Ishihama, Y. Phase transfer surfactant-aided trypsin digestion for membrane proteome analysis. *J Proteome Res* **7**, 731–740 (2008).
18. Li, G.-Z. *et al.* Database searching and accounting of multiplexed precursor and product ion spectra from the data independent analysis of simple and complex peptide mixtures. *Proteomics* **9**, 1696–1719 (2009).
19. Silva, J. C., Gorenstein, M. V., Li, G. Z., Vissers, J. P. & Geromanos, S. J. Absolute quantification of proteins by LCMS^E: a virtue of parallel MS acquisition. *Mol Cell Proteomics* **5**, 144–156 (2006).
20. Silva, J. C. *et al.* Simultaneous qualitative and quantitative analysis of the *Escherichia coli* proteome: a sweet tale. *Mol Cell Proteomics* **5**, 589–607 (2006).
21. Silva, J. C. *et al.* Quantitative proteomic analysis by accurate mass retention time pairs. *Anal Chem* **77**, 2187–2200 (2005).
22. Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J Mol Biol* **157**, 105–132 (1982).
23. Dimmer, E. C. *et al.* The UniProt-GO Annotation database in 2011. *Nucleic Acids Res* **40**, D565–D570 (2012).
24. Binns, D. *et al.* QuickGO: a web-based tool for Gene Ontology searching. *Bioinformatics* **25**, 3045–3046 (2009).
25. Futagami, Y. *et al.* Role of thrombospondin-1 in T cell response to ocular pigment epithelial cells. *J Immunol* **178**, 6994–7005 (2007).
26. Lee, S. B. *et al.* ADAM10 is upregulated in melanoma metastasis compared with primary melanoma. *J Invest Dermatol* **130**, 763–773 (2010).
27. Liu, P. C. C. *et al.* Identification of ADAM10 as a major source of HER2 ectodomain sheddase activity in HER2 overexpressing breast cancer cells. *Cancer Biol Ther* **5**, 657–664 (2006).

28. Aoka, Y. *et al.* The embryonic angiogenic factor Del1 accelerates tumor growth by enhancing vascular formation. *Microvasc Res* **64**, 148–161 (2002).
29. Sun, J.-C. *et al.* High expression level of EDIL3 in HCC predicts poor prognosis of HCC patients. *World J Gastroenterol* **16**, 4611–4615 (2010).
30. Zhang, L. *et al.* VEGF is essential for the growth and migration of human hepatocellular carcinoma cells. *Mol Biol Rep* **39**, 5085–5093 (2012).
31. Li, C., Liu, B., Dai, Z. & Tao, Y. Knockdown of VEGF receptor-1 (VEGFR-1) impairs macrophage infiltration, angiogenesis and growth of clear cell renal cell carcinoma (CRCC). *Cancer Biol Ther* **12**, 872–880 (2011).
32. Suzuki, S., Heldin, C. H. & Heuchel, R. L. Platelet-derived growth factor receptor-beta, carrying the activating mutation D849N, accelerates the establishment of B16 melanoma. *BMC Cancer* **7**, 224 (2007).
33. Pirraco, A. *et al.* Imatinib targets PDGF signaling in melanoma and host smooth muscle neighboring cells. *J Cell Biochem* **111**, 433–441 (2010).
34. Mears, R. *et al.* Proteomic analysis of melanoma-derived exosomes by two-dimensional polyacrylamide gel electrophoresis and mass spectrometry. *Proteomics* **4**, 4019–4031 (2004).

CHAPTER 7

The early and latent effects of ionizing radiation on lymphoblastoid cells with functional and non-functional ATM.

ABSTRACT

Ataxia Telangiectasia, Mutated (ATM) is a nuclear protein kinase at the core of the cellular response to double-stranded DNA breaks induced by ionizing radiation (IR). Loss of functional ATM results in neurodegeneration, cancer predisposition and sensitivity to IR. Comparative proteomics utilizing label free quantitative LC- MS^E is employed to study differential regulation of gene expression at early and late time points following the exposure of lymphoblastoid cell lines (LCLs), with functional or aberrant ATM, to 5 Gy IR. Analysis of quantitative data revealed a number of differentially regulated proteins and pathways involved in cellular repair and survival resulting from exposure to IR or to radioprotector. Cytoprotective proteins involved in the Nrf2-mediated oxidative response are quickly up-regulated following IR exposure in an effort to quench damaging free radicals. TCA cycle and oxidative phosphorylation proteins are downregulated following IR exposure, which reduces mitochondrial ROS production. Glycolysis proteins are upregulated following IR exposure, potentially enabling the cell to compensate for the decrease in ATP production from suppressed cellular respiration and to generate sufficient ATP for cellular defense and repair. The observed up-regulation of lactate dehydrogenases would allow for NAD⁺ regeneration for further rounds of glycolysis. Western blotting has been used for validation of protein

abundance changes. In the absence of functional ATM, cells are still able to induce the expression of proteins that function in survival or cell cycle modulation. The mechanism of these radiation-induced changes is not clear, although, based on transcription factor analysis, there is an overlap with genes targeted by TP53, which is essential in ATM-dependent radiation responses. Thrombin signaling may also play a part in this alternative mechanism of survival and repair following IR insult. These indicate redundant pathways, that do not require ATM, which can effect responses to ionizing radiation exposure.

TABLE OF CONTENTS

INTRODUCTION.....	182
BACKGROUND	184
Ionizing radiation.....	184
Complexity reduction and sample analysis.....	185
Label-free absolute quantification with LC-MS ^E	186
Data Extraction and Correction (DECO)	187
EXPERIMENTAL.....	188
Strains and cell growth	188
Processing of 4-hour LCL samples.....	189
OFFGEL fractionation of peptides	189
Processing of 24-hour LCLs	190
Analysis by nanoLC-MS ^E	190
Data processing and protein identification for 4 hour LCLs	191
Data processing and protein identification for 24 hour LCLs	191
Protein classification and pathway analysis.....	192
Gel electrophoresis.....	192
Western blotting.....	193
RESULTS AND DISCUSSION FOR EARLY IR EFFECTS	195
Early effects of radiation: NRF2-oxidative stress response	198
Integrin signaling.....	199
Energy metabolism	201
Early effects of radiation: radiation modulated proteins	202

RESULTS AND DISCUSSION: LATENT EFFECTS	204
Latent effects of radiation on ATM-deficient LCLs by ERLIC-LC-MS ^E	204
Cell survival	204
Thrombin signaling is upregulated in OFFGEL analysis (AT24+IR vs. AT24-IR).....	205
Transcription factors	206
Latent effects of radiation on Normal LCLs by ERLIC-LC-MS ^E	207
IL-8 and the bystander effect	207
DNA repair and damage response	209
Latent oxidative stress response	210
Transcription factors	211
REFERENCES.....	212

INTRODUCTION

Ataxia Telangiectasia, Mutated (ATM) is a 350 kDa nuclear protein kinase at the core of the cellular response to double-stranded DNA breakage. ATM is responsible for sensing double-stranded DNA damage, due to genotoxic agents (ionizing radiation, anticancer therapeutics, etc.) or normal physiological processes (occurring during meiosis, immune system maturation and telomere maintenance), and initiating an appropriate cellular response. Signaling to pathways involved in cell cycle control, DNA repair and apoptosis are essential for ATM to exert the appropriate cellular effects.

Ataxia telangiectasia (A-T) is caused by loss-of-function mutations in the ATM gene. A-T is characterized early in childhood by ataxia and ocular telangiectasias. Patients experience progressive neurodegeneration, radiosensitivity, and have a propensity for lymphoid cancers ¹. The absence of functional ATM has been shown to result in a buildup of cellular oxidative stress, independent of IR exposure, in A-T patients as well as ATM deficient cell lines and animals ²⁻⁶. The cause of this buildup is not entirely clear; however, it is thought that chromosomal breaks may lead to oxidative stress indirectly through altered gene expression or directly through disruption of mitochondrial function and oxygen metabolism ³. Additionally, it is hypothesized that ATM may have a role in the regulation of cellular redox homeostasis since cells from A-T patients are sensitive to agents that induce ROS generation and oxidative DNA damage, and siRNA against ATM in normal cells decreases cellular respiration and the introduction of normal ATM increases respiration ⁷.

Two major studies have analyzed the effects of radiation on cells in culture, and used comparative proteomics to quantify the differential cellular effects. Matsuoka et al. utilized SILAC for a proteomic analysis of protein phosphorylation in response to DNA

damage by ionizing radiation in human embryonic kidney 193T cells⁸. Over 700 proteins were found to be phosphorylated following exposure to IR. Many of the identified kinase substrates were previously unassociated with the DNA damage response pathway. Although some of these substrates may be attributed to the oxidative stress response, and not direct IR DNA damage, the results highlight the large and complex network of proteins involved in cellular responses to IR damage whose functional significance has yet to be determined.

Stokes et al. investigated the phosphorylation cascades in response to DNA damage due to UV-radiation, and utilized SILAC to find the UV-inducible substrates for ATR⁹. The use of ATR-deficient cell lines allowed for them to come up with a list of about 192 phosphorylation sites that are induced by UV-radiation but absent without functional ATR. Together with Matsuoka's work, more clearly defined signaling pathways involved in cellular response to different kinds of DNA damage is possible. There was a large overlap between phosphorylation sites in IR and UV-radiation damage signaling.

A-T is a paradigm for cancer and neurodegenerative disorders, and will be utilized as a model for DNA damage and repair in order to elucidate the molecular processes fundamental to cellular sensitivity to IR, cellular protection from oxidative stress, and cellular response to IR induced DNA damage. The aims of this work are to quantify the changes in protein abundance following exposure to ionizing radiation, at levels consummate to high exposure, and to investigate radioprotector compounds shown to enhance survival following IR exposure. Results from these aims will help to better define the human immune cell response to IR, to determine pathways and

components that depend on ATM activity, and to describe pathways and mechanisms of radioresistance. To achieve these aims, we employ label free quantitative mass spectrometry and comparative proteomics to interrogate lymphoblastoid cell lines collected after exposure to radiation or drug treatment.

BACKGROUND

Ionizing radiation

Ionizing radiation refers to that portion of the electromagnetic spectrum with sufficient energy to pass through matter and to dislodge orbital electrons to form ions. There are two forms of IR: the wave form, gamma and X rays, and the particle form, alpha and beta particles and high energy neutrons.

There are many sources of IR exposure, from environmental, including, radon gas and cosmic rays. Medical diagnostics, including X rays, CT scans and angiograms make account for about 15% the yearly IR exposure for the average person. High doses of IR may be encountered in radiotherapy, with the most intense exposure occurring during industrial accidents or exposure to nuclear fallout.

IR affects all organs of the human body, but the immune system, the gastrointestinal system, and the brain are the most susceptible to IR. When talking about IR, it is convenient to refer to it in units of absorbed dose, 1 Grey is the equivalent of 1 joule of radiation absorbed per kilogram of body weight. Dosages between 2 and 6 Gy induce radiation sickness, with chances of survival decreasing steeply with higher dose. Anything over 10 Gy makes survival unlikely.

As radiation passes through a biological system, its energy is deposited in discreet regions known as spurs. Within a spur, free radicals can be produced through the radiolysis of water. These radicals can diffuse into the system and damage macromolecules, including proteins, lipids and DNA. IR can also induce direct DNA damage through the breakage of phosphodiester bonds, inducing double stranded breaks, which are the most lethal to the cell ¹⁰.

Complexity reduction and sample analysis

The dynamic range of protein concentrations and the quantity of different proteins is high, which results in even greater peptide digest complexity. Reduction of peptide complexity is essential when working with higher order organisms, especially human, in order to minimize the masking of low abundance peptides, and to enable in-depth proteome analysis. Two dimensions of complexity reduction are employed.

The first dimension separates the peptides by isoelectric point in liquid phase in the OFFGEL fractionation apparatus. Liquid fractions are collected and the peptides in each fraction are further separated based on hydrophobicity on a reversed phase UPLC column. As peptides elute off the column, they are ionized by ESI and enter a quadrupole-time-of-flight (qTOF) instrument, which alternately samples the peptide precursor masses, and the peptide fragment masses. The comprehensive collection of precursor ion and fragment ion masses is stored in a .RAW data file, and post-processed in ProteinLynx Global Server (PLGS) for peptide and protein identification and quantification information.

In order to recapitulate the original unfractionated peptide digest, the .RAW file for each fraction must be merged to produce one file that maintains data for all the fractions. This merged file is then searched in PLGS, against a protein database for protein identification, and then analyzed for determination of peptide quantities, from which protein abundance is inferred.

Label-free absolute quantification with LC-MS^E

LC-MS^E is a strategy for the acquisition of precursor and fragment ion data that increases the depth of analysis and enables absolute quantification of proteins from complex peptide mixtures. Unlike data dependent acquisition (DDA), in which a set number of precursor ions are selected for fragmentation, this non-selective approach continually collects precursor and fragment ion data in parallel by alternating between low and high collision energy MS functions ¹¹. Post-processing of the data with ProteinLynx Global Server (PLGS, Waters) provides a comprehensive inventory of precursor ions along with their respective fragment ions, which can then be search against a database for peptide and protein identification.

The high mass accuracy and comprehensive nature of LC-MS^E allows for the determination of absolute quantities of proteins from LC-MS data of complex peptide mixtures ^{12,13}. A known quantity of pre-digested protein is utilized as an internal standard and is spiked into a sample prior to LC-MS^E analysis. Once the sample data has been acquired, PLGS calculates the average MS signal for the 3 most intensely ionizing peptides (Hi3) from each protein identified, including the internal standard. Since the average signal response from the Hi3 for any given protein is similar, the

average Hi3 value for the internal standard is utilized as a reference point for absolute quantitation. The observed Hi3 value for any identified protein is quantified by normalizing to the average Hi3 of the internal standard, which has a known concentration. This method is advantages because it does not require isotope labeling of cells and will afford a more comprehensive qualitative and quantitative view of the cellular proteome.

Expression Analysis is used to quantify all proteins present in the sample, and can compare the sample to any, or all, other processed samples. The ratio, the natural log of the ratio, standard deviation, and the probability of up- or down-regulation are provided.

Data Extraction and Correction (DECO)

As with many data analysis software tools under development, there are issues that arise which require custom software to enable correct processing. The first problem concerns those peptides that can be identified in multiple fractions. Due to the way in which PLGS handles fraction merging, and the requirement for absolute consistency in retention time between LC-MS^E runs, there are instances in which the same peptide, found in multiple fractions, are not correctly merged (Chapter 2). In these cases, rather than the merge file containing 1 precursor ion for a peptide, which represents the summation of ion volumes for the peptide in all fractions it was identified in, the actual result is a merged file that contains information on 2 or more precursor ions, each with only a fraction of the ion volume data. One of the peaks will be identified correctly, and the other will not be identified, or will be identified as a much

lower scoring peptide. This can have great effects on the accuracy of protein quantification, as well as the possibility of increased false positive protein identifications. This is an issue that must be handled properly for experiments involving peptide or protein fractionation prior to LC-MS analysis.

The second problem relates to data accessibility, and can be an issue for any post-processing done with proprietary software, or open-source software that requires tools to interpret or access processed data. DECO enables the exportation of low level peptide data, as well as protein data, from PLGS and from Expression Analysis, which may be used for an analysis or visualization performed in other statistical software packages, like R, for the production of heatmaps, and other statistical tests of data quality or probability of differential regulation.

EXPERIMENTAL

Strains and cell growth

LEPU LCLs and NAT8 LCLs were grown in RPMI-1640 with 10% FBS and 1% PSG (pen/strep/glutamine) to a density of 10^7 cells (approximately 1mg of total protein). Duplicate plates for each LCL and timepoint were utilized. Specified flasks pretreated with tetracycline, or were exposed to 5 Gy ionizing radiation. LCLs were harvested at 4 hours and 24 hours following exposure of certain flasks to 5 Gy of IR.

Harvested LCLs were lysed by sonication in the presence of Laemmli Buffer (2% SDS, 0.1% DCA, 5 mM TCEP, 100 mM ABC, 1:100 protease inhibitor, 1:100 phosphatase inhibitor). 4-hour samples were centrifuged to remove cellular debris and the soluble fraction was subjected to overnight acetone precipitation. 24-hour samples

were processed by eFASP, without precipitation. The BCA Protein Assay (Pierce, 23227) was utilized to determine sample protein concentration.

Processing of 4-hour LCL samples

150 µg of protein was re-solubilized by sonication and vortexing in 100 µl of 1% sodium deoxycholate (Sigma Aldrich, D6750). Soluble proteins were reduced with DTT, alkylated with IAN, and subjected to overnight trypsin digestion (1:30 trypsin:protein). SDC was removed from the peptide samples using the phase transfer method, which involves the addition of ethyl acetate and the acidification with TFA ¹⁴. The aqueous phase was removed, and suspended in OFFGEL Buffer prior to IEF.

OFFGEL fractionation of peptides

Peptides were separated by OFFGEL isoelectric focusing into 12 or 24 fractions (pH 3-10) on 12-cm or 24-cm pH 3-10 non-linear IEF strips (GE Healthcare) with peptide fractionation performed with default parameters (maximum current 50 µA, with voltages ranging from 200-4000 until 50 kVh was reached). Fractions were recovered, dried down, and suspended in 50 µl of 5% formic acid. Each fraction was mixed in a 1:1 ratio with MassPREP Enolase Digestion Standard (Waters) to a final enolase digest concentration of 10 fmol/µl.

Processing of 24-hour LCLs

One set of samples was processed with 4-hour LCLs, and fractionated by OFFGEL IEF, and a second set of samples were processed by eFASP and fractionated with ERLIC as described in Chapter 8.

Analysis by nanoLC-MS^E

nanoUPLC of sample peptides was performed on an Waters nanoACQUITY UPLC system equipped with a nanoACQUITY UPLC Symmetry C18 180 μ m x 20 mm trap column (Waters, 186003514), and a nanoACQUITY UPLC BEH C18 75 μ m x 150 mm reversed phase analytical column (Waters, 186003543). The aqueous mobile phase (Mobile Phase A) contained 0.1% formic acid in water and the organic mobile phase (Mobile Phase B) contained 0.1% formic acid in acetonitrile. 5 μ l of the sample was loaded onto the trap with and washed with 99% A and 1% B for 5 minutes, and then injected onto the column with 3% Mobile Phase B. Peptides were eluted from the column with a gradient of 3-40% B over 90 minutes at 300 nl/min, followed by a 5 minute rinse with 95% B. The column was re-equilibrated to 97% A for 20 minutes.

The lock mass compound, [Glu¹]fibrinopeptide (0.5 pmol/ μ l, 25% ACN, 0.1% formic acid), and the analyte were delivered at 300 nl/min to the NanoLockSpray dual electrospray ion source (Waters, UK) which was equipped with a microcapillary for the lockmass compound, and a PicoTip Emitter (New Objective) for the analyte. Two electrosprays are produced which may be selectively sampled by positioning of the baffle.

The ion source was interfaced with a Waters Synapt HDMS for mass spectrometric analysis. Acquisition was performed in positive nanoESI mode and the instrument was operated in V-mode with an average resolution of 9500. Accurate mass data for precursor and fragment ions were collected over the 100-2000 m/z range, with the mode of acquisition alternating between low and elevated energy at 1-second intervals, for the duration of analysis. The lock mass channel was sampled every 30 seconds for calibration of MS and MS^E data.

Data processing and protein identification for 4 hour LCLs

Raw data files were processed in ProteinLynx Global Server 2.4 (PLGS, Waters) to provide an inventory of precursor ions along with their respective fragment ions, to search against the Swissprot Human database (April 2010) for peptide and protein identifications, and to provide quantitative information¹⁵. The Waters Expression System Informatics component of the PLGS software was utilized to cluster data^{12,13,16}. Methionine oxidation was set to variable modification and carbamidomethyl set to fixed modification variable modifications.

Data processing and protein identification for 24 hour LCLs

Raw data files for the 24-hour samples were processed in ProteinLynx Global Server 2.5 (PLGS, Waters) and analyzed as described in Chapter 8.

Protein classification and pathway analysis

Data Extraction and Correction (DECO) software was used to correct merged ion volumes for peptides identified in multiple fractions. Uniprot-GOA¹⁷ was searched in the QuickGO! browser to obtain compartment and function annotations for all proteins¹⁸. Data sets were further analyzed through the use of IPA (Ingenuity® Systems, www.ingenuity.com) for functional and pathway information.

Gel electrophoresis

Each 4-hour sample (3 µg) was aliquoted into tubes for SDS-PAGE analysis. Samples were boiled for 5 minutes with NuPAGE® 4X LDS Sample Buffer (Invitrogen) and NuPAGE® 10X Sample Reducing Agent (Invitrogen), and spun down. A NuPAGE® Novex 4-12% Bis-Tris Gel (1.0 mm, 12 well, Invitrogen) was placed into the Novex XCell SureLock® Mini-Cell (Invitrogen, EI001) and filled with running buffer. The samples were loaded along with a ladder mix (3:1, SeeBlue® Plus2 Pre-Stained Standard: MagicMark™ XP Western Protein Standard). Proteins were separated on ice, at 200 V, with 125 mA per gel, for 90 minutes, with MES buffer system, and 200 V, with 80 mA per gel, for 80 minutes, with MOPS buffer system. The tank was placed in an ice during electrophoresis. The gel was removed from the cassette, and the proteins were fixed in 50% methanol / 7% acetic acid for 15 minutes. The gel was washed in water three times, and then incubated with GelCode Blue Reagent (Pierce) and processed by the manufacturers instructions. The gel was visualized in a flat bed scanner.

Western blotting

Immobilon 0.45 μ m PVDF membranes, Whatman paper and sponges were soaked in 1X Transfer Buffer (Invitrogen). One PAGE gel was rinsed in water, and placed onto a PVDF membrane, and sandwiched on both sponges by Whatman paper and sponges. The sandwich was placed into the Novex XCell II™ Blot Module (Invitrogen), and then inserted into the Novex XCell SureLock® Mini-Cell (Invitrogen). The Blot Module was filled with Transfer Buffer until the buffer level just exceeded the top of the sponges. Water was added to the tank. Proteins were transferred from the SDS-PAGE gel to the PVDF membrane at 35 V, 50 mA/gel, overnight in a cold room.

After transfer, the PVDF membrane was removed from the sandwich and washed with TBST for 5 minutes. Even protein transfer was verified using the Pierce Reversible Protein Stain Kit for PVDF Membranes Kit (Pierce, 24585) as described. The membrane was blocked with Blocking Buffer (5% Blotto (Santa Cruz Biotechnology) or 3% BSA in TBST) for 1 hour, and washed three times, for 5 minutes each.

Antibody	Product	Source	Primary Dilution	Secondary Dilution
Plastin L [LPL4A.1]	ab80729	abCam	1:40000	1:10000
GSTP1 (3F2)	3369S	Cell Signaling	1:5000	1:10000
Nucleophosmin [FC82291]	ab10530	abCam	1:50000	1:50000
MDH2 antibody	ab92867	abCam	1:2000	1:5000
GAPDH (7B)	sc-69778	Santa Cruz	1:100000	1:50000
SSRP1 (3E4)	sc-56782	Santa Cruz	1:2000	1:5000
PARP-1 (F-2)	sc-8007	Santa Cruz	1:2000	1:5000
SPT16 (A-1)	sc-165987	Santa Cruz	1:2000	1:5000
Peroxiredoxin 3 [12B]	ab16751	abCam	1:1000	1:5000
Lamin A/C (636)	sc-7292	Santa Cruz	1:500	1:2500
Triosephosphate isomerase	ab58327	abCam	1:665	1:2500
Lamin B1 (ZL-5)	sc-56145	Santa Cruz	1:500	1:2500
LDH-B (431.1)	sc-100775	Santa Cruz	1:5000	1:2500
Ku-70 (E-5)	sc-17789	Santa Cruz	1:10000	1:5000
Nucleophosmin [FC82291]	ab10530	abCam	1:50000	1:3000
CD48 (4H9)	ab21058	abCam	1:50000	1:3000
SAHH (A-11)	sc-8397	Santa Cruz	1:500	1:3000
BTF3a/b (628C5a)	sc-271389	Santa Cruz	1:5000	1:10000
beta Tubulin (HRP)	sc-81206	Santa Cruz	1:2000	

Table 7-1. Mouse antibodies for Western Blotting. Product information, vendor source, and dilutions of primary antibody, and secondary anti-mouse antibody.

The membrane was incubated with a primary antibody at 4°C (Table 7-1). Membranes were washed with TBST, three times, for 5 minutes each, and then incubated with the appropriate secondary antibody prior to visualizing.

Re-probing of blots occurred after stripping with either mild or harsh stripping reagents, Restore Western Blot Stripping Buffer (Thermo) or Restore PLUS Western Blot Stripping Buffer (Thermo), respectively.

Membranes ready for visualization were incubated with ECL Plus Western Blotting Detection Reagent (1000 cm²) (RPN2132) as described by the manufacturer. Incubated membranes were inserted into a development cassette. In a developing dark room, Pierce CL-Xposure* Film (Thermo Scientific, 8x10, PI34091), was placed on top

of the membrane and exposed for a set duration, and then the film was removed and developed. This was repeated until the optimum exposure was obtained.

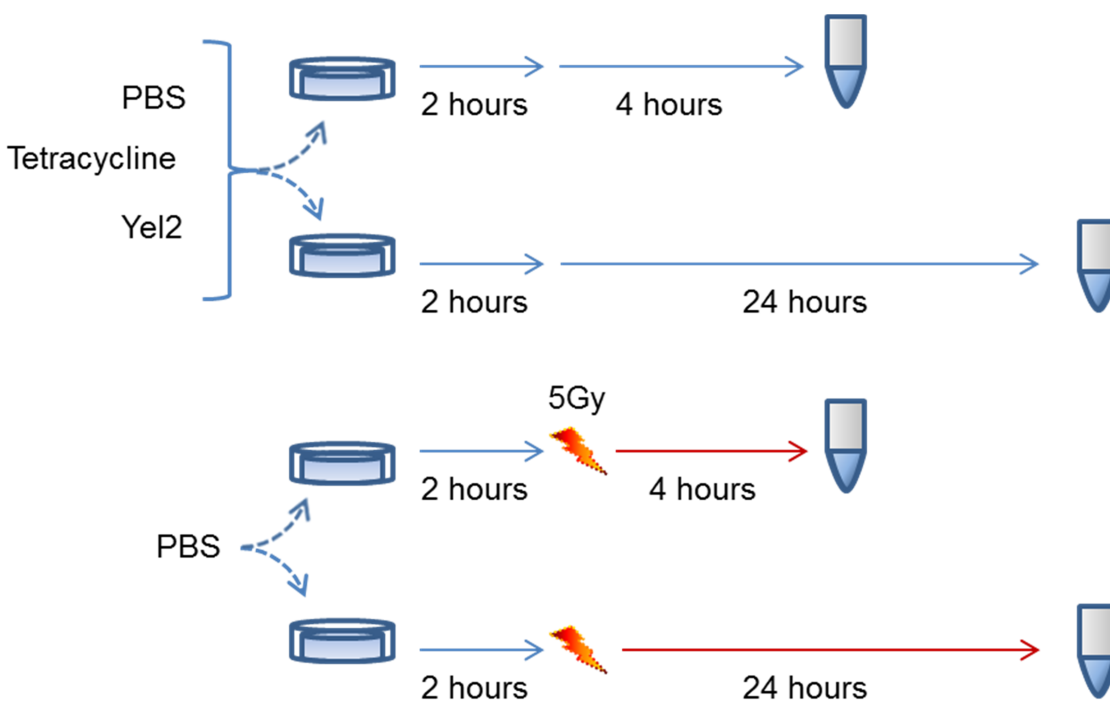


Figure 7-1. Timecourse analysis of drug and radiation proteome perturbations. Treated cells will be grown and treated with PBS, tetracycline, or YEL2, incubated with drug for 2 hours, and then harvested after 4 or 24 hours. Irradiated cells will incubate with PBS treated media for 2 hours, be exposed to 5 Gy IR, and will then be harvested after 4 and 24 hours. Biological duplicates for each condition will be generated and analyzed by OFFGeLC-MS^E.

RESULTS AND DISCUSSION FOR EARLY IR EFFECTS

Biological duplicate flasks of ATM-deficient LCLs and Normal LCLs were incubated with tetracycline or PBS for 2 hours, and then the cells were harvested either before or 4 hours after exposure to 5 Gy IR (Figure 7-1). Proteins from harvested LCL cultures were isolated and subjected to in-solution trypsin digestion. Peptides were separated by isoelectric point using OFFGEL isoelectric focusing into 12 fractions; each

fraction was analyzed in quadruplicate by LC-MS^E on a Synapt HDMS QTOF mass spectrometer. Raw data files were processed in ProteinLynx Global Server (PLGS) to produce an inventory of associated precursor ions and fragment ions, to search the inventory against the Swissprot database for peptide and protein identifications, and to provide quantitative protein information. DECO, a software resource we have created, facilitates the proper merging of fractionated peptide data, assuring accurate summed ion volume calculations. Ingenuity pathway analysis, as well as GO and KEGG databases, were used to group differentially regulated proteins by function or associated pathways. Western blotting has been used for validation of specific protein abundance changes.

Approximately 1,200 proteins were identified in each technical replicate, of which more than half were quantified. In normal LCLs, four hours after radiation exposure, 90 proteins are downregulated ($\ln$ ratio < -0.5), and 101 are upregulated ($\ln$ ratio > 0.5). In normal LCLs treated with tetracycline, four hours after radiation exposure, 165 are downregulated, and 71 are upregulated, with the same ratio cut-offs.

The average MS signal for the 3 most intensely ionizing peptides (Hi3) for each protein in a sample was utilized to determine the abundance in each sample and the differences between the samples. 395 proteins were quantified in at least 2 technical replicates of each biological sample (Figure 7-2). IR exposure has widespread effects on protein abundance, while tetracycline is subtler.

Ingenuity Pathway Analysis was used to analyze the quantitative data and put the results in the context of biological processes for radio-responsive proteins. The data input were 440 proteins that were upregulated or uniquely identified post-IR, and 620

proteins that were found to be downregulated or uniquely identified in non-exposed cells. Three of the main pathways affected were NRF2 oxidative stress response, major components of energy metabolism, and integrin signaling.

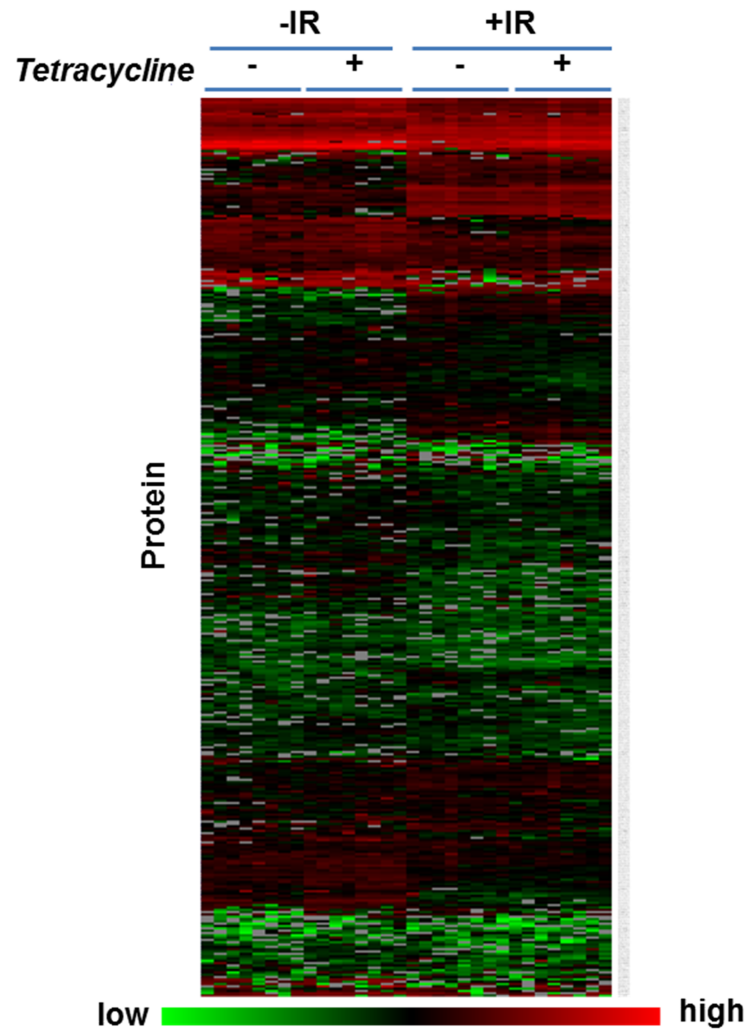


Figure 7-2. Ionizing radiation induces significant early proteome shifts. Ionizing radiation induces significant proteome shifts. The heat map reflects the quantitative data for the 395 most well represented proteins across all samples. Normal LCLs were pre-treated with control or tetracycline, and then harvested either before (N0) or after (N4) 5 Gy IR exposure. Each column represents protein quantities determined from one OFFGeLC-MS^E technical replicate. 32 technical replicates are represented, 8 technical replicates per condition. Red indicates high protein abundance, while green represents low protein abundance in a sample.

Early effects of radiation: NRF2-oxidative stress response

The NRF2 pathway is involved in the response to oxidative stress, and has been found upregulated in our dataset following IR exposure (Figure 7-3). NRF2 is a transcription factor that translocates to the nucleus, and binds to antioxidant response elements on genes that code for detoxifying and antioxidant genes. Genes activated by NRF2 include glutathione S-transferase pi (GST-Pi), superoxide dismutase (SOD), and catalase (CAT). The upregulation of the Nrf2 pathway following radiation exposure in mice has been identified, but the effect was only detectable 5-days after ¹⁹. Further work showed that a loss of Nrf2 enhances cellular radiosensitivity.

Components of the NRF2 pathway have been previously associated with radiation exposure, including GST-Pi, in CHO cells exposed to 4.5 Gy IR ²⁰. This upregulation was found at the mRNA and protein level ^{20,21}.

The upregulation detected in human LCLs is much earlier than previously found in murine systems. GST-Pi maintains peroxidase activity and is important for dealing with IR-produced lipid peroxides and DNA hydroperoxides. Following IR exposure, normal LCLs increase levels of GST-Pi 2.5-fold, while no change is observed in radiosensitive LCLs.

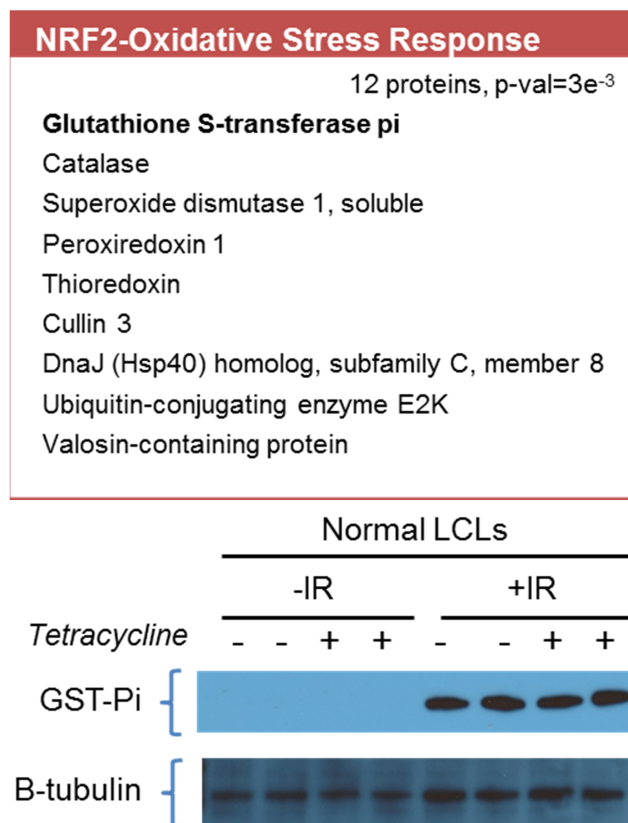


Figure 7-3. Protein components of NRF-2 oxidative stress response signaling that are upregulated following IR exposure. Ingenuity Pathway Analysis identified NRF-2 oxidative stress response as being significantly enriched in the irradiated data set, with 12 proteins clustered, and a p-value of enrichment of $3e^{-3}$. A number of the proteins involved are presented along with the cellular compartment to which they are associated. To the right is a western blot validation of GST-Pi upregulation following IR exposure.

Integrin signaling

The integrins are a major class of transmembrane molecules that mediate cell adhesion and cell-ECM interactions. They can transmit biochemical cues that can facilitate multiple cell fates, including apoptosis and survival. B1 integrins have been implicated in cancer cell resistance to IR and chemotherapy, presumably via the PI3K pathway and AKT ²². Components involved in integrin signaling, spanning many cellular

locations, are upregulated following radiation exposure, particularly B2 integrin, and downstream effectors, MAPKs and PI3Ks (Table 7-2). This upregulation post-IR lends evidence to the theory that, in addition to DNA damage, multiple cellular mechanisms, like cell interactions with neighboring cells and the microenvironment, fundamentally influence cell fate in response to IR exposure.

Integrin Signaling	
	24 proteins, p-val=3e ⁻¹²
Plasma Membrane	Cytosol
B2-integrin	Actin related protein 2/3 complex, subunit 2, 34kDa
FYN oncogene related to SRC, FGR, YES	Actinin, alpha 1
ARP3 actin-related protein 3 homolog	Actinin, alpha 3
Vinculin	ADP-ribosylation factor 3
Zyxin	Cell division cycle 42 (GTP binding protein, 25kDa)
	Mitogen-activated protein kinase 1
	Mitogen-activated protein kinase 3
	Phosphoinositide-3-kinase, class 3
	Phosphoinositide-3-kinase, regulatory subunit 4
	Protein phosphatase 1, catalytic subunit, beta isozyme
	Protein phosphatase 1, regulatory subunit 12A
	Rho guanine nucleotide exchange factor (GEF) 7
	V-crk sarcoma virus CT10 oncogene homolog (avian)-like
	WAS/WASL interacting protein family, member 1
	Wiskott-Aldrich syndrome (eczema-thrombocytopenia)
Nucleus	
Actinin, alpha 2	
Formin binding protein 1	
Talin 2	

Table 7-2. Protein components of integrin signaling that are upregulated following IR exposure. Ingenuity pathway Analysis identified Integrin signaling as being significantly enriched in the irradiated data set, with 24 proteins clustered, and a p-value of enrichment of 3e⁻¹⁵. A number of the proteins involved are presented along with the cellular compartment to which they are associated.

Energy metabolism

Following IR exposure, the main energy pathways are differentially regulated, with an observed increase in glycolytic proteins and a decrease in proteins involved in the TCA cycle and oxidative phosphorylation.

The downregulation or inhibition of the TCA cycle and oxidative phosphorylation has been shown to confer radioresistance in radiosensitive bacteria, rats, hamster. Decreased TCA cycle components post IR, rats ²³, CHO cells ²⁴, *D. radiodurans* ²⁵. ETC deficiency prevents radiation induced ROS/RNS in CHO cells ²⁶. The consensus is that decreasing these pathways prevents additional reactive oxygen species production for the cell to deal with. While glycolysis has less information on regulation in this context, its involvement in radioresistance was established in studies showing that inhibiting glycolysis at different points conferred radiosensitivity, presumably due to a lack of ATP for cellular repair processes ²⁷⁻²⁹.

Our data set brings these two findings together (Figure 7-4). We propose that, following radiation exposure, LCLs respond by suppressing cellular respiration to reduce ROS production, and by up-regulating glycolytic proteins to compensate for the decrease in ATP production and to generate sufficient ATP for cellular defense and repair. The observed up-regulation of lactate dehydrogenases allows for NAD⁺ regeneration for further rounds of glycolysis. Western blotting has been used for validation of protein abundance changes.

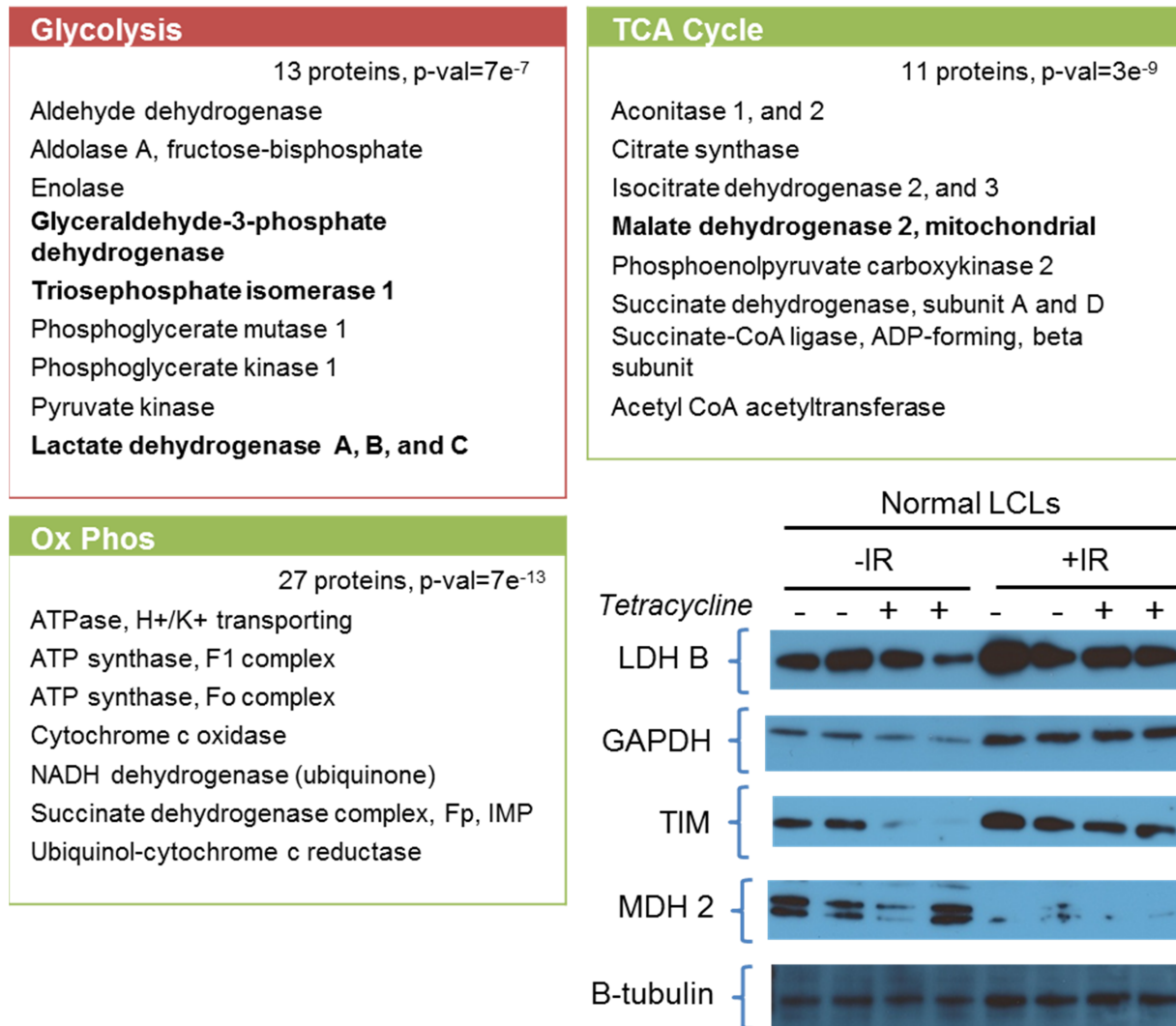


Figure 7-4. Differential regulation of pathways involved in energy metabolism following IR exposure. Ingenuity Pathway Analysis identified glycolysis as significantly enriched following IR exposure, whereas IPA found the TCA cycle and oxidative phosphorylation to be depleted following IR exposure. The total number of proteins identified by IPA, and their p-value are displayed. Western blot validation of proteins in bold is shown at the bottom.

Early effects of radiation: radiation modulated proteins

Proteins were annotated with information functional and structural information from the KEGG and GO databases. Differentially regulated proteins with common attributes were selected and validated. Proteins that were up or down regulated

following IR exposure include those involved in transcription and translation, anti-apoptosis, immune function, and structure and maintenance (Figure 7-5).

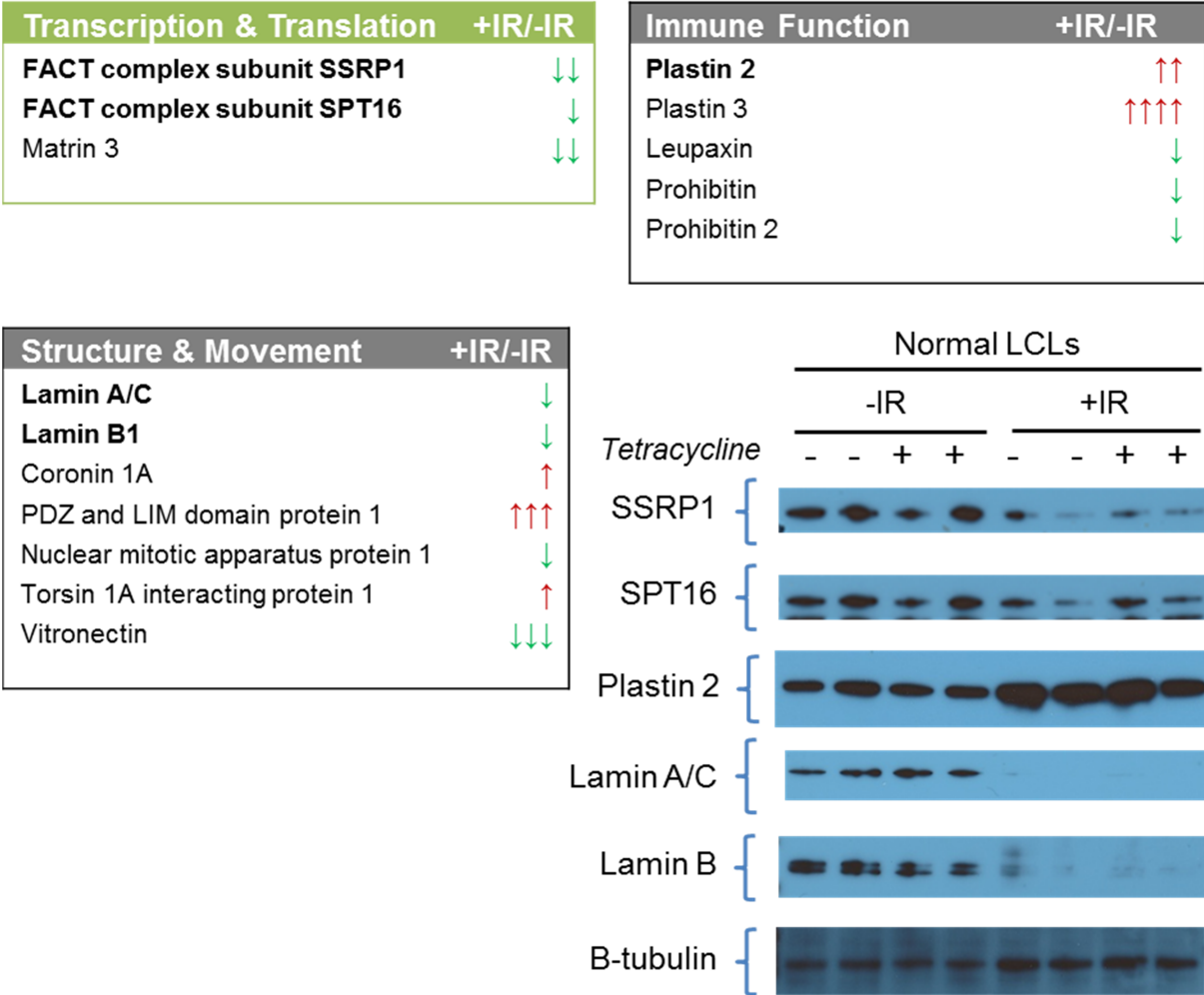


Figure 7-5. Radiation modulated proteins. Proteins differentially regulated in irradiated cells, with arrows indicating the magnitude of change 1 arrow (+/- 0.3 to +/- 0.9), 2 arrows (+/- 1.0 to +/-1.5) , 3 arrows (+/- 1.5 to +/-2.0), or 4 arrows (< -2.0 or > 0.2). Western blot validation of expression changes of selected proteins (lower right) in response to radiation exposure.

RESULTS AND DISCUSSION: LATENT EFFECTS

Flasks of ATM-deficient LCLs and Normal LCLs were irradiated with 5 Gy, or not irradiated. Cells were harvested from the control and irradiated conditions after 24 hours. These cells were processed by eFASP, and the peptides separated by ERLIC (Chapter 8). Data were merged and search using PLGS, and compared, head to head, in Expression analysis.

The threshold for upregulation is a natural log ratio of 0.45, and a p-value of upregulation over 0.95, for the comparison of N24+IR to N24-IR, and AT24+IR to AT24-IR. Proteins flagged as unique in a condition must have been identified with a PLGS score over 100. In most cases, unique protein identifications had PLGS score well over 1,000. 2,393 unique proteins were identified and quantified in both the N24-IR and the N24+IR samples. 238 proteins were upregulated in the N24+IR sample, and 418 proteins were unique to N24+IR. 2,435 unique proteins were identified and quantified in both the AT24-IR and the AT24+IR samples. 143 proteins were upregulated in the AT24+IR sample, and 473 proteins were unique to AT24+IR.

Latent effects of radiation on ATM-deficient LCLs by ERLIC-LC-MS^E

Cell survival

A number of cancer related pathways are enriched in ATM-deficient LCLs following exposure to radiation. Enriched proteins were involved in melanoma signaling (p-value=0.00295) and breast cancer signaling (p-value=0.00646). Many of these components were only identified in AT24+IR cells, and include major players in cell growth and cell cycle control, including B-RAF, H-RAS, and CDK1.

B-RAF (Q13878, AT24+IR) is a serine-threonine protein kinase which is involved in cells signaling for growth. The V600E mutation in B-Raf allows for unhindered cellular growth in a large number of melanomas ³⁰. The usage of a B-Raf inhibitor, PLX4032, was shown to enhance the radiosensitivity in highly radioresistant melanoma cell lines ³¹.

H-RAS (P01112, AT24+IR) is a GTPase that regulates cell division and has been found to be involved in cell transformation ³². H-Ras plays an early role in the signal transduction initiated by growth factors. Recently H-Ras has been found to mediate radioresistance in cervical cancer cells ³³. It has been found that mutant H-Ras can induce DNA damage repair and this has been associated with its ability to the production of reactive oxygen species ³⁴. This is thought to be related to oncogene induced senescence ³⁵.

CDK1 (P06493, AT24+IR) is a serine/threonine protein kinase whose phosphorylation of its substrates leads to cell cycle progression. CDK1 phosphorylates the cell cycle inhibitor, Emi1, to promote its degradation by SCF ³⁶. The SCF complex is important for the ubiquitination of proteins destined for proteasome degradation, particularly those involved in the cell cycle, and the control of G1/S through G2/M transitions ³⁷. The S-phase kinase-associated protein 1 (SKP1, P63208), which is only identified in the AT24+PBS sample, is integral to the proper functioning of the SCF.

Thrombin signaling is upregulated in OFFGEL analysis (AT24+IR vs. AT24-IR)

Thrombin is a protease that is involved in a multitude of pathophysiological processes, including blood clotting, inflammation, repair processes and tumor

metastasis. In ATM-deficient cells, 12 proteins involved in thrombin signaling are upregulated, with a total pathway upregulation p-value of $9.77e^{-7}$. Some of the major proteins and pathways that are upregulated in ATM-deficient LCLs after radiation exposure are CaMKII (Calcium/calmodulin-dependent protein kinase II α , β , δ , γ) and SAPK/JNK.

CAMKII is necessary for initiating the S phase checkpoint after radiation exposure, and its inhibition confers cellular radiosensitivity^{38 39}. The halting of DNA synthesis by CAMKII, via a non-ATM-dependent mechanism, is a possible means for ATM-deficient cells to elicit control of the cell cycle without ATM signaling. Halts DNA synthesis via non-ATM-dependent mechanism.

Thrombin induces phosphorylation of SAPK/JNK, which mediates effects on proliferation, apoptosis, motility, and DNA repair⁴⁰. SAPK/JNK signaling is upregulated following IR exposure (6 proteins, $9.77e^{-7}$) in ATM-deficient LCLs. This pathway has previously been found to be activated by environmental stresses like UV irradiation.

Transcription factors

Based on analysis of all proteins submitted to Ingenuity, a list of predicted transcription factors that could account for these networks was generated. These included MYC, and TP53, which had p-values of overlap of 0.0005 and 0.00552, respectively.

In the absence of functional ATM, cells are still able to induce the expression of proteins that function in survival or cell cycle modulation. The mechanism of these radiation-induced changes is not clear, although, based on transcription factor analysis,

there is an overlap with genes targeted by TP53, which is essential in ATM-dependent radiation responses. Thrombin signaling may also play a part in this alternative mechanism of survival and repair following IR insult. These indicate redundant pathways, that do not require ATM, which can effect responses to ionizing radiation exposure.

Latent effects of radiation on Normal LCLs by ERLIC-LC-MS^E

IL-8 and the bystander effect

The most significantly enriched pathway in normal LCLs, 24 hours have radiation exposure, is IL-8 Signaling (p-value= 7.94×10^{-6}) (Table 7-3). Exposure to gamma-radiation activates ATM, which in turn activated NF-kb. NF-kb induces expression of numerous cytokines, including IL-8 and IL-6, in addition to a multitude of pro-survival genes. These released factors can interact with nearby cells, and effect changes in expression and signaling. In this manner, nearby cells that were not directly affected by radiation, will manifest some of the biological consequences of radiation exposure⁴¹. In the context of cancer radiotherapy, this can be troubling, since irradiated cells can release anti-apoptotic factors, which can interfere with cancer cell death by apoptosis or necrosis. This strong enrichment of IL-8 signaling components is possibly the initiation of or the result of the bystander effect, and includes Raf and Src kinases, and pro-inflammatory cytokines.

Raf kinases and proteins associated with their activation are upregulated following irradiation. They are part of the Ras/MAPK signaling cascade, and are activated by growth factor stimulation to promote growth and development. A Raf, Raf-1, and B-Raf are strongly activated by growth factor signaling, the latter two are

activated via Ras and Src (ln ratio (N24+IR/N24+PBS) = 0.68, p-value=1.00). The upregulation of a complete interacting module of Raf and Ras proteins highlights their importance to cell survival in the context of direct radiation insult, and in the cellular effects induced by the bystander effect.

Aside from growth signaling, IL-8 can effect the expression of genes involved in inflammation and cellular organization. Interleukin 1 receptor associated kinase 4 (IRAK4) is involved in innate immune response signaling against foreign pathogens, and in the context of radiation maybe be important in preventing opportunistic infection ⁴². LIM and SH3 domain protein 1 (LASP1, ln ratio (N24+IR/N24+PBS) = 0.42, p-value=1.00)) is a p53 target that affects cytoskeletal architecture and organization ⁴³. It is involved I normal development, but has been associated with migration and proliferation of beast tumors and medulloblastoma ^{44,45}.

IL-8 Signaling		
Accession	Protein description	12 proteins, p-val=3e ⁻³
P10398	Proto-oncogene A-Raf	N24+IR unique
P04049	Proto-oncogene c-RAF	N24+IR unique
Q13878	Proto-oncogene B-Raf	N24+IR unique
P01116	GTPase KRas	N24+IR unique
P12931	Proto-oncogene c-Src	Upregulated
Q9NWZ3	Interleukin-1 receptor-associated kinase 4	N24+IR unique
Q14847	LIM and SH3 domain protein 1	Upregulated

Table 7-3. Protein components of IL-8 signaling that are upregulated 24-hours after IR exposure. Ingenuity pathway Analysis identified Integrin signaling as being significantly enriched in the irradiated data set, with 12 proteins clustered, and a p-value of enrichment of 3e⁻³. A number of the proteins involved are presented along with the cellular compartment to which they are associated.

DNA repair and damage response

Following exposure to 5 Gy of IR, normal LCLs upregulate a panel of proteins important for DSB recognition and repair, including components of DNA repair complexes, proteins that function independently for repair or regulation of repair processes.

RAD50 (Q92878, N24+IR) is a component of the MRN complex. The MRN complex is composed of MRE11, NBS1, and RAD50, and is responsible for binding to DNA around the foci of strand break to participate in repair by homologous recombination and non-homologous recombination ⁴⁶.

poly[ADP-ribose] synthase 1 (PARP1, P09874) (ln ratio (N24+IR/N24+PBS) = 0.4, p-value=1.00) is involved in many aspects of proliferation and DNA damage recovery. PARP functions by modifying nuclear proteins, by ADP-ribosylation, to effect single stranded break repair, but it can also function on double strand breaks ^{47,48}. ATM is activated through ADP-ribosylation by PARP1, following IR damage ⁴⁹.

Bifunctional polynucleotide phosphatase kinase (PNKP, Q96T60) (ln ratio (N24+IR/N24+PBS) = 0.42, p-value=0.95) is a protein that contains both phosphatase and kinase activities ⁵⁰. PNKP is involved in single strand break repair, double strand break repair, and base excision repair dependent on 3'-phosphatase activity. In yeast, the loss of PNKP activity renders cells more sensitive to gamma-radiation and DNA damaging drugs ⁵¹.

Latent oxidative stress response

NRF2 oxidative stress response genes are quickly upregulated 4 hours after IR exposure, as well as 24 hours later (p-value=0.0158). While the pathways are upregulated, the major players change, which may be a result of differing antioxidant needs at timepoints farther from the exposure point. Some of the upregulated components include GRase, SOD, PRXN5, PRXN6, and catalase. These components have been either uniquely identified, or identified but below the upregulation threshold of 0.45, indicating a weakening in the NRF2-oxidative stress response after 24 hours.

Glutathione reductase (GRase, P00390, N24+IR) reduces oxidized glutathione to produce the antioxidant reduced form. Superoxide dismutase Mn (P04179, In ratio (N24+IR/N24+PBS) = 0.31) catalyzes the production of oxygen and hydrogen peroxide from the more harmful superoxide. Peroxiredoxin 5 (P30044, In ratio (N24+IR/N24+PBS) = 0.12) and peroxiredoxin 6 (P30041, In ratio (N24+IR/N24+PBS) = 0.22) participate in the reduction of hydrogen peroxide and short chain hydroperoxides to protect against oxidative injury. Catalase (P04040, In ratio (N24+IR/N24+PBS) = 0.18) also participates in the removal of hydrogen peroxide, through breakdown of hydrogen peroxide into water and oxygen.

There is a loss of many antioxidant proteins. There is a striking reduction in GST-Pi expression (P09211) (In ratio (N24+IR/N24+PBS) = -0.27, p-value=0.00), thioredoxin (P10599) (In ratio (N24+IR/N24+PBS) = -0.57, p-value=0.00),

These data indicate a possible wave of antioxidant activity early after radiation, when oxidative stress is more important to reduce. After a certain amount of time,

oxidative stress becomes less important, and the repair of DNA damage is the paramount function for cellular recovery.

Transcription factors

Based on analysis of all proteins submitted to Ingenuity, a list of predicted transcription factors that could account for these networks was generated. These included TP53 and ESR1, which have p-values of overlap of 0.00587 and 0.0104, respectively.

REFERENCES

1. McKinnon, P. J. ATM and ataxia telangiectasia. *EMBO Rep* **5**, 772–776 (2004).
2. Barzilai, A., Rotman, G. & Shiloh, Y. ATM deficiency and oxidative stress: a new dimension of defective response to DNA damage. *DNA Repair* **1**, 3–25 (2002).
3. Barzilai, A. & Yamamoto, K. I. DNA damage responses to oxidative stress. *DNA Repair* **3**, 1109–1115 (2004).
4. Barlow, C. *et al.* Loss of the ataxia-telangiectasia gene product causes oxidative damage in target organs. *Proc Natl Acad Sci U S A* **96**, 9915–9919 (1999).
5. Kamsler, A. *et al.* Increased oxidative stress in ataxia telangiectasia evidenced by alterations in redox state of brains from Atm-deficient mice. *Cancer Res* **61**, 1849–1854 (2001).
6. Chen, P. *et al.* Oxidative stress is responsible for deficient survival and dendritogenesis in purkinje neurons from ataxia-telangiectasia mutated mutant mice. *J Neurosci* **23**, 11453–11460 (2003).
7. Ambrose, M., Goldstine, J. V. & Gatti, R. A. Intrinsic mitochondrial dysfunction in ATM-deficient lymphoblastoid cells. *Hum Mol Genet* **16**, 2154–2164 (2007).
8. Matsuoka, S. *et al.* ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA Damage. *Science* **316**, 1160–1166 (2007).
9. Stokes, M. P. *et al.* Profiling of UV-induced ATM/ATR signaling pathways. *Proc Natl Acad Sci U S A* **104**, 19855–19860 (2007).
10. Morgan, W. F. & Sowa, M. B. Effects of ionizing radiation in nonirradiated cells. *Proc Natl Acad Sci U S A* **102**, 14127–14128 (2005).
11. Geromanos, S. J. *et al.* The detection, correlation, and comparison of peptide precursor and product ions from data independent LC-MS with data dependant LC-MS/MS. *Proteomics* **9**, 1683–1695 (2009).
12. Silva, J. C., Gorenstein, M. V., Li, G. Z., Vissers, J. P. & Geromanos, S. J. Absolute quantification of proteins by LCMS^E: a virtue of parallel MS acquisition. *Mol Cell Proteomics* **5**, 144–156 (2006).
13. Silva, J. C. *et al.* Simultaneous qualitative and quantitative analysis of the *Escherichia coli* proteome: a sweet tale. *Mol Cell Proteomics* **5**, 589–607 (2006).
14. Masuda, T., Tomita, M. & Ishihama, Y. Phase transfer surfactant-aided trypsin digestion for membrane proteome analysis. *J Proteome Res* **7**, 731–740 (2008).

15. Li, G. Z. *et al.* Database searching and accounting of multiplexed precursor and product ion spectra from the data independent analysis of simple and complex peptide mixtures. *Proteomics* **9**, 1696–1719 (2009).
16. Silva, J. C. *et al.* Quantitative proteomic analysis by accurate mass retention time pairs. *Anal Chem* **77**, 2187–2200 (2005).
17. Dimmer, E. C. *et al.* The UniProt-GO Annotation database in 2011. *Nucleic Acids Res* **40**, D565–D570 (2012).
18. Binns, D. *et al.* QuickGO: a web-based tool for Gene Ontology searching. *Bioinformatics* **25**, 3045–3046 (2009).
19. McDonald, J. T. *et al.* Ionizing radiation activates the Nrf2 antioxidant response. *Cancer Res* **70**, 8886–8895 (2010).
20. Cholon, A., Giaccia, A. J., Lewis, A. D., Hickson, I. & Brown, J. M. What role do glutathione S-transferases play in the cellular response to ionizing radiation? *Int J Radiat Oncol Biol Phys* **22**, 759–763 (1992).
21. Cho, Y. W., Park, E. H., Fuchs, J. A. & Lim, C. J. A second stress-inducible glutathione S-transferase gene from *Schizosaccharomyces pombe*. *Biochim Biophys Acta* **1574**, 399–402 (2002).
22. Park, C. C., Zhang, H. J., Yao, E. S., Park, C. J. & Bissell, M. J. B1 integrin inhibition dramatically enhances radiotherapy efficacy in human breast cancer xenografts. *Cancer Res* **68**, 4398–4405 (2008).
23. Lanz, C. *et al.* Radiation metabolomics. 3. biomarker discovery in the urine of gamma-irradiated rats using a simplified metabolomics protocol of gas chromatography-mass spectrometry combined with random forests machine learning algorithm. *Radiat Res* **172**, 198–212 (2009).
24. Miller, J. H. J. *et al.* Profiling mitochondrial proteins in radiation-induced genome-unstable cell lines with persistent oxidative stress by mass spectrometry. *Radiat Res* **169**, 700–706 (2008).
25. Qiu, X. X. *et al.* Transcriptome analysis applied to survival of *Shewanella oneidensis* MR-1 exposed to ionizing radiation. *J Bacteriol* **188**, 1199–1204 (2006).
26. Leach, J. K., Van Tuyle, G., Lin, P. S., Schmidt-Ullrich, R. & Mikkelsen, R. B. Ionizing radiation-induced, mitochondria-dependent generation of reactive oxygen/nitrogen. *Cancer Res* **61**, 3894–3901 (2001).
27. Seymour, C. B. & Mothersill, C. The effect of glycolysis inhibitors on the radiation response of CHO-K1 cells. *Radiat Environ Biophys* **27**, 49–57 (1988).

28. Xu, R. H. *et al.* Inhibition of glycolysis in cancer cells: a novel strategy to overcome drug resistance associated with mitochondrial respiratory defect and hypoxia. *Cancer Res* **65**, 613–621 (2005).
29. Heminger, K., Jain, V., Kadakia, M., Dwarakanath, B. & Berberich, S. J. Altered gene expression induced by ionizing radiation and glycolytic inhibitor 2-deoxy-glucose in a human glioma cell line: implications for radio sensitization. *Cancer Biol Ther* **5**, 815–823 (2006).
30. Flaherty, K. T. *et al.* Inhibition of mutated, activated BRAF in metastatic melanoma. *N Engl J Med* **363**, 809–819 (2010).
31. Sambade, M. J. *et al.* Melanoma cells show a heterogeneous range of sensitivity to ionizing radiation and are radiosensitized by inhibition of B-RAF with PLX-4032. *Radiother Oncol* **98**, 394–399 (2011).
32. Wong-Staal, F., Dalla-Favera, R., Franchini, G., Gelmann, E. P. & Gallo, R. C. Three distinct genes in human DNA related to the transforming genes of mammalian sarcoma retroviruses. *Science* **213**, 226–228 (1981).
33. Huang, E. Y. *et al.* A novel radioresistant mechanism of galectin-1 mediated by H-Ras-dependent pathways in cervical cancer cells. *Cell Death Dis* **3**, e251 (2012).
34. Di Micco, R. *et al.* Oncogene-induced senescence is a DNA damage response triggered by DNA hyper-replication. *Nature* **444**, 638–642 (2006).
35. Mallette, F. A., Gaumont-Leclerc, M.-F. & Ferbeyre, G. The DNA damage signaling pathway is a critical mediator of oncogene-induced senescence. *Genes Dev* **21**, 43–48 (2007).
36. Vodermaier, H. C. APC/C and SCF: controlling each other and the cell cycle. *Curr Biol* **14**, R787–96 (2004).
37. Nakayama, K. I. & Nakayama, K. Regulation of the cell cycle by SCF-type ubiquitin ligases. *Semin Cell Dev Biol* **16**, 323–333 (2005).
38. Mirzayans, R., Famulski, K. S., Enns, L., Fraser, M. & Paterson, M. C. Characterization of the signal transduction pathway mediating gamma ray-induced inhibition of DNA synthesis in human cells: indirect evidence for involvement of calmodulin but not protein kinase C nor p53. *Oncogene* **11**, 1597–1605 (1995).
39. Enns, L., Murray, D. & Mirzayans, R. Effects of the protein kinase inhibitors wortmannin and KN62 on cellular radiosensitivity and radiation-activated S phase and G1/S checkpoints in normal human fibroblasts. *Br J Cancer* **81**, 959–965 (1999).

40. Zhu, W. *et al.* Thrombin induces endothelial arginase through AP-1 activation. *Am J Physiol Cell Physiol* **298**, C952–C960 (2010).
41. Hamada, N., Maeda, M., Otsuka, K. & Tomita, M. Signaling pathways underpinning the manifestations of ionizing radiation-induced bystander effects. *Curr Mol Pharmacol* **4**, 79–95 (2011).
42. Li, S., Strelow, A., Fontana, E. J. & Wesche, H. IRAK-4: a novel member of the IRAK family with the properties of an IRAK-kinase. *Proc Natl Acad Sci U S A* **99**, 5567–5572 (2002).
43. Wang, B., Feng, P., Xiao, Z. & Ren, E. C. LIM and SH3 protein 1 (Lasp1) is a novel p53 transcriptional target involved in hepatocellular carcinoma. *J Hepatol* **50**, 528–537 (2009).
44. Grunewald, T. G. P. & Butt, E. The LIM and SH3 domain protein family: structural proteins or signal transducers or both? *Mol Cancer* **7**, 31 (2008).
45. Traenka, C. *et al.* Role of LIM and SH3 protein 1 (LASP1) in the metastatic dissemination of medulloblastoma. *Cancer Res* **70**, 8003–8014 (2010).
46. Stracker, T. H., Theunissen, J. W., Morales, M. & Petrini, J. H. The Mre11 complex and the metabolism of chromosome breaks: the importance of communicating and holding things together. *DNA Repair* **3**, 845–854 (2004).
47. Helleday, T., Bryant, H. E. & Schultz, N. Poly(ADP-ribose) polymerase (PARP-1) in homologous recombination and as a target for cancer therapy. *Cell Cycle* **4**, 1176–1178 (2005).
48. Thompson, L. H. Recognition, signaling, and repair of DNA double-strand breaks produced by ionizing radiation in mammalian cells: the molecular choreography. *Mutat Res* **751**, 158–246 (2012).
49. Valenzuela, M. T. *et al.* PARP-1 modifies the effectiveness of p53-mediated DNA damage response. *Oncogene* **21**, 1108–1116 (2002).
50. Audebert, M., Salles, B., Weinfeld, M. & Calsou, P. Involvement of polynucleotide kinase in a poly(ADP-ribose) polymerase-1-dependent DNA double-strand breaks rejoining pathway. *J Mol Biol* **356**, 257–265 (2006).
51. Kashkina, E., Qi, T., Weinfeld, M. & Young, D. Polynucleotide kinase-phosphatase, Pnk1, is involved in base excision repair in *Schizosaccharomyces pombe*. *DNA Repair* **8**, 676–683 (2012).

CHAPTER 8

The immune cell responses to the radiomitigator drug, YEL2 and to ATM-deficiency.

ABSTRACT

In this study, we probe the proteomic effects of the potent radiomitigator compound, YEL2, which significantly decreases DNA double strand breaks (DSBs), and increases mouse survival following exposure to lethal doses of ionizing radiation (IR). Exposure to IR can cause wide spread cellular and tissue damage. Ataxia Telangiectasia, Mutated (ATM) is a nuclear protein kinase at the heart of the cellular response to DSBs induced by ionizing radiation (IR). Loss of functional ATM results in neurodegeneration, cancer predisposition and radiosensitivity.

Label-free quantitative proteomics, utilizing Electrostatic Repulsion-Hydrophilic Interaction Chromatography (ERLIC) and LC-MS^E, is employed to study the cellular responses to YEL2 in immune cells with functional ATM (N) or mutated ATM (AT). This analysis provides the first catalogue of radiomitigator-modulated proteins and pathways, and the cellular effects of chronic deficiency in ATM activity. These data will be instrumental to the investigation of radioprotective pathways, to the development of radiomitigator drugs for clinical and military applications, and to the understanding of ATM-deficiency on a proteome-wide level.

TABLE OF CONTENTS

INTRODUCTION.....	218
BACKGROUND	220
Radioprotectors: the importance of post-exposure therapeutics.....	220
Ataxia telangiectasia.....	221
Oxidative stress and intrinsic mitochondrial dysfunction.....	223
Proposed treatments for A-T	224
EXPERIMENTAL.....	225
Strains and cell growth	225
Enhanced filter aided sample preparation of LCLs	225
ERLIC fractionation of eFASP peptides.....	227
Analysis by nanoLC-MS ^E	228
Data processing and protein identification	229
RESULTS AND DISCUSSION.....	229
The major proteomic effects of YEL2 on DNA damage repair.....	234
Upregulation of ATM and components of the MRN Complex	234
Master signaling by ATM	235
ATM Activation, the MRN complex, and the cell cycle in the context of IR damage	235
Direct DNA repair enzymes	237
Structural maintenance of chromosomes and homologous recombination.....	238
Cyclin dependent kinases.....	240
Rho family GTPases enriched with YEL2 treatment.....	242
YEL2 in the absence of functional ATM.....	245
YEL2 Transcription factors in both Normal and ATM-deficient LCLs.....	245
Proteomic disturbances due to chronic lack of functional ATM	246
Proteomic consequences of ATM deficiency	246
Pathways upregulated in the presence of functional ATM	247
Previous work on ATM-dependent proteome effects	252
Lack of proteome depth yields dimensionless conclusions.....	253
Data deficiency results in inappropriate assumptions	254
REFERENCES.....	256

INTRODUCTION

Since World War II, there has been great interest in developing agents to protect against the deleterious effects of IR exposure. Thus far, amifostine, is the only drug approved by the FDA for radioprotection during radiotherapy for head and neck cancer. It functions as an antioxidant that selectively protects non-tumor tissues. Its major benefit is the prevention of xerostomia, a condition caused by the destruction of the salivary glands, which can severely decrease a patient's quality of life. When dosed immediately before radiation exposure, it has modest efficacy; however, the doses for efficacy can be toxic, and do not have applicability to radiation emergencies in which exposure is unanticipated.

A number of drugs have been discovered through high throughput screening of drug libraries which show potential as radioprotectors or radiomitigators, which reduce double strand DNA breaks, and increase mouse survival tremendously after lethal doses of radiation exposure. The first class of drugs showing promise is the tetracyclines and ciprofloxacin. Tetracycline was shown to protect cells before and after radiation exposure, in human cells, and to enhance mouse survival following lethal doses of radiation ¹. In its capacity as a radioprotector, tetracycline was shown not to function as an antioxidant, or as an antimicrobial, which lead to the hypothesis that it may be activating upstream regulators/effectors of ATM. The second compound is YEL2, from Robert Schiestl's lab at UCLA. In mice, it has been shown to increase survival at day 30, by 80%, and to reduce DNA damage by 50%, and to reduce incidence of leukemia ². The mechanism of YEL2 is still unknown.

Ataxia telangiectasia (A-T) is caused by mutations in the ATM gene, and is a paradigm for cancer and neurodegenerative disorders. A-T will be utilized as a model for DNA damage and repair in order to elucidate the molecular processes fundamental to radiosensitivity, cellular protection from oxidative stress, and the dependence of YEL2 activity on ATM interaction.

The main goal of this work is twofold: (i) to describe the proteins and networks that are affected by treatment with the YEL2, and (ii) to detail the proteomic consequences of ATM deficiency.

To achieve these aims, we employ label-free quantitative mass spectrometry and comparative proteomics to analyze lymphoblastoid cell lines, with normal and mutated ATM, grown with control or radiomitigator. Ingenuity Pathway Analysis (IPA) groups differentially regulated proteins by related by pathway and allows for a higher-level analysis of the data. We provide the first description of the potent radiomitigator, YEL2, and a comprehensive account of proteome changes due to ATM-deficiency. This work will be of interest for clinical and military applications aimed at developing radiomitigators and treatments for ATM-deficiency.

Normal LCLs and radiosensitive ATM-deficient LCLs were incubated with PBS, tetracycline, or YEL2, for 2 hours, and then the cells were harvested and processed by eFASP with 0.2% DCA. Peptides were fractionated by Electrostatic Repulsion and Hydrophilic Interaction Chromatography (ERLIC), and then each fraction was analyzed by LC-MS^E. Approximately 1,400 proteins were identified in each technical replicate, based on the identification of around 36,000 peptides. 754 proteins were quantified

across all Normal LCL sample conditions, and 828 proteins were quantified across all AT LCL sample conditions in at least 1 technical replicates per biological sample.

BACKGROUND

Radioprotectors: the importance of post-exposure therapeutics

Studies on ATM and its scope of effects may lead to the development of effective radioprotector or radiomitigator compounds. Since the mid-1950s there has been great interest in the development of therapeutic agents for the protection against the deleterious effects of IR. IR has practical uses in many fields including nuclear medicine and scientific research, but poses health hazards if improperly used.

Cancer therapeutic agents destroy rapidly dividing cells, a characteristic of cancer but also of stem and bone marrow cells. Amifostine is the only drug that is FDA approved for the purpose of preventing IR-induced damage by preferentially protecting normal tissues. In patients undergoing radiation therapy for head and neck cancer, amifostine is beneficial in mitigating xerostomia, a life altering side-effect in which a patient's salivary glands are damaged^{3,4}. Amifostine must be administered almost immediately before radiotherapy in order to have a preventative effect, and is ineffective in dealing with unanticipated exposure scenarios.

People may also be exposed to IR during industrial accidents and dirty bomb explosions. In these cases, the inability to predict exposure prevents prophylactic drug therapies; therefore, the need to find a drug that not only counters the deleterious effects of IR, but also shows efficacy when administered post-exposure, is greatly important. Collaborating laboratories involved in the discovery and characterization of

putative radioprotector and radiomitigator compounds have developed high-throughput screening assays to identify candidates (McBride and Schiestl, UCLA). Such compounds may prove to be integral in preventing side effects of radiotherapy and nuclear medicine as well as treatment of patients exposed to IR industrially or on the battlefields.

Ataxia telangiectasia

Ataxia telangiectasia (A-T) is an autosomal recessive neurodegenerative disease that commences in early childhood. The first signs of the disease include uncoordinated movements (ataxia), dilated blood vessels (telangiectasias) in the corners of the eyes or the surface of the cheeks, slurred speech, and delayed development of motor skills ^{5,6}. Estimates on worldwide incidence range from 1:40,000, to 1:1,000,000 ^{6,7}.

As the disease progresses further detrimental conditions become apparent, including immunodeficiency, predisposition to cancer (hematopoietic malignancies), progressive cerebellar neurodegeneration and sterility ⁶. Patients may live into their 20s and 30s, although severely handicapped ⁶. On a molecular level, cells in patients with A-T have decreased lifespan, defective cell cycle checkpoint activation, reduced capacity to repair double stranded DNA double strand breaks (DSBs), defective cytoskeletons, and greater sensitivity to ionizing radiation ^{6,8}. This array of effects makes A-T a perplexing disease and links diverse pathologies which may give insight into fields of study including cancer, radiation protection, and neurodegeneration.

Ataxia Telangiectasia was first described by Syllaba and Henner in 1926 and was established as a disease 30 years later by Broder and Sedgewick^{9,10}. Research over the next 60 years helped to define the genetic, clinical, and molecular aspects A-T. Landmark findings include: the discovery that A-T patients have adverse responses to radiotherapy¹¹, cellular hypersensitivity to X- and gamma rays¹², cancer predisposition¹³, mapping of the A-T gene to chromosome C11q22-23¹⁴, and the identification of the ATM (Ataxia Telangiectasia, Mutated) gene that gives rise to A-T when mutated⁶.

The quality of A-T research and discovery was enhanced when Savitsky et al. discovered that the mutation in the ATM gene was the underlying cause of A-T^{6,7}. ATM spans approximately 150kb of genomic DNA; it contains 66 exons and a 9,168-bp coding region^{6,15,16}. With over 500 different mutations identified in the coding region for the ATM gene, patients are generally compound heterozygotes, inheriting a different mutation from each parent. Thus, there is much allelic diversity and consequent heterogeneity in clinical presentation and severity^{7,17}. Most ATM gene mutations give rise to truncated ATM protein resulting from nonsense or splicing mutations, or secondary premature terminations due to frameshift mutations⁷.

In milder forms of A-T, immunodeficiency or cancer susceptibility may be absent, which may be due to the production of either decreased numbers of ATM protein or impaired ATM activity. There are suggestions that in individuals carrying only one mutated allele for the ATM gene, there may be an increased risk for breast cancer as well as other cancers⁷.

Oxidative stress and intrinsic mitochondrial dysfunction

Ionizing radiation (IR) causes DNA breakage and cellular damage in two ways. First, IR directly transfers energy to DNA causing breaks and second, IR causes the production of reactive oxygen species (ROS). While the prevention of the direct effects of IR may be implausible, there is interest in enhancing the protection from free radical damage and increasing cellular antioxidants to soak up excess free radicals.

Oxidative stress results in macromolecular damage resulting in genome instability, cancer and neurodegeneration ⁷. IR passing through tissues interacts with water to generate ROS, including hydroxyl radicals, hydrogen radicals and hydrogen peroxide, which interact with and damage DNA, proteins and membranes ¹⁶. The endogenous defenses against free radicals are insufficient to handle the amount of ROS resulting from IR exposure; therefore, the degree of deleterious cellular effects from IR depends on the dose and the sensitivity of the organ.

The absence of functional ATM has been shown to result in a buildup of cellular oxidative stress, independent of IR exposure, in A-T patients as well as ATM deficient cell lines and animals ¹⁸⁻²². The cause of this buildup is not entirely clear; however, it is thought that chromosomal breaks may lead to oxidative stress indirectly through altered gene expression or directly through disruption of mitochondrial function and oxygen metabolism ¹⁸. Additionally, it is hypothesized that ATM may have a role in the regulation of cellular redox homeostasis since cells from A-T patients are sensitive to agents that induce ROS generation and oxidative DNA damage, and cells from healthy patients when treated with siRNA against ATM show decreases in cellular respiration

until normal ATM is introduced and increases respiration ²³. The roles of extra-nuclear ATM are largely unexplored as well as its role in mitochondrial respiration ²³.

Proposed treatments for A-T

While no curative approaches for A-T currently exist there are a number of treatments aimed at ameliorating manifestations of A-T. Therapies are aimed at treating tumors with radiotherapy while minimizing tissue damage, countering immunodeficiency and infection through the usage of immunoglobulins, and reducing neurological problems (loss of balance and tremors) using dopaminergic agents, such as L-DOPA ⁷.

A number of experimental treatments for A-T have been proposed to hinder the progression of neurodegeneration through the reduction of cellular oxidative stress. CTMIO (5-carboxy-1,1,3,3-tetramethylisoindolin-2-yloxy), a nitroxide antioxidant, acts similarly to superoxide dismutase and was shown in a mouse model of A-T to correct the neurobehavioral phenotype and diminish tumor progression ²⁴. The iron chelator EGCG (epigallocatechin gallate) has been shown to improve genomic stability in ATM deficient mice ²⁵.

Geneticin, and other aminoglycoside antibiotics, have been shown to induce read-through of premature termination codons (PTC), restore the levels of full-length ATM protein and return ATM functions to A-T cells in culture ²⁶; however, this is only effective in about 14% of the population of A-T patients who have mutations with single-nucleotide changes resulting in primary PTCs.

The need for effective, mutation non-specific, drug therapies targeted towards rescuing fully functional ATM protein production and abating the destructive effects of chronically high levels of oxidative stress is apparent. Drugs found to be beneficial for A-T may also have efficacy in health research with relation to DNA damage and oxidative stress.

EXPERIMENTAL

Strains and cell growth

LEPU LCLs and NAT8 LCLs were grown in RPMI-1640 with 10% FBS and 1% PSG (pen/strep/glutamine) to a density of 10^7 cells (approximately 1mg of total protein). Duplicate flasks for each condition were utilized. Specified flasks were incubated with PBS, 1000 μ M YEL2, or 33.7 mM tetracycline, for 2 hours. Specified flasks were then exposed to 5 Gy ionizing radiation, and all cells were harvested 24 hours later.

Harvested cells were washed with cold 5% dextrose solution. Laemmli buffer (2% SDS, 0.1% deoxycholic acid (DCA), 5 mM TCEP, 100 mM ABC, 1:100 protease inhibitor, 1:100 phosphatase inhibitor) was added to all cell pellets, and the cells were lysed by ultrasonication and snap-freezing in liquid nitrogen. Protein concentration was determined with the reducing agent compatible BCA assay kit (Pierce, PI23250). 250ug of each sample was processed by eFASP.

Enhanced filter aided sample preparation of LCLs

Aliquots of each sample (250 μ g) were brought to 4% SDS, 0.2% DCA, and 50 mM TCEP. Glycogen phosphorylase B (rabbit muscle, Protea, PS-126-1) was added to

each sample at a 1:100 ratio. Each sample was sonicated for three 10-second intervals with a Sonic Dismembrator Model 100 (Fisher Scientific), and then spun at 16,000 rcf for 10 minutes in an Eppendorf centrifuge; this step was repeated once. The supernatant and any residual pellet, was sonicated and allowed to cool to 37°C. Proteins were alkylated by the addition of 4-vinylpyridine (4-VP) at 30 mM final concentration, and incubated for 1 hour on a shaker at 37°C. DTT was added to a final concentration of 50 mM to deactivate residual 4-VP. Based on a 20 mg:1 mg wet weight:protein estimate, 250 µg of each sample was transferred to a new tube for further processing.

Eight volumes of Solubilization Buffer B (8M urea, 0.2% deoxycholic acid, 100 mM Tris-HCL, pH 8) was added to the digest to begin buffer exchange of SDS. The diluted samples were applied to a Passivated Microcon Filter Unit (described below), and spun at 14,000 rcf for 10 minutes. The eluate was discarded, and 200 µl of Solubilization Buffer B was added to the filter unit and then spun at 14,000 rcf for 10 minutes. This step was repeated twice. To remove urea, three buffer exchange steps were performed with Digestion Buffer (0.2% deoxycholic acid in 50 mM ABC). The Passivated Microcon Filter Unit (30,000 Da cutoff), and collection tube, were pre-incubated in 5% Tween-20 overnight, and then thoroughly rinsed with large volumes of MS-grade water.

Digestion Buffer (100 µl) was added to the Microcon unit, followed by the addition of 5 µg of trypsin (1 µg trypsin:50 µg protein). Digestion proceeded for 12 hours on a shaker set to 37°C. The Microcon unit was transferred to a new Passivated Collection Tube, and spun at 14,000 rcf for 10 minutes. To assure complete removal of peptides,

100 µl of 50 mM ABC buffer was added to the Microcon unit, and collected by centrifugation, two times.

Removal of deoxycholic acid from the digest was accomplished by acidification and phase transfer of insoluble deoxycholate to an ethyl acetate layer ²⁷. Instead of removing the aqueous layer as in Masuda et al. ²⁷, the organic layer was removed, and then two more rounds of addition and removal of ethyl acetate facilitate the removal of deoxycholate from the solution, similar to the strategy Yeung et al. employ for the removal of n-octylglucoside ²⁸. Collected peptides were dried down in a SpeedVac and re-suspended in 50% methanol, three times.

ERLIC fractionation of eFASP peptides

Dried peptide samples were suspended in a small volume of 0.1% acetic acid, to which acetonitrile and Mobile Phase A (90% ACN, 0.1% acetic acid) were added to a final concentration of 85% ACN. Peptides were fractionated using a PolyWAX LP anion-exchange (200 x 2.1 mm, 5-µm, 300-Å, PolyLC, Columbia, MD (202WX0503)), on a Shimadzu Prominence HPLC. 87 fractions were collected with a 140 minute gradient of 100% Mobile Phase A for 10 minutes, 0-12% Mobile Phase B (30% ACN, 0.2% formic acid) for 10 minutes, 12-25% Mobile Phase B for 37 minutes, 25-40% Mobile Phase B for 13 minutes, 40-47% Mobile Phase B for 18 minutes, 47-100% Mobile Phase B for 22 minutes, 100% Mobile Phase B for 20 minutes, and then 0% Mobile Phase B for 30 minutes. The flow rate was 0.25 ml/min, with 400 µl fractions collected into a 96-well plate, every 1.6 minutes. Fractions were pooled into 25 fractions based on abundance from UV-Vis absorbance data collected at 215nm.

Analysis by nanoLC-MS^E

nanoUPLC of sample peptides was performed on an Waters nanoACQUITY UPLC system equipped with a nanoACQUITY UPLC Symmetry C18 180 μm x 20 mm trap column (Waters, 186003514), and a nanoACQUITY UPLC BEH C18 75 μm x 150 mm reversed phase analytical column (Waters, 186003543). The aqueous mobile phase (Mobile Phase A) contained 0.1% formic acid in water and the organic mobile phase (Mobile Phase B) contained 0.1% formic acid in acetonitrile. 5 μl of the sample was loaded onto the trap with and washed with 99% A and 1% B for 5 minutes, and then injected onto the column with 3% Mobile Phase B. Peptides were eluted from the column with a gradient of 3-40% B over 90 minutes at 300 nl/min, followed by a 5 minute rinse with 95% B. The column was re-equilibrated to 97% A for 20 minutes.

The lock mass compound, [Glu¹]fibrinopeptide (0.5 pmol/ μl , 25% ACN, 0.1% formic acid), and the analyte were delivered at 300 nl/min to the NanoLockSpray dual electrospray ion source (Waters, UK) which was equipped with a microcapillary for the lockmass compound, and a PicoTip Emitter (New Objective) for the analyte. Two electrosprays are produced which may be selectively sampled by positioning of the baffle.

The ion source was interfaced with a Waters Synapt HDMS for mass spectrometric analysis. Acquisition was performed in positive nanoESI mode and the instrument was operated in V-mode with an average resolution of 9500. Accurate mass data for precursor and fragment ions were collected over the 100-2000 m/z range, with the mode of acquisition alternating between low and elevated energy at 1-second

intervals, for the duration of analysis. The lock mass channel was sampled every 30 seconds for calibration of MS and MS^E data.

Data processing and protein identification

Raw data files were processed in ProteinLynx Global Server 2.5 (PLGS, Waters) to provide an inventory of precursor ions along with their respective fragment ions, to search against the Uniprot Swissprot Human database (April 2011) for peptide and protein identifications²⁹. Database search settings included a minimum of 3 peptides per protein identification, two missed cleavages, a 4% false positive rate, fixed S-pyridylethyl modification of cysteine residues, and variable oxidation modification of methionine residues. The Waters Expression Informatics component of the PLGS software was utilized to quantify data using glycogen phosphorylase B (Uniprot accession P00489) as the quantification reference³⁰⁻³².

Data Extraction and Correction (DECO) software was used to correct merged ion volumes for peptides identified in multiple fractions, and to extract all quantification data for heatmap generation. Uniprot-GOA³³ was searched in the QuickGO! browser to obtain compartment and function annotations for all proteins³⁴. Data sets were further analyzed through the use of IPA (Ingenuity® Systems, www.ingenuity.com) for functional and pathway information.

RESULTS AND DISCUSSION

Flasks of ATM-deficient LCLs and Normal LCLs were incubated with PBS, tetracycline, or YEL2 for two hours. Specified flasks were exposed to 5 Gy γ -irradiation,

and then cells from all experimental flasks were harvested after 24 hours. Cells were lysed and proteins extracted, solubilized, and digested by eFASP with 0.2% DCA. The resulting peptides from eFASP were separated based on isoelectric point and hydrophilicity using ERLIC on a PolyWAX LP column. Fractions were pooled based on the 215 nm UV-Vis absorbance trace, yielding 25 to 30 fractions. Each fraction was injected onto a reverse phase column and separated by UPLC on a nanoACQUITY. Eluting peptides were analyzed by LC-MS^E on a Synapt HDMS. Raw data files produced for all fractions of a sample were merged and searched using PLGS. DECO was utilized to correct peptide ion volume data for any peptide found in multiple ERLIC fractions.

ERLIC-LC-MS^E enabled deeper proteome coverage than previously obtained with OFFGEL-LC-MS^E. 939 proteins were quantified in both cell lines, under all conditions, and in all technical replicates (Figure 8-1). 3,484 unique proteins were quantified in at least one technical replicate analysis. 1,080 proteins were quantified in at least one technical replicate of all Normal LCL samples (Figure 8-2). 1,314 proteins were quantified in at least one technical replicate of all three AT LCL samples (Figure 8-3).

For comparative analysis, each processed data set was compared in a head-to-head fashion using Expression Analysis for normalization, ratio determination, and statistical analysis of expression differences. All ratios for N LCLs or AT LCLs were determined with protein abundance for the IR, the TET, or the YEL2 condition as the numerator, and the protein abundance for the PBS control as the denominator. The threshold for protein upregulation was a natural log ratio greater than 0.45, with a p-

value of upregulation greater than 0.95 (Table 8-1). Proteins flagged as unique in a condition must have been identified with a PLGS score over 100. In most cases, unique protein identifications had PLGS score well over 1,000. For comparisons between each condition in the N24 LCLs and the AT24 LCLs, the ratios were determined with the protein abundance from the AT24 LCL as the denominator (Table 8-2). The threshold for upregulation was the same, and the threshold for downregulation was a natural log ratio less than -0.45, with a p-value of downregulation less than 0.05.

Ingenuity pathway analysis of the differentially expressed proteins from the head-to-head comparisons allows for the determination of wide proteomic effects of IR, TET, and YEL2, and their dependence on ATM. Additionally, we can compare the cell lines without experimental conditions, to explore the proteomic consequences of ATM-deficiency.

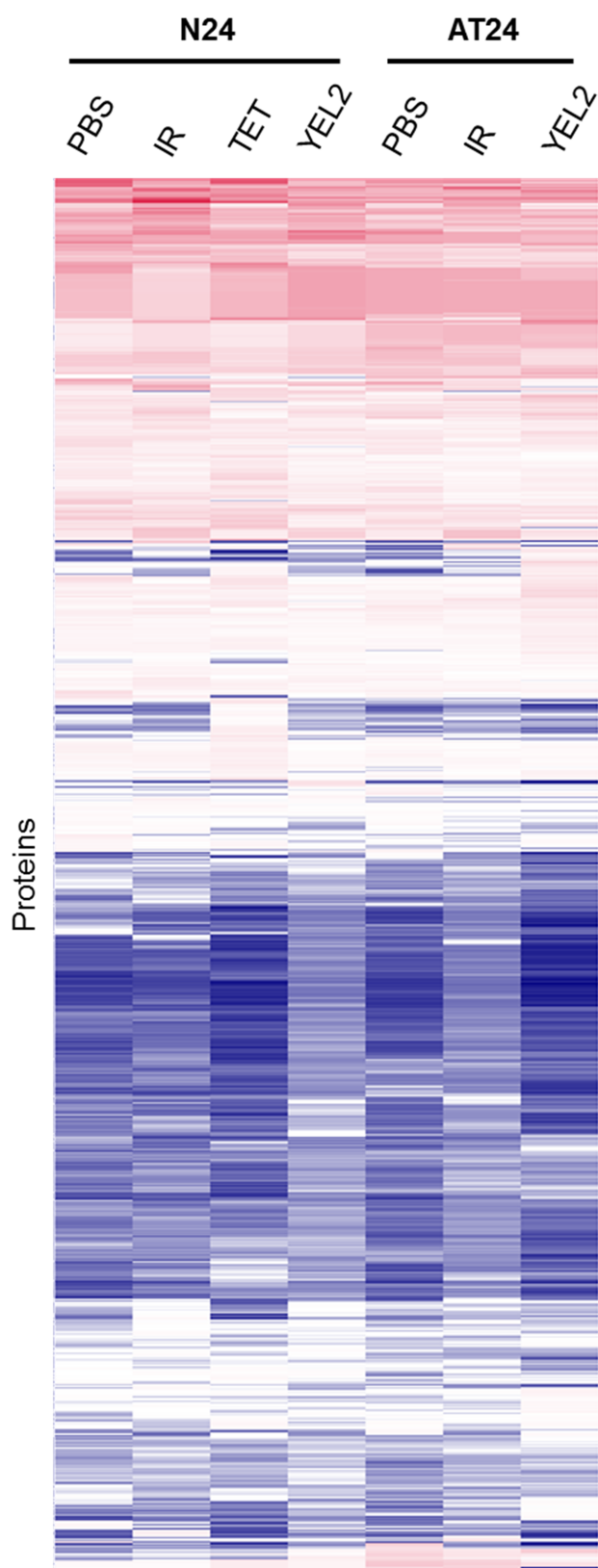


Figure 8-1. 939 proteins were quantified across all sample conditions, and all technical replicates, for both N and AT LCLs. 3,484 proteins were quantified in at least one sample replicate analysis. The heatmap represents the average protein abundance calculated from the acquired technical replicates for each sample. Each row represents the intensity for a specific protein across the conditions. The data were normalized and clustered by average linkage distance. Crimson indicates high protein abundance, while navy represents low protein abundance.

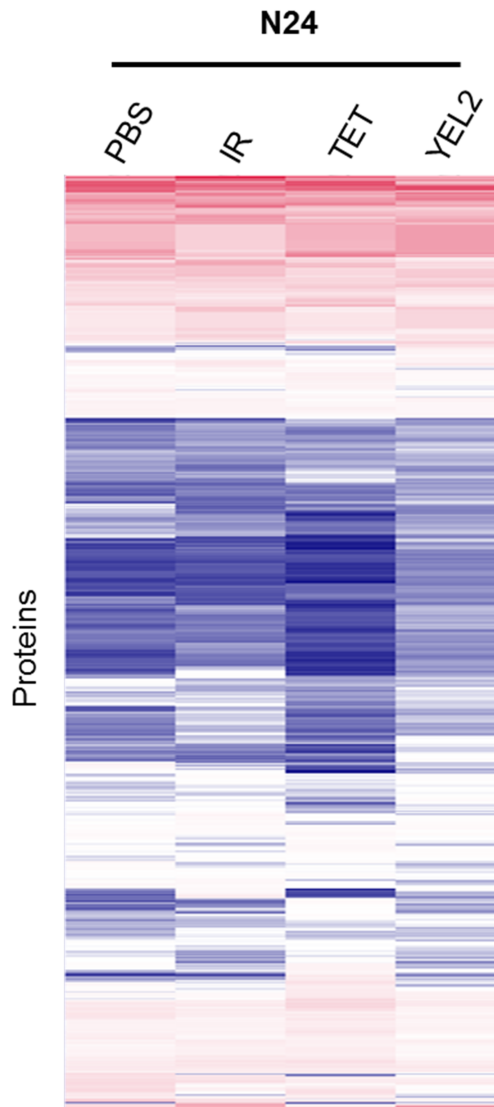


Figure 8-2. Protein expression levels in Normal LCLs, 24 hours after treatment or after 5Gy IR. This heatmap represents 1,080 proteins that were quantified in one technical replicate of all 4 samples. Each row represents the intensity for a specific protein across the conditions. The data were normalized and clustered by average linkage distance. Quantitative information for each protein was determined by the average protein intensity in 3 technical replicates. Crimson indicates high protein abundance, while navy represents low protein abundance in a sample.

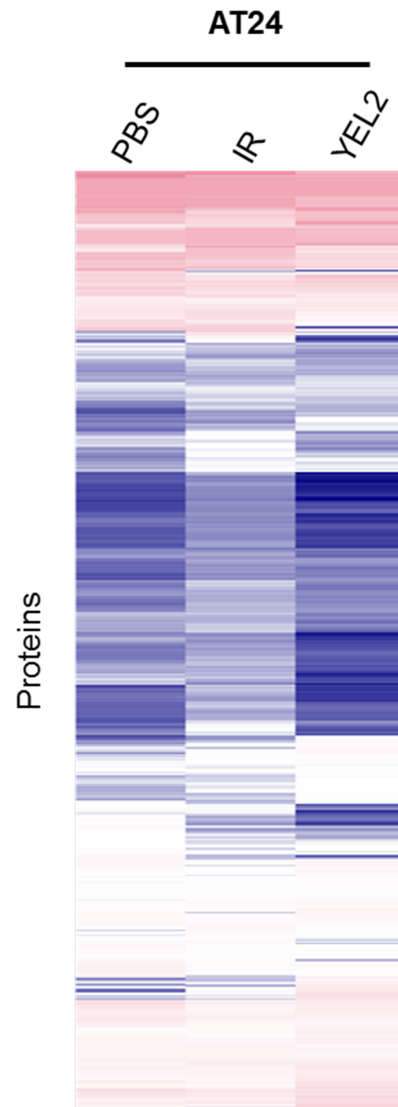


Figure 8-3. Protein expression levels in ATM-deficient LCLs, 24 hours after treatment or after 5Gy IR. This heatmap represents 1,314 proteins that were quantified in one technical replicate of all 3 samples. Each row represents the intensity for a specific protein across the conditions. The data were normalized and clustered by average linkage distance. Quantitative information for each protein was determined by the average protein intensity in 2 technical replicates. Crimson indicates high protein abundance, while navy represents low protein abundance in a sample.

	N24			AT24	
	IR / PBS	TET / PBS	YEL2 / PBS	IR / PBS	YEL2 / PBS
Upregulated	238	96	244	143	307
Unique	418	344	341	473	470

Table 8-1. Summary of differentially regulated proteins in each condition, for each cell line. Expression analysis enabled the head-to-head comparisons between treated and control in each cell line. The threshold for protein upregulation was a natural log ratio greater than 0.45, with a p-value of upregulation greater than 0.95. Unique identifications must have a PLGS score greater than 100.

	N24 / AT24		
	PBS	IR	YEL2
Upregulated	307	198	275
Downregulated	308	170	228
Unique N24	399	474	378
Unique AT24	336	508	665

Table 8-2. Summary of differentially regulated proteins in each condition, between the cell lines. Expression analysis enabled the head-to-head comparison of protein abundance between cell lines, for each condition. The threshold for protein upregulation was a natural log ratio greater than 0.45, with a p-value of upregulation greater than 0.95. The threshold for protein upregulation was a natural log ratio less than -0.45, with a p-value of downregulation less than 0.05. Proteins unique to either cell line must have a PLGS score greater than 100.

The major proteomic effects of YEL2 on DNA damage repair

Upregulation of ATM and components of the MRN Complex

After YEL2 treatment there is enrichment in components of ATM Signaling (p-value=5.75E-05), including components of DSB repair by homologous recombination (p-value=7.08E-04) and non-homologous recombination (p-value=4.47E-05) (Table 8-3). Among the most important are ATM (N24+YEL2, PLGS score 604), MRE11 (ln ratio

(N24+PBS/N24+YEL2) = 0.42, p-value =1.00), NBN (N24+YEL2, PLGS score 678), and RAD50 (N24+YEL2, PLGS score 713). In the variety of experiments we have done, outside of this work, these components are rarely identified due to low copy number, especially ATM.

Master signaling by ATM

ATM is a 350 kDa nuclear protein kinase at the core of the cellular response to double-stranded DNA breakage. ATM is responsible for sensing double-stranded DNA damage, due to genotoxic agents (ionizing radiation, anticancer therapeutics, etc.) or normal physiological processes (occurring during meiosis, immune system maturation and telomere maintenance), and initiating an appropriate cellular response. Signaling to pathways involved in cell cycle control, DNA repair and apoptosis are essential for ATM to exert the appropriate cellular effects.

ATM exists as an inactive dimer with three sites for autophosphorylation integral to its functioning and activity: S1981 (critical for recruitment and activation), S367 and S1893³⁵. A multitude of kinases and phosphatases are involved in the delicate balance of ATM activity⁸. The details of ATM activation, the extranuclear role of ATM and the function of other ATM post translational modifications remain elusive³⁵.

ATM Activation, the MRN complex, and the cell cycle in the context of IR damage

DNA DSBs are deleterious and potentially lethal to the cell and must be repaired to prevent genetic damage and to assure that damaged DNA is not passed on to daughter cells. Repair of DNA DSBs occurs through either non-homologous end joining (NHEJ), which joins the ends of a DNA break without the need for a homologous

sequence, or homologous recombination (HR), which requires a homologous sequence to guide repair ⁸. Myriad pathways and effector molecules are involved in the proper signaling involved in the cellular response to DNA DSB damage following ionizing radiation, termed the DNA damage response (DDR) ⁸. ATM plays a central role in the DDR, and it activates (via direct or indirect phosphorylation) many substrates and effector molecules involved in signaling to DNA repair machinery and cell cycle checkpoints; therefore, cells deficient in functional ATM lack proper DDR signaling.

Recognition and repair of DNA DSBs involves many proteins with redundant functions. The Mre11 complex senses double strand breaks (primary sensor), migrates to the foci of the DNA lesion and tethers the broken ends of the DNA together to prepare for repair. Mre11 recruits ATM through multiple contacts, and activates ATM, which in turn phosphorylates members of the Mre11-Rad50-Nbs1 (MRN) complex as well as many (30 known) substrates that signal different processes involved in DDR ^{8,35}. Interestingly, ATM and the MRN complex are dependent on one another for recognition and signaling, and subsequently the MRN complex binds ATM and activates it, and in turn becomes a substrate and an adaptor for ATM-signaling ³⁵.

The activity and activation of ATM is not solely influenced by DNA DSBs, but also by modifications to chromatin structure and DNA conformation, which explains the presence of ATM adjacent to sites of DNA damage ^{8,35}. At sites of chromatin modification it is thought that ATM is partially activated and is only capable of activating p53, thus requiring MRN for full activity ³⁵. Breast cancer-1, early onset (BRCA1) is a substrate of ATM and is involved in the unfolding of chromatin. This leads to the “DNA-

repair-restore hypothesis” in which disrupted chromatin structure is required before DNA damage is recognized and DNA repair is initiated ⁸.

ATM regulates progression through the cell cycle at many steps in response to DNA damage. During the G1/S transition, ATM affects the stability of p53, which is responsible for selecting whether to initiate apoptosis or begin DNA repair ⁸. If the amount of DNA damage is too large, then p53 initiates apoptosis otherwise; p53 signals for DNA repair and synthesis, and blocks the transition from G1 to S phase ⁸. ATM activation directly affects two pathways during S phase: hindering the initiation of DNA synthesis and controlling processes involved in replication forks and DNA repair ⁸. In the G2/M transition ATM activates Chk2 which hinders the cell from progressing to mitosis ⁸. Questions remain regarding the differential activity of ATM during different cell phases, the activation of ATM in post-mitosis and post-meiosis, as well as the role of activated ATM in cells devoid of external DNA damage.

Direct DNA repair enzymes

YEL2 treatment induces an increase in the abundance of proteins that directly affect DNA damage repair via detection, excision repair and ligation. These proteins include APEX1, DDB1, and XRCC4.

APEX nuclease 1 (APEX1) is a multifunctional enzyme that functions in the base excision repair of DNA lesions induced by oxidative damage. After a lesion is identified, DNA glycosylases remove the base. APEX1 recognizes apurinic/apyrimidic sites, and catalyzes the cleavage of the phosphodiester backbone adjacent to the damage. The created single strand break can then be repaired by creating a single strand break

which can be repaired by long path or short patch BER ³⁶. Expression of APEX1 has been associated with chemoresistance in human osteosarcoma, and was found to be a predictor of recurrence or metastasis ³⁷. Moreover, a study on APEX1 and radioresistance showed that increasing APEX1 levels via transfection, was directly correlated with increased radioresistance ³⁸.

Damage-specific DNA binding protein 1 (DDB1) is a component of complexes that function in genome repair and in the transcriptional control of DNA repair proteins. One of the major functions of DDB1 is to recognize UV-induced DNA damage, and to initiate base excision repair ³⁹. DDB1 can complex with a ubiquitin-E3 ligase, to target proteins involved in DNA repair and transcription for ubiquitination and degradation ⁴⁰. Association of DDB1 with Cockayne syndrome group A protein can modulate the transcription of genes induced by UV exposure ⁴⁰.

X-ray repair complementing defective repair in Chinese hamster cells 5 (XRCC5), is integral to DSB repair. It binds to DNA, and in conjunction with Ligase IV, ligates the ends of the damaged DNA. This is the final step in non-homologous end joining, which is the major pathway of DNA repair from DSBs ^{41,42}. In yeast, the loss of XRCC5 renders cells incapable of NHEJ ⁴³.

Structural maintenance of chromosomes and homologous recombination

Structural maintenance of chromosomes (SMC) proteins are key components of complexes that have a variety of cellular functions involving chromosome structure and remodeling during the cell cycle and DNA damage repair ⁴⁴. SMC3, a component of

Cohesin, and SMC2 and 4, components of Condensin, are upregulated following YEL2 treatment.

The Cohesin complex plays an important role during the cell cycle and in DNA damage repair. During mitosis, Cohesin enables sister chromatids to remain together until anaphase, when the complex is cleaved, and the sister chromatids can separate. In the role of DNA repair, SMC3, is an indispensable component of the Cohesion complex, and is required for homologous recombination (HR)⁴⁵. The sister chromatid tethering functions of Cohesin are important DNA DSB repair after ionizing radiation exposure, since sister chromatids serve as templates for HR^{46,47}. Cohesion can also bind near DNA damage sites, and controls the degree of γ -H2AX spreading and the subsequent chromatin remodeling⁴⁵. RAD21 has been found to associate with Cohesion and to be important for DNA repair, as insufficiencies in Rad21-Cohesin decrease radioresistance in mice⁴⁸.

The other two SMCs, 2 and 4, bind each other and associate with either the Condensin I or the Condensin II complex. These complexes condense chromosomes during mitosis, and, in conjunction with Cohesion, can control chromosome structure during metaphase⁴⁹. In humans, Condensin I can interact with the PARP1-XRCC1 complex to effect single strand DNA breaks, while Condensin II is involved with HR of DSBs induced by IR^{50,51}.

Condensin and cohesin exhibit important roles in sister chromatid cohesion, and chromosome condensation, in addition to effects on DNA damage repair by homologous recombination. The upregulation of these components that are integral for genome stability and general chromosome biology are important.

Cyclin dependent kinases

Cyclin dependent kinases are the master regulators of the cell. Progression through the major cell cycle phases is regulated by CDKs and their associated cyclins. Mutations in the CDKs are found in a variety of cancers, in which there is over activity of the cell cycle and a loss of checkpoint integrity ⁵².

The DNA damage response can shut down the cell cycle via checkpoint 1 and checkpoint 2 activation. Recent studies have shown that the CDKs can coordinate or mediate DNA repair. Additionally, the centrosomes have been found to be important in proper DNA repair after IR exposure, with various DNA repair proteins localizing to the centrosome to regulate centrosome duplication, either directly, or via CDK2 ⁵³. The involvement of CDKs in HR in mammalian systems has yet to be determined. CDKs act directly to influence DNA repair, or influence the pathway activated in response to DNA damage. Following YEL2 treatment, there is an upregulation of CDK1, 2, 8, 11B, 12, and 13.

CDK1 is involved in modulating the onset of mitosis, promoting the G2-M and the G1-S transitions. It is also required in higher cells for entry into S-phase and mitosis, and has been found to be able to drive the cell cycle alone ⁵⁴. In yeast, CDK1 can promote DNA repair through HR during the G2 phase ⁵⁵.

CDK2 is a serine threonine protein kinase interphase control. It promotes the E2F transcriptional program at the G1-S transition and initiates DNA synthesis. A range of phosphorylation substrates exist, including MYC and TP53, and is involved in balancing proliferation, DNA repair, and apoptosis. In mice, deletion of Cdk2 increased radiosensitivity and delayed DSB repair ⁵⁵.

CDK8 has a number of roles outside of cell cycle control, including the regulation of RNA polymerase II-dependent genes. CDK8 and CDK11B are involved in transcription⁵⁶. CDK11B is involved in cell cycle progression, cytokinesis and apoptosis.

CDK12 and CDK13 are required for RNA splicing, and display kinase activity towards RNA polymerase II. CDK12 regulates transcription elongation through phosphorylation of RNA polymerase II, and is involved in RNA splicing⁵².

The CDKs have a variety of functions outside of cell cycle regulation. Of these, the most important are related to DNA repair from UV-radiation, IR, or DNA damaging agents. The upregulation in these CDKs may implicate their function in increasing cellular resilience to lethal doses of radiation, or increasing their ability to repair DNA and faithfully duplicate their genome.

ATM Signaling and Cell Cycle Control

			N24+YEL2 / N24+PBS		
Accession	Gene	Description	In ratio	p-value	PLGS Score
P45973	CBX5	Chromobox protein homolog 5	N24+YEL2		1,540
P06493	CDK1	Cyclin dependent kinase 1	N24+YEL2		814
Q9NZS5	CDK11B	Cyclin dependent kinase 11B	N24+YEL2		520
Q9NYV4	CDK12	Cyclin dependent kinase 12	N24+YEL2		806
Q9UDR4	CDK13	Cyclin dependent kinase 13	N24+YEL2		942
P24941	CDK2	Cyclin dependent kinase 2	N24+YEL2		1,252
P49336	CDK8	Cyclin dependent kinase 8	N24+YEL2		678
P27695	APEX	DNA apurinic or apyrimidinic site lyase	0.48	0.98	2,102
Q16531	DDB1	DNA damage binding protein 1	0.22	0.95	1,909
Q92878	RAD50	DNA repair protein RAD50	N24+YEL2		713
Q13426	XRCC4	DNA repair protein XRCC4	N24+YEL2		715
P49959	MRE11A	Double strand break repair protein MRE11A	0.49	1.00	1,647
O60934	NBN	Nibrin	N24+YEL2		678
P40937	RCF5	Replication factor C subunit 5	N24+YEL2		703
Q13315	ATM	Serine protein kinase ATM	N24+YEL2		604
Q9Y2H1	NDR2	Serine threonine protein kinase 38 like	4.66	1.00	1,058
		Serine threonine protein phosphatase 2A 55			
P63151	PPP2R2A	kDa regulatory subunit B alpha isoform	N24+YEL2		954
O95347	SMC2	SMC2	0.61	1.00	868
Q9UQE7	SMC3	SMC3	0.66	1.00	1,407
Q9NTJ3	SMC4	SMC4	N24+YEL2		537

Table 8-3. YEL2-modulated proteins in N24 LCLs related to ATM Signaling and Cell Cycle regulation. This table represents identification and quantification information for proteins from 4 related pathways identified as enriched in the YEL2 treated sample by IPA. All proteins were upregulated or uniquely identified in the YEL2 treated condition.

Rho family GTPases enriched with YEL2 treatment

The Rho family of guanine nucleotide binding proteins (G proteins) that regulate many aspects of intracellular actin dynamics, proliferation, and gene expression⁵⁷. A variety of these proteins are upregulated or unique to YEL2 treated normal LCLs (Signaling by Rho Family of GTPases, p-value=9.77E-09) (Table 8-4). These G proteins participate in transducing signals from a variety of systems, including olfactory, visual, and hormonal. Each G protein is composed of 3 subunits (α , β , and γ). After

YEL2 treatment, many members and of this family, and related proteins, are found in greater abundance and potentially affect a wide range of cellular functions.

Several of these have been associated with disease states. The Rho-associated kinase, ROCK2, operates directly downstream of Rho proteins ⁵⁸, and have already been associated with proliferation, inflammation, and oxidative stress ⁵⁹. GNAT12 and GNAT13 are involved in a variety of diseases and cancer metastasis ⁶⁰. GFAP is a cell marker involved in differentiation of astrocytes from other glial cells during development, and mutations in GHFAP can cause Alexander's Disease ⁶¹. GNB2L1 (RACK1) can inhibit the activity of SRC kinases, and prolong the G0/G1 phase of the cell cycle ⁶².

NF- κ B is a transcription factor involved in many biological processes, from stress response, to inflammation, to apoptosis. NF- κ B dimerizes with itself, or a variety of partners, and then binds κ B sites on DNA near its target genes. Studies on cancer radiotherapy implicate prosurvival signaling by NF- κ B as a means of tumor cell acquired radioresistance, or exhibit radioadaptation ⁶³. This effect is thought to be the reason that low doses of radiation, prior to high doses, results in reduced DNA damage and increased survival in many organisms, including mice, rabbits, and human lymphocytes ⁶⁴⁻⁶⁶. Elevated levels of NF- κ B have been associated with chemoresistance and radioresistance in certain cancers ⁶⁷. Blocking NF- κ B, or silencing it, in these contexts is a promising avenue in abrogating tumor cell resistance to radiation exposure.

Signaling by Rho Family GTPases

			N24+YEL2 / N24+PBS		PLGS
Accession	Gene	Description	ln ratio	p-value	Score
P61160	ACTR2	Actin related protein 2	1.03	1.00	4,983
Q9UQB8	BAIAP2	Brain specific angiogenesis inhibitor 1 associated protein 2	N24+YEL2		772
Q96KS4	GFAP	Glial fibrillary acidic protein	N24+YEL2		862
P16520	GNB3	Guanine nucleotide binding protein G I G S G T subunit beta 3	1.97	0.99	1,812
P63096	GNAI1	Guanine nucleotide binding protein G i alpha 1	0.90	1.00	4,477
P04899	GNAI2	Guanine nucleotide binding protein G i alpha 2	0.48	1.00	12,251
P08754	GNAI3	Guanine nucleotide binding protein G k alpha	0.71	1.00	7,193
P09471	GNAO1	Guanine nucleotide binding protein G o alpha	0.76	1.00	3,589
P38405	GNAL	Guanine nucleotide binding protein G olf alpha	0.78	1.00	3,831
Q5JWD5	GNAS	Guanine nucleotide binding protein G s alpha isoforms short	0.75	1.00	4,355
Q5JWF2	GNAS	Guanine nucleotide binding protein G s alpha isoforms XLas	0.50	1.00	4,966
P11488	GNAT1	Guanine nucleotide binding protein G t alpha 1	0.85	1.00	3,787
P19087	GNAT2	Guanine nucleotide binding protein G t alpha 2	0.91	1.00	4,074
A8MTJ3	GNAT3	Guanine nucleotide binding protein G t alpha 3	0.76	1.00	3,938
P63211	GNAT1	Guanine nucleotide binding protein G T gamma T1	N24+YEL2		632
Q03113	GNA12	Guanine nucleotide binding protein alpha 12	0.60	1.00	7,487
Q14344	GNA13	Guanine nucleotide binding protein alpha 13	0.78	1.00	3,336
Q6FH47	GNB2L1	Guanine nucleotide binding protein beta 2 like 1	0.48	1.00	4,243
Q9HAV0	GNB4	Guanine nucleotide binding protein beta 4	1.84	0.96	882
P27361	MAPK3	Mitogen activated protein kinase 3	N24+YEL2		672
P19105	MYL12A	Myosin regulatory light chain 12A	1.14	1.00	5,789
O14950	MYL12B	Myosin regulatory light chain 12B	1.18	1.00	8,903
Q96A32	MYLPP	Myosin regulatory light chain 2 skeletal muscle	N24+YEL2		1,602
Q00653	NFKB2	Nuclear factor NF kappa B p100 subunit	N24+YEL2		586
Q3ZK23	DIAPH3	Protein diaphanous homolog 3	N24+YEL2		595
O75116	ROCK2	Rho associated protein kinase 2	N24+YEL2		769
Q92888	ARHGEF1	Rho guanine nucleotide exchange factor 1	0.27	0.91	1,327
Q92974	ARHGEF2	Rho guanine nucleotide exchange factor 2	0.81	1.00	1,508
Q92599	SEPT8	Septin 8	N24+YEL2		537

Table 8-4. Signaling by Rho Family GTPases is enriched in YEL2 treated Normal LCLs. The table represents identification and quantification information for those proteins grouped into this pathway by IPA.

YEL2 in the absence of functional ATM

Without functional ATM, YEL2 exerts effects on many DNA repair proteins involved in replication, recombination, or homologous recombination. These include AKT1, BAX, BAG1, ARF1, CDK1/2/4, DAP3, BSG, H-RAS, RAC1, S11, and REL. There is also an upregulation of mTOR signaling components AKT, H-RAS, R-RAS, MAPK1, RAC1, RHO A/C/G/J/T2. However, we do not identify the same profile, which may be due to the dependence of those effects on proper ATM signaling, or opposing regulation mediated by effectors present as a consequence of ATM-deficiency.

YEL2 Transcription factors in both Normal and ATM-deficient LCLs

The incubation of either cell line with YEL2, resulted in enrichment of proteins that overlap with MYC transcription factor. MYC is a multifunctional nuclear protein that is involved in cell cycle progression and apoptosis. Recent work has shown the requirement of MYC for the activation of checkpoints by ATM, and the lack of MYC resulted in a loss of ATM substrate phosphorylation following IR exposure ⁶⁸. MYC can initiate DNA repair during S phase in MYC overexpressing cells, which can be mediated by ATM, W1P1, and T60 ⁶⁹.

MYC overexpression can be pathological and has been associated with a variety of tumors. When overexpressed in immortalized cell lines, alters the G1/S checkpoint, which then allows cells to enter into S phase with unrepaired DNA ⁷⁰. A later study showed that overexpression of MYC was responsible for attenuating post irradiation G2/M arrest, which enhanced apoptosis due to damaged DNA ⁷¹.

Target genes of MYC, found in Normal LCLs include: APEX1, ATAD3A/ATAD3B, EXOSC8, KRAS, MRE11A, NAP1L1, NBN, PDK1. MYC target genes found in ATM-deficient cells: ATAD3A/ATAD3B, BAX, BCAT1, CDK4, ENO1, FBXO32, HK2, IGF2BP1, NBN, RCC1, SHMT2, TMSB10/TMSB4X.

Proteomic disturbances due to chronic lack of functional ATM

A comparison of baseline proteome profiles of Normal and ATM-deficient LCLs reveals many pathways that are differentially enriched. These protein networks may have involvement in AT biology and its clinical manifestations, compensatory functions, as well as the pleiotropic effects of ATM outside the context of DNA damage repair.

Proteomic consequences of ATM deficiency

In the absence of ATM, cells experience a great deal of oxidative stress. This can result in many primary and secondary effects on protein expression and signaling (Table 8-5). The enrichment of the Rho Family of GTPases was shown in our previous study in irradiated Normal LCLs, and it is also enriched in the AT LCLs cells without treatment. ATM activity, in response to DNA damage, can activate RhoA GTPase, and increase survival; however, hypoxia and ROS can mediate its activation via superoxide signaling⁷². The upregulation in this signaling family, without ATM, highlights a compensatory, or redundant, mechanism that ATM- deficient cells may utilize in the absence of ATM for dealing with oxidative stress and damage. In addition to Rho Family of GTPases, there are many protein pathways that function in preventing

apoptosis, and controlling the cell. These mechanisms are important in the chronic stress conditions

The enrichment profiles in ATM-deficiency overlap with the targets of a number of transcription factors, including E2F1, E2F4, FUBP1, and TP53. E2F1 and E2F4 are members of the E2F transcription family that controls cell cycle and tumor suppressor proteins. E2F1 is involved in the DNA damage response in the context of viral infection, γ H2AX foci formation, and the blockage of the cell cycle through CDK2 phosphorylation⁷³. Its role in these processes is thought to be mediated through the activation of ATM, by a non DSB-dependent mechanism⁷⁴. In fibroblasts deficient in ATM, E2F1 was unable to phosphorylate p53 and induce apoptosis⁷⁵.

Surprisingly, the enrichment of a variety of TP53 target genes is found in the ATM-deficient cell line. The main mode of activation of TP53 is by phosphorylation by ATM, in addition to ATR. Recent studies have shown that in the absence of functional ATM, ATR and DNA-PKcs are able to trigger DNA damage responses by phosphorylation of P53, SMC1, CHK1, and CHK2⁷⁶. The upregulation we find here may be a result of the low level activation of ATM substrates by ATR. Cells may be adapting to ATM-deficiency, and effecting survival in the presence of oxidative stress and DNA damage through the activation and upregulation of pathways that have redundant functions or that activate similar targets as ATM.

Pathways upregulated in the presence of functional ATM

In the presence of functional ATM, cells are able to cope with oxidative stress and DNA damage more effectively. ATM has many non-nuclear functions, and this

comprehensive analysis attempts to highlight some of the broad effects of ATM (Table 8-6).

Chronic oxidative stress can induce expression of hypoxia inducible factor (HIF1) ⁷⁷. HIF1 promotes the expression of various genes involved in adaptation to low oxygen conditions, including glycolytic enzymes and VEGF, which increase ATP production and blood vessel formation. HIF1 is thought to be a link between ATM deficiency and its clinical manifestations.

Many components of glycolysis and gluconeogenesis are enriched in N24 cells. In a previous study we have shown that ATM can suppress glycolysis in the presence of IR exposure and DNA damage, presumably in an attempt to reduce ROS production through that pathway. Switching from glycolysis to the pentose phosphate pathway can produce NADPH to function in antioxidant reactions, and increase nucleotide production ⁷⁸. In ATM-deficient cells, the lower abundance of glycolytic enzymes is likely an effort to reduce the production of ROS, as these cells are chronically stressed already ⁷⁹.

In the presence of functional ATM there is enrichment in components of integrin signaling. Recently, it was found that chemotherapy resistance in tumors could be mediated by integrin $\alpha\beta3$ upregulation. The upregulation was a consequence of constant activation of ATM and NF- κ B ⁸⁰. This observed upregulation of integrin signaling components in N24 LCLs is probably a normal feature that may be exploited in the context of chemoresistance.

T sterol regulatory element-binding protein-1 (SREBF1) is a transcription factor that binds the sterol regulatory element-1 (SRE1) on target genes related to glucose metabolism and fatty acid biosynthesis. It is thought that insulin can mediate its effects

via Akt activation and nuclear sequestration of SREBP-1⁸¹. mTOR is involved in the integration of energy status signals and the advancement of cell growth and metabolism, and is involved in the cleavage of SREBP-1⁸¹. Interestingly, mTOR signaling is enriched in the presence of ATM. This implies an association between cell growth control and ATM function, although it is possibly a secondary effect of better cellular conditions and less oxidative stress.

In the presence of functional ATM, LCLs exhibit the enrichment of many pathways involved in energy metabolism, and transcription factors related to growth. These effects may be independent of ATM, when activation is a consequence of oxidative stress, or dependent on ATM, when regulation is mediated by ATM or one of its substrates.

Enriched in Normal LCLs

Canonical pathways	<u>p-value</u>
<i>Integrin Signaling</i>	5.0E-11
FYN, MPRIP, CDC42, MAPK1, PIK3R1, ACTA2, RHOT2, PAK1, RHOG, ARF4, ARPC3, ARPC1A, ACTG2, VASP, ACTR2, SRC, RRAS, RAC1, RHOJ, RAC3, ROCK1, PIK3R3, ARF1, ARF5, TLN2, PAK3, ARF3, ARPC2, PLCG2, PPP1R12A, ZYX	
<i>Cdc42 Signaling</i>	1.0E-06
B2M, ACTR2, SRC, MPRIP, MAPK1, CDC42, HLA-DRB1, PAK1, PAK3, EXOC4, HLA-A, ARPC2, HLA-B, ARPC3, PPP1R12A, VAV1, ARPC1A, HLA-G, HLA-DRB5	
<i>Glycolysis/Gluconeogenesis</i>	3.0E-06
ALDH4A1, GAPDH, ENO2, PGAM2, TPI1, PGAM4, PFKL, ADH5, ALDH1A1, HK2, ALDOB, PGAM1, ALDH1A2, DHRS4	
<i>mTOR Signaling</i>	1.5E-05
MAPK1, RRAS, RPS17/RPS17L, RPS28, PIK3R1, RHOT2, RAC1, EIF3J, EIF4A2, RHOJ, RPS21, PRKAG1, RPS24, PIK3R3, PPP2R1A, RHOG, RPS5, RPS3, PPP2R1B, EIF3K, PRKCB	
<i>NRF2-mediated Oxidative Stress Response</i>	3.1E-04
GSTM1, MAPK1, RRAS, PIK3R1, ACTA2, DNAJC19, GSR, PIK3R3, GSTM2, SOD2, DNAJB11, UBE2K, GSTM4, MAP2K3, ACTG2, TXN, GSTK1, PRKCB	
<i>Oxidative Stress</i>	1.3E-05
GSR, GSTM1, VCAM1, GSTM2, ICAM1, SOD2, PRDX5, GPX1, MPO, STAT3, NFKB1	
<i>Mitochondrial Dysfunction</i>	3.0E-04
NDUFA4, PRDX5, COX6C, RHOT2, NDUFB4, NDUFB10, GSR, SOD2, UQCRC2, CYB5R3, COX5A, UQCRC1, NDUFS2, CYCS, UQCRC1	
Transcription Regulators	<u>p-value</u>
<i>SREBF1</i>	3.8E-03
CAPZB, CYP51A1, GSR, LGALS3, PLS3, RPS24, UQCRC1	
<i>HMGB1</i>	1.2E-02
BAX, HLA-DRB1, ICAM1, VCAM1	
<i>HSF2</i>	2.1E-02
HSPA7, TXN (includes EG:116484)	
<i>HIF1A</i>	2.7E-02
EIF5A, ENO2, ERGIC1, GAPDH, HK2, LGALS1, NAE1, NDRG1, NPM1, SESN2, VASP	
<i>CTCF (includes EG:10664)</i>	3.1E-02
BAX, HLA-DRB1	
<i>IRF4</i>	4.0E-02
B2M, GBP1, MANF	

Table 8-5. Canonical pathways and transcription regulators enriched in Normal LCLs versus ATM-deficient LCLs. Transcription regulators represent those transcription factors whose target genes overlap with proteins found in the enrichment data. Ingenuity pathway analysis p-values for enrichment significance are given.

Enriched in ATM-deficient LCLs

Canonical pathways	p-value
<i>Signaling by Rho Family GTPases</i>	1.6E-05
SEPT8, MYL5, GNAI1, VIM, NFKB2, DES, ACTG1, GNAI2, GNAI3, MYL12A, STMN1, RHOQ, MYL12B, CFL2, GNAO1, NCF2, BAIAP2, ARHGEF2, GFAP, GNA13, ACTC1, ACTA1, PI4KA	
<i>Cdc42 Signaling</i>	2.6E-05
MYL5, HLA-DRB1, HLA-DRB3 (includes others), HLA-DQB1, MYL12A, DIAPH1, HLA-DRB4, MYL12B, CFL2, BAIAP2, HLA-DRA, HLA-B, GSK3B, HLA-DPB1, HLA-DRB5, HLA-C	
<i>ILK Signaling</i>	1.3E-04
FLNB, MYH6, CASP3, PPP2R2A, MYH13, MYL5, VIM, NFKB2, RHOQ, FLNA, CFL2, KRT18, MYH3, GSK3B	
<i>Role of NFAT in Regulation of the Immune Response</i>	1.7E-04
CAMK4, NRAS, GNAI1, HLA-DRB1, KRAS, ITPR1, NFKB2, HLA-DQB1, GNAI2, RCAN1, GNAI3, SYK, GNAO1, HLA-DRA, GNA13, GSK3B, HLA-DRB5	
<i>Nur77 Signaling in T Lymphocytes</i>	5.8E-04
CAMK4, CASP3, HDAC2, HLA-DRA, HDAC1, HLA-DRB1, HLA-DQB1, HLA-DRB5	
<i>RhoA Signaling</i>	7.9E-04
SEPT8, MYL12A, MYL12B, CFL2, MYLK2, MYL5, BAIAP2, GNA13, ACTC1, ACTG1, PI4KA	
<i>RhoGDI Signaling</i>	2.1E-03
MYL5, GNAI1, GNAI2, GNAI3, RHOQ, MYL12B, CFL2, GNAO1, ARHGEF2, GNA13, PI4KA	
<i>Cell Cycle Control of Chromosomal Replication</i>	3.0E-03
MCM3, MCM6, MCM2, MCM4, MCM7	
<i>Anti-Apoptosis</i>	3.1E-02
TNFAIP8, TMX1, HDAC1, BAG2	
Transcription Regulators	
<i>E2F4</i>	4.4E-04
EXOSC8, HMGB1, MAD2L1, MCM3, MCM6, MKI67, MRTO4, MSH2, NDC80, RFC2, RRM1, RRM2, SLC3A2, SMC4, SPAG7, SRPR, TK1	
<i>FUBP1</i>	2.1E-03
STMN1, STMN3	
<i>TP53 (includes EG:22059)</i>	6.1E-03
ARHGEF2, CARS, EIF2B1, GNA13, GSK3B, HDAC2, KRT8, MAD2L1, MCM2, MCM3, MCM4, MCM6, MCM7, MRPL46, MSH2, MSH6, MYO6, NDC80, NFKB2, PDHX, PDK1, RRM2, RRM2B, SEC62 (includes EG:294912), SMC3, SMC4, SRSF3, STMN1, TOP2B, TPX2, TRIM22	
<i>E2F1</i>	1.8E-02
HMGB1, MCM3, MCM6, MSH2, NDC80, RFC2, RRM1, RRM2, SAFB, SLC3A2, SMC4, SRPR, TK1	

Table 8-6. Canonical pathways and transcription regulators enriched in ATM-deficient LCLs versus Normal LCLs. Transcription regulators represent those transcription factors whose target genes overlap with proteins found in the enrichment data. Ingenuity pathway analysis p-values for enrichment significance are given.

Previous work on ATM-dependent proteome effects

Marzano et al. published work defining the characteristics of ATM-proficient and ATM-deficient cell lines using quantitative proteomics⁸². They utilized human ATM-deficient LCLs that were transfected either L6pCDNA or L6-Flag-ATM-WT, in an attempt to collect direct proteomic changes due to ATM⁸². This is conceptually interesting; however, their use of shotgun LC-MS^E on unfractionated digests with such a complex proteome has major effects on the scope of their analysis and the reliability of their quantitative conclusions.

In their study, cells were processed and analyzed by shotgun LC-MS^E on a Q-ToF Premier (Waters, UK). Swissprot and PLGS were used for database searching and quantitative comparisons. They identified a total of 442 unique protein identifications from all samples analyzed; however, the frequency of protein identifications was very low (Table 8-8)⁸². Comparing ERLIC-LC-MS^E data from this study to the data presented by Marzano et al. quickly highlights the complications of interpreting shotgun LC-MS^E data of a complex samples.

One of the main complications is the assumption that if a protein is only identified in one sample, then it is likely upregulated in that sample. This estimation is fine when proteome coverage is good; however, in cases without pre-fractionation and poor proteome coverage, many reasons exist, other than differential regulation, to explain why a protein may not be identified in all samples. The most likely explanation is the large number of co-eluting species at any one time during the LC gradient, which can mask and suppress the ionization of other eluting species, and the resulting fragment ion spectra can be so complex that the search algorithm will not be able to properly

associate fragment ions with precursor ions. These two issues will inhibit a proper analysis by reducing protein identifications, and decreasing quantification quality and reliability.

Lack of proteome depth yields dimensionless conclusions

The lack of proteome coverage can be detrimental to the quantity and quality of identified proteins. A comparison of ERLIC-LC-MS^E with data from Marzano et al.⁸² yield some concordant results, even though many of these are not actually quantified in their data set.

Glutathione S-transferase Pi (GSTPi) is only identified in their ATM-deficient cell line, which implicates ATM-deficiency with the upregulation of GSTPi. In our data set, GSTPi is identified in both conditions, and downregulated in N24+PBS (ln ratio (N/AT) = -0.51, and a p-value<0.01), which is concordant with their general conclusion; however, there is no indication of magnitude of change in their study.

Fumarate hydratase (FH) is only identified in their ATM-proficient cells, and is associated with ATM presence. In our data, this protein is quantified in both samples, and is upregulated in N24+PBS (ln ratio (N/AT) = 0.73, p-value=1.00). Again, the general conclusion is concordant.

There is also a low representation of proteins that play a significant role in cell fate, regulation and expression, like DNA damage proteins and transcription factors. Increased proteome coverage can enable more robust and reliable quantification, provide magnitude information, and enable more reliable differential regulation determination based on unique identifications.

Data deficiency results in inappropriate assumptions

While the conclusions reached by Marzano et al. based on unique identifications can be concordant with our data set, there are many instances where the assumption of differential regulation proves to be misleading.

The Interferon induced GTP binding protein, Mx1, is only identified in their ATM-deficient cell line, which leads to the conclusion that it is upregulated in the absence of ATM. Mx1 is identified and quantified in both our cell lines, with upregulation in N24+PBS (ln ratio (N/AT) = 0.83, with p-value=1.00). So, in the absence of ATM, Mx1 is actually downregulated.

The DNA repair protein, XRCC6, is identified only in their ATM-proficient cells, and is included in upregulated proteins associated with functional ATM. In our data set, XRCC6 is quantified in both cell lines, and is downregulated in N24+PBS (ln ratio (N/AT) = -0.19, p-value<0.01). This indicates downregulation in the presence of ATM, which is opposite their conclusion, and based on our threshold of differential regulation, would be considered unchanged.

Protein disulfide isomerase A3 is upregulated in their data set with a natural log ratio ATM-proficient/ATM-deficient of 0.39. In our data set, this protein is not significantly altered. In a case of quantification discrepancy, is the discordance due to a lack of protein coverage in their data set, or suppression of peptides that were utilized for quantification? The data presented by Marzano et al. is misleading, at best.

Count	Frequency	%
12	52	11.8
11	11	2.5
10	14	3.2
9	4	0.9
8	9	2.0
7	11	2.5
6	14	3.2
5	12	2.7
4	27	6.1
3	29	6.6
2	47	10.6
1	212	48.0
<i>total unique IDs</i>	442	

Table 8-8. Frequency of protein identifications across the samples from Marzano et al.⁸². The number of times a protein was identified in each technical replicate analysis was counted. 52 proteins (11.8%) were identified in all of their samples, out of 442 unique protein identifications.

REFERENCES

1. Kim, K. *et al.* High-throughput screening identifies two classes of antibiotics as radioprotectors: tetracyclines and fluoroquinolones. *Clin Cancer Res* **15**, 7238–7245 (2009).
2. Hershman, J. M., Okunyan, A., Rivina, Y., Cannon, S. & Hogen, V. Prevention of DNA double-strand breaks induced by radioiodide-131i in frtl-5 thyroid cells. *Endocrinology* **152**, 1130–1135 (2011).
3. Hosseinimehr, S. J. Trends in the development of radioprotective agents. *Drug Discov Today* **12**, 794–805 (2007).
4. Grdina, D. J., Murley, J. S. & Kataoka, Y. Radioprotectants: current status and new directions. *Oncology* **63**, 2–10 (2002).
5. McKinnon, P. J. ATM and ataxia telangiectasia. *EMBO Rep* **5**, 772–776 (2004).
6. Savitsky, K. *et al.* A single ataxia telangiectasia gene with a product similar to PI-3 kinase. *Science* **268**, 1749–1753 (1995).
7. Lavin, M. F., Gueven, N., Bottle, S. & Gatti, R. A. Current and potential therapeutic strategies for the treatment of ataxia-telangiectasia. *Br Med Bull* **81-82**, 129–147 (2007).
8. Lavin, M. F. & Kozlov, S. ATM activation and DNA damage response. *Cell Cycle* **6**, 931–942 (2007).
9. Syllaba, L. & Henner, K. Contribution à l'indépendance de l'athétose double idiopathique et congénitale. Atteinte familiale, syndrome dystrophique, signe du réseau vasculaire conjonctival, intégrité psychique. *Rev Neurol* **1**, 541–562 (1926).
10. Boder, E. & Sedgwick, R. Ataxia-telangiectasia. A familial syndrome of progressive cerebellar ataxia, oculocutaneous telangiectasia and frequent pulmonary infection. A preliminary report on 7 children, an autopsy, and a case history. *Univ Soc Calif Med Bull* **9**, 15–28 (1957).
11. Morgan, J. L., Holcomb, T. M. & Morrissey, R. W. Radiation reaction in ataxia telangiectasia. *Am J Dis Child* **116**, 557–558 (1968).
12. Taylor, A. M. *et al.* Ataxia telangiectasia: a human mutation with abnormal radiation sensitivity. *Nature* **258**, 427–429 (1975).
13. Chen, P. C., Lavin, M. F., Kidson, C. & Moss, D. Identification of ataxia telangiectasia heterozygotes, a cancer prone population. *Nature* **274**, 484–486 (1978).

14. Gatti, R. A. *et al.* Localization of an ataxia-telangiectasia gene to chromosome 11q22-23. *Nature* **336**, 577–580 (1988).
15. Platzer, M. *et al.* Ataxia-telangiectasia locus: sequence analysis of 184 kb of human genomic DNA containing the entire ATM gene. *Genome Res* **7**, 592–605 (1997).
16. Uziel, T. *et al.* Genomic Organization of the ATM gene. *Genomics* **33**, 317–320 (1996).
17. Eng, L. *et al.* Nonclassical splicing mutations in the coding and noncoding regions of the ATM Gene: Maximum entropy estimates of splice junction strengths. *Hum Mutat* **23**, 67–76 (2003).
18. Barzilai, A. & Yamamoto, K. I. DNA damage responses to oxidative stress. *DNA Repair* **3**, 1109–1115 (2004).
19. Barzilai, A., Rotman, G. & Shiloh, Y. ATM deficiency and oxidative stress: a new dimension of defective response to DNA damage. *DNA Repair* **1**, 3–25 (2002).
20. Barlow, C. *et al.* Loss of the ataxia-telangiectasia gene product causes oxidative damage in target organs. *Proc Natl Acad Sci U S A* **96**, 9915–9919 (1999).
21. Kamsler, A. *et al.* Increased oxidative stress in ataxia telangiectasia evidenced by alterations in redox state of brains from Atm-deficient mice. *Cancer Res* **61**, 1849–1854 (2001).
22. Chen, P. *et al.* Oxidative stress is responsible for deficient survival and dendritogenesis in purkinje neurons from ataxia-telangiectasia mutated mutant mice. *J Neurosci* **23**, 11453–11460 (2003).
23. Ambrose, M., Goldstine, J. V. & Gatti, R. A. Intrinsic mitochondrial dysfunction in ATM-deficient lymphoblastoid cells. *Hum Mol Genet* **16**, 2154–2164 (2007).
24. Gueven, N. *et al.* Dramatic extension of tumor latency and correction of neurobehavioral phenotype in Atm-mutant mice with a nitroxide antioxidant. *Free Radic Biol Med* **41**, 992–1000 (2006).
25. Shackelford, R. E. *et al.* Iron chelators increase the resistance of Ataxia telangiectasia cells to oxidative stress. *DNA Repair* **3**, 1263–1272 (2004).
26. Lai, C. H. *et al.* Correction of ATM gene function by aminoglycoside-induced read-through of premature termination codons. *Proc Natl Acad Sci U S A* **101**, 15676–15681 (2004).
27. Masuda, T., Tomita, M. & Ishihama, Y. Phase transfer surfactant-aided trypsin digestion for membrane proteome analysis. *J Proteome Res* **7**, 731–740 (2008).

28. Yeung, Y. G., Nieves, E., Angeletti, R. H. & Stanley, E. R. Removal of detergents from protein digests for mass spectrometry analysis. *Anal Biochem* **382**, 135–137 (2008).
29. Li, G. Z. *et al.* Database searching and accounting of multiplexed precursor and product ion spectra from the data independent analysis of simple and complex peptide mixtures. *Proteomics* **9**, 1696–1719 (2009).
30. Silva, J. C., Gorenstein, M. V., Li, G. Z., Vissers, J. P. & Geromanos, S. J. Absolute quantification of proteins by LCMS^E: a virtue of parallel MS acquisition. *Mol Cell Proteomics* **5**, 144–156 (2006).
31. Silva, J. C. *et al.* Simultaneous qualitative and quantitative analysis of the *Escherichia coli* proteome: a sweet tale. *Mol Cell Proteomics* **5**, 589–607 (2006).
32. Silva, J. C. *et al.* Quantitative proteomic analysis by accurate mass retention time pairs. *Anal Chem* **77**, 2187–2200 (2005).
33. Dimmer, E. C. *et al.* The UniProt-GO Annotation database in 2011. *Nucleic Acids Res* **40**, D565–D570 (2012).
34. Binns, D. *et al.* QuickGO: a web-based tool for Gene Ontology searching. *Bioinformatics* **25**, 3045–3046 (2009).
35. Lavin, M. F. ATM and the Mre11 complex combine to recognize and signal DNA double-strand breaks. *Oncogene* **26**, 7749–7758 (2007).
36. Liu, Y. *et al.* Coordination of steps in single-nucleotide base excision repair mediated by apurinic/apyrimidinic endonuclease 1 and DNA polymerase beta. *J Biol Chem* **282**, 13532–13541 (2007).
37. Yang, J. *et al.* APEX1 gene amplification and its protein overexpression in osteosarcoma: correlation with recurrence, metastasis, and survival. *Technol Cancer Res Treat* **9**, 161–169 (2010).
38. Naidu, M. D., Mason, J. M., Pica, R. V., Fung, H. & Pena, L. A. Radiation resistance in glioma cells determined by DNA damage repair activity of Ape1/Ref-1. *J Radiat Res* **51**, 393–404 (2010).
39. Scrima, A. *et al.* Detecting UV-lesions in the genome: The modular CRL4 ubiquitin ligase does it best! *FEBS Lett* **585**, 2818–2825 (2011).
40. Iovine, B., Iannella, M. L. & Bevilacqua, M. A. Damage-specific DNA binding protein 1 (DDB1): a protein with a wide range of functions. *Int J Biochem Cell Biol* **43**, 1664–1667 (2011).

41. Li, Z. & Alt, F. W. Identification of the XRCC4 gene: complementation of the DSB and V(D)J recombination defects of XR-1 cells. *Curr Top Microbiol Immunol* **217**, 143–150 (1996).
42. Yano, K. I., Morotomi-Yano, K. & Akiyama, H. Cernunnos/XLF: a new player in DNA double-strand break repair. *Int J Biochem Cell Biol* **41**, 1237–1240 (2009).
43. Jeggo, P. A. Identification of genes involved in repair of DNA double-strand breaks in mammalian cells. *Radiat Res* **150**, S80–91 (1998).
44. Nasmyth, K. & Haering, C. H. The structure and function of SMC and kleisin complexes. *Annu Rev Biochem* **74**, 595–648 (2005).
45. Caron, P. *et al.* Cohesin protects genes against gammaH2AX induced by DNA double-strand breaks. *PLoS Genet* **8**, e1002460 (2012).
46. Bauerschmidt, C. *et al.* Cohesin promotes the repair of ionizing radiation-induced DNA double-strand breaks in replicated chromatin. *Nucleic Acids Res* **38**, 477–487 (2010).
47. McAleenan, A. *et al.* SUMOylation of the alpha-kleisin subunit of cohesin is required for DNA damage-induced cohesion. *Curr Biol* **22**, 1564–1575 (2012).
48. Xu, H. *et al.* Rad21-cohesin haploinsufficiency impedes DNA repair and enhances gastrointestinal radiosensitivity in mice. *PLoS ONE* **5**, e12112 (2010).
49. Stephens, A. D., Haase, J., Vicci, L., Taylor, R. M. 2. & Bloom, K. Cohesin, condensin, and the intramolecular centromere loop together generate the mitotic chromatin spring. *J Cell Biol* **193**, 1167–1180 (2011).
50. Wood, J. L., Liang, Y., Li, K. & Chen, J. Microcephalin/MCPH1 associates with the Condensin II complex to function in homologous recombination repair. *J Biol Chem* **283**, 29586–29592 (2008).
51. Heale, J. T. *et al.* Condensin I interacts with the PARP-1-XRCC1 complex and functions in DNA single-strand break repair. *Mol Cell* **21**, 837–848 (2006).
52. Lapenna, S. & Giordano, A. Cell cycle kinases as therapeutic targets for cancer. *Nat Rev Drug Discov* **8**, 547–566 (2009).
53. Shimada, M. & Komatsu, K. Emerging connection between centrosome and DNA repair machinery. *J Radiat Res* **50**, 295–301 (2009).
54. Satyanarayana, A. & Kaldis, P. Mammalian cell-cycle regulation: several Cdks, numerous cyclins and diverse compensatory mechanisms. *Oncogene* **28**, 2925–2939 (2009).

55. Wohlbold, L. & Fisher, R. P. Behind the wheel and under the hood: functions of cyclin-dependent kinases in response to DNA damage. *DNA Repair* **8**, 1018–1024 (2009).
56. Malumbres, M. & Barbacid, M. Cell cycle, CDKs and cancer: a changing paradigm. *Nat Rev Cancer* **9**, 153–166 (2009).
57. Boureux, A., Vignal, E., Faure, S. & Fort, P. Evolution of the Rho family of ras-like GTPases in eukaryotes. *Mol Biol Evol* **24**, 203–216 (2007).
58. Croft, D. R. & Olson, M. F. Transcriptional regulation of Rho GTPase signaling. *Transcription* **2**, 211–215 (2011).
59. Liao, J. K., Seto, M. & Noma, K. Rho kinase (ROCK) inhibitors. *J Cardiovasc Pharmacol* **50**, 17–24 (2007).
60. Kozasa, T., Hajicek, N., Chow, C. R. & Suzuki, N. Signalling mechanisms of RhoGTPase regulation by the heterotrimeric G proteins G12 and G13. *J Biochem* **150**, 357–369 (2011).
61. Brenner, M. *et al.* Mutations in GFAP, encoding glial fibrillary acidic protein, are associated with Alexander disease. *Nat Genet* **27**, 117–120 (2001).
62. Chang, B. Y., Conroy, K. B., Machleder, E. M. & Cartwright, C. A. RACK1, a receptor for activated C kinase and a homolog of the beta subunit of G proteins, inhibits activity of src tyrosine kinases and growth of NIH 3T3 cells. *Mol Cell Biol* **18**, 3245–3256 (1998).
63. Ahmed, K. M. & Li, J. J. NF-kappa B-mediated adaptive resistance to ionizing radiation. *Free Radic Biol Med* **44**, 1–13 (2008).
64. Liu, S. Z., Cai, L. & Sun, S. Q. Induction of a cytogenetic adaptive response by exposure of rabbits to very low dose-rate gamma-radiation. *Int J Radiat Biol* **62**, 187–190 (1992).
65. Shadley, J. D. & Wolff, S. Very low doses of X-rays can cause human lymphocytes to become less susceptible to ionizing radiation. *Mutagenesis* **2**, 95–96 (1987).
66. Bhattacharjee, D. Role of radioadaptation on radiation-induced thymic lymphoma in mice. *Mutat Res* **358**, 231–235 (1996).
67. Orłowski, R. Z. & Baldwin, A. S. J. NF-kappaB as a therapeutic target in cancer. *Trends Mol Med* **8**, 385–389 (2002).
68. Guerra, I. *et al.* Myc is required for activation of the ATM-dependent checkpoints in response to DNA damage. *PLoS ONE* **5**, e8924 (2010).

69. Campaner, S. & Amati, B. Two sides of the Myc-induced DNA damage response: from tumor suppression to tumor maintenance. *Cell Div* **7**, 6 (2012).
70. Sheen, J. H. & Dickson, R. B. Overexpression of c-Myc alters G(1)/S arrest following ionizing radiation. *Mol Cell Biol* **22**, 1819–1833 (2002).
71. Sheen, J. H., Woo, J. K. & Dickson, R. B. c-Myc alters the DNA damage-induced G2/M arrest in human mammary epithelial cells. *Br J Cancer* **89**, 1479–1485 (2003).
72. Broughton, B. R., Jernigan, N. L., Norton, C. E., Walker, B. R. & Resta, T. C. Chronic hypoxia augments depolarization-induced Ca²⁺ sensitization in pulmonary vascular smooth muscle through superoxide-dependent stimulation of RhoA. *Am J Physiol Lung Cell Mol Physiol* **298**, L232–L242 (2010).
73. E, X. *et al.* An E2F1-mediated DNA damage response contributes to the replication of human cytomegalovirus. *PLoS Pathog* **7**, e1001342 (2011).
74. Li, Y. *et al.* E2F1-dependent pathways are involved in amonafide analogue 7-d-induced DNA damage, G2/M arrest and apoptosis in p53-deficient K562 cells. *J Cell Biochem* **113**, 3165–3177 (2012).
75. Powers, J. T. *et al.* E2F1 uses the ATM signaling pathway to induce p53 and Chk2 phosphorylation and apoptosis. *Mol Cancer Res* **2**, 203–214 (2004).
76. Tomimatsu, N., Mukherjee, B. & Burma, S. Distinct roles of ATR and DNA-PKcs in triggering DNA damage responses in ATM-deficient cells. *EMBO Rep* **10**, 629–635 (2009).
77. Ousset, M. *et al.* Loss of ATM positively regulates the expression of hypoxia inducible factor 1 (HIF-1) through oxidative stress: role in the physiopathology of the disease. *Cell Cycle* **9**, 2814–2822 (2010).
78. Cosentino, C., Grieco, D. & Costanzo, V. ATM activates the pentose phosphate pathway promoting anti-oxidant defence and DNA repair. *EMBO J* **30**, 546–555 (2011).
79. Stern, N. *et al.* Accumulation of DNA damage and reduced levels of nicotine adenine dinucleotide in the brains of Atm-deficient mice. *J Biol Chem* **277**, 602–608 (2002).
80. Jinushi, M. *et al.* ATM-mediated DNA damage signals mediate immune escape through integrin- α v β 3-dependent mechanisms. *Cancer Res* **72**, 56–65 (2012).
81. Laplante, M. & Sabatini, D. M. An emerging role of mTOR in lipid biosynthesis. *Curr Biol* **19**, R1046–52 (2009).

82. Marzano, V. *et al.* Proteomic profiling of ATM kinase proficient and deficient cell lines upon blockage of proteasome activity. *J Proteomics* **75**, 4632–4646 (2012).