AN ANALYSIS OF DYNAMIC PROTEOMES WITH ENHANCED SAMPLE PREPARATION AND COMPUTATION

by

Erin M. M. Weisenhorn

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

(Biomolecular Chemistry)

at the

UNIVERSITY OF WISCONSIN–MADISON

2019

Date of final oral examination: 04/02/2019

The dissertation is approved by the following members of the Final Oral Committee:
Joshua J. Coon, Professor, Chemistry and Biomolecular Chemistry
Melissa M. Harrison, Associate Professor, Biomolecular Chemistry
Christina M. Hull, Professor, Biomolecular Chemistry
Lloyd M. Smith, Professor, Chemistry

ACKNOWLEDGMENTS

First I would like to thank Prof. Coon for his mentorship, support, and guidance during my time in graduate school. The Coon laboratory is an inspiring place to come to work every day due to his continual efforts to acquire the best resources, instrumentation, and personnel to be a part of the team. The diversity of high quality research I have been exposed to continues to motivate me and broaden my thinking. I also would like to thank my committee Christina Hull, Melissa Harrison, Lloyd Smith, and Doug Weibel for their insight and recommendations throughout my time in graduate school.

When I first joined the lab Tim Rhoads, Alicia Richards, and Alex Hebert were instrumental in teaching me the fundamentals of the lab's methods as well as important trouble-shooting skills. Upon starting in the lab I had never touched a mass spectrometer and it is due to their continual patience that my confidence and competence began to grow. Throughout my time in graduate school Alex Hebert and Mike Westphall have answered many questions as well as provided much needed comic relief. Jason Russell was also an invaluable mentor who guided some of my first collaborative projects and later provided beneficial counseling as I navigated my next steps after graduate school. I would also like to thank Katie Overmyer for her guidance and assistance in many of the projects described here. I was fortunate to start graduate school one year after IPiB students Elyse Freiburger and Evgenia Shishkova who have expertly bridged the gap between role models and friends. Having remained in the lab as a staff scientist, Evgenia has continued to support me whether

by reviewing papers and presentations, or prescribing a shot of tequila. In reality, all the past and present members of the Coon lab have enabled this thesis and provided wonderful companionship.

Before I arrived at graduate school I was lucky to have many other individuals who encouraged my career in science. All the way in high school I had a wonderful teacher, Chuck Schelle, who favored independent thinking and creative problem solving even if it meant filling his classroom with goat bones. Having the opportunity to complete an undergraduate thesis prepared me for graduate school much more than I realized, and this was possible through the mentorship of Danielle Cass and Arthur Glasfeld. During college I also had the amazing opportunity to spend summers at NASA Ames where Chad Paavola gave me my first introduction to laboratory science. The time spent in the lab with Suzanne Chan and Julia DeSimone were some of my most fun and entertaining times doing science and they had a strong impact on my decision to pursue graduate school.

Outside the lab in Madison, I was lucky to start graduate school with a small and tight-knit group of students that made up the 2014 IPiB cohort. All of them have played an integral role in making my graduate school career an enjoyable experience and I am grateful for our many annual traditions including Friendsgiving, Secret Santa, and 4^{th} of July boat rides. I also met an incredible group of people in the local improv community and slaying imaginary monsters with them provided a wonderful respite from lab.

Throughout all of these times my family has provided unwavering support, encouragement, and acceptance of my time at the University. My parents Nora and Dave always

believed in my abilities and fostered an independence and perseverance (they might say stubbornness) which have served me well in graduate school and life. The opportunities in my life they made possible have truly shaped me in every sense. I am also grateful to my mother and father-in-law Jenice and Curt who have welcomed me to the family and Wisconsin.

Finally, I want to thank my husband Andy whose patience and love have kept me happy through the sometimes trying times in graduate school. He has selflessly agreed to move across the country with me twice to support my career and always steps up in handling many of our shared responsibilities when I am overwhelmed. Whenever I come home he never fails to put a smile on my face, although hopefully he is not disappointed that I have not completed his daily admonishment to “Cure something!" I can’t wait to share our next adventures together as well as the rest of my life.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS AND ACRONYMS

2,3-DPG	2,3-disphosphoglycerate
3PG	3-phosphoglycerate
6PGD	6-phosphogluconate dehydrogenase
ACN	Acetonitrile
ADP	Adenosine diphosphate
AGC	Automatic gain control
ALDOA	fructose bisphosphate aldolase
AMP	Adenosine monophosphate
ATP	Adenosine triphosphate
BPGM	Bisphosphoglycerate mutase
C	Celsius
C#	C sharp (a programming language)
CAA	Chloroacetamide
CA1	Carbonic anhydrase
CAD	Collision-activated dissociation
Da	Dalton (the atomic mass unit)
DHAP	Dihydroxyacetone phosphate
DMI2	Does not Make Infection 2
DNA	Deoxyribonucleic acid

DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
ENO1	Alpha-enolase
ESI	Electrospray Ionization
ETD	Electron transfer dissociation
FC	Fold-change
FDP	Fructose diphosphate
FDR	False discovery rate
g	Gram
G6P	Glucose 6 phosphate
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GCLC	Glutamate-cysteine ligase
GO	Gene ontology
GPx1	Glutathione peroxidase 1
GPx4	Glutathione peroxidase 4
GPI	Glucose phosphate isomerase
GSH	Glutathione reduced
GSSG	Glutathione oxidized
GST	Glutathione S transferase
h	Hour

HCD	Higher-energy collisional dissociation
IPA	Isopropanol
K000	Lysine with no isotopic substituents
K080	Lysine with eight hydrogen replaced with deuterium
K602	Lysine with six carbons replaced with carbon 13 and two nitrogen replaced with nitrogen 15
L	Liter
L000	Leucine with no isotopic substituents
L601	Leucine with six carbons replaced with carbon 13 and one nitrogen replaced with nitrogen 15
LC-MS/MS	Liquid chromatography coupled online with tandem mass spectrometry
LDH	Lactate dehydrogenase
LFQ	Label free quantitation
LysM	Lysin motif
LYK3	Lysin motif receptor-like kinase 3
m	Meter
mg	Milligram
M	Molar concentration
mM	Millimolar concentration
m	Mass

m/z	Mass-to-charge ratio
min	Minute
mRNA	Messenger ribonucleic acid
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
MS^1	Survey mass analysis
MW	Molecular weight
NEM	N-ethylmaleimide
NH4Ac	Ammonium acetate
nLC	Nanoflow liquid chromatography
NORK	Nodulation Receptor Kinase
OMSSA	Open Mass Spectrometry Search Algorithm
PBS	Phosphate buffered saline
PC	Principal component
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PEP	Phosphoenolpyruvate
PFK	Phosphofructokinase
PGAM1	Phosphoglycerate mutase
PGK1	Phosphoglycerate kinase

PGM2	Phosphoglucomutase-2
pH	Potential of hydrogen
PKLR	Pyruvake kinase
PMSF	Phenylmethane sulfonyl fluoride
ppm	Part per million
psi	Pounds per square inch
PSM	Peptide-spectrum match
PTM	Post-translational modification
p-value	Probability value
PUB	Plant U-box type E3 ubiquitin ligases
RBC	Red blood cells
RIA	Ratio of isotope abundance
RIA_{inf}	Ratio of isotope abundance at infinite time
RNA	Ribonucleic acid
RNAPI	RNA polymerase I
RT	Retention time
s	Second
s.d.	Standard deviation
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
S/N	Signal to noise

SWATH	Sequential window acquisition of all theoretical mass spectra
TAG	Triacylglycerol
TEAB	Triethylammonium bicarbonate
TCEP	Tris(2-carboxyethyl)phosphine
TFA	Trifluoroacetic acid
Th	Thomson (the unit of the mass-to-charge ratio)
TIC	Total-ion chromatogram
TMT	Tandem mass tags
Ub	Ubiquitin
UND	U-box N-terminal Domain
UPS	Ubiquitin-proteasome system
WT	Wild-type

ABSTRACT

Rapid advances in proteomic technologies have facilitated new and unprecedented biological studies. Despite these advances, both the preparation and analysis of large data sets require improvement to derive meaning. This dissertation addresses sample preparation and computational improvements to expedite the analysis of particularly challenging biological samples. **Chapter 1** provides an overview of the current methods employed to robustly quantitate proteins using mass spectrometry. In **Chapter 2**, proteomic technologies are applied to clinical blood samples from twins in a high-throughput fashion. These results are integrated with the metabolomic phenotype to create a multi-faceted model of red blood cell behavior in storage, and predict its efficacy for transfusion. For **Chapter 3,** the analysis shifts to the model organism *Saccharomyces cerevisiae* and particularly its response to a zinc deficient environment. In **Chapter 4**, immunoaffinity enrichment is employed to study the low abundance modification ubiqutination. This modification plays an important role in the symbiotic relationship between a model legume and a nitrogen-fixing bacterium. Finally, in **Chapter 5**, an isotope labeling method is applied to trace protein turnover across tissues in a mouse. **Chapter 6** provides an overview of the work contained herein and potential future directions for the field.

Chapter 1

INTRODUCTION

Background

In the last decade mass spectrometry has revolutionized the types of questions scientists interested in studying an organism can ask. Technical innovations have facilitated routine sequencing of thousands of proteins from simple organisms and cell lines in record time[1–3]. However, applying these technologies to many eukaryotic tissues poses remaining challenges such as broad dynamic range or limited sample quantity. Access to this wealth of data furthermore permits new analyses and opportunities beyond the identification and quantitation of proteins. By combining high throughput computation and mass spectrometry instrumentation, we can generate the most complete snapshot of an organism's phenotype. This chapter will provide an overview of the complexity challenges facing those that study the proteome, as well as instrumentation and data analysis tools that are currently used to surmount them.

Biological flow of information

Genetic information classically flows from DNA to RNA to protein transmitting the data needed to maintain life as described by the central dogma of biology. The sequencing of the human genome and the subsequent rapid sequencing of nearly 15,000 additional organisms permitted the use of mass spectrometry experiments to rapidly probe their proteomes[4,5].

The human genome contains approximately 20,000 genes with most estimates ranging from 19,000-22,000[6]. This was a surprisingly small number at the time as estimates prior to

the completion of the human genome proposed ~ 100,000 genes[7]. However, further genome sequencing revealed the number of genes in an organism only weakly correlates with its complexity. For example, many plants contain significantly more genes than humans[8]. Despite this surprising lack of genes, humans have exceptionally large genomes with 3.2E9 base pairs contained within 23 chromosomes[9]. Much of the genome actually contains non-coding repeats (>50%) and regions which regulate genes through non-coding RNA or other regulatory elements[4].

Sequencing the genome of an organism provides information on all of the proteins that it is able to synthesize and any mutations that may be present. However, simply because the sequence of a protein is present in the genome does not mean that the gene is transcribed and translated into a functional protein. While mass spectrometry can be used to measure proteins, RNA can also be directly measured with RNAseq and transcriptomics. This gives a measurement of which genes are being transcribed. Transcriptomics and proteomics have shown surprisingly poor correlation suggesting that substantial regulation occurs at the translational level[10–14]. Some of this discrepancy is explained by many mRNAs undergoing additional regulation, with not all initiating translation and others serving unique regulatory roles[15,16]. Proteomics allows the direct quantitation of the molecular machines that carry out the business of life. For this reason it can be a valuable complement to RNAseq experiments, and also illuminating on its own.

Proteomics faces additional technical challenges compared to the sequencing of nucleotides. The structure of proteins contains much greater chemical diversity and thus are

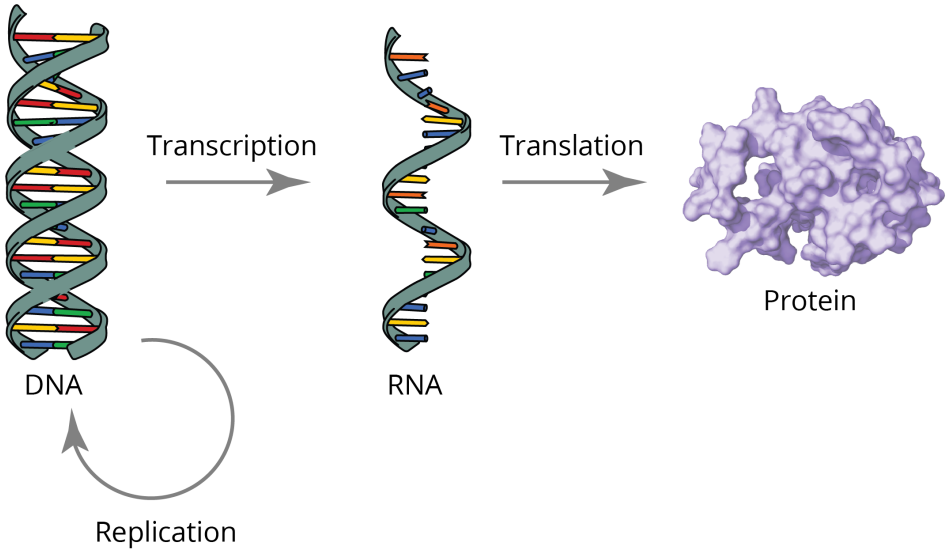


Figure 1.1: Central dogma of biology. Genetic information classically flows from DNA to RNA to protein transmitting the data needed to maintain life as described by the central dogma of biology.

more difficult to sequence than DNA and RNA. Additionally, no method currently exists to amplify low abundance proteins in the same way that PCR can be used with DNA. For these reasons proteomics is challenging, but remains a valuable way to capture important physiological changes.

Proteome Complexity

Genes that are transcribed into RNA and translated into protein make up the proteome and the total cohort of molecules that can be potentially measured with protein technologies. Of the ~ 20,000 human genes, evidence for ~ 17,000-18,000 proteins has been observed[17,18]. Despite the relatively small number of genes in the human genome, a number of factors combine to exponentially increase the number of proteoforms, or different protein forms produced from the genome, which can be detected with the mass spectrometer[19,20]. First, single amino acid polymorphisms (SAPs) arise from canonical differences in gene sequence that result in the replacement of one codon with another. These SAPs are derived from single nucleotide polymorphisms (SNPs) in the genome although many more SNPs exist that do not contribute to an amino acid substitution and thus are invisible to proteomic methods. It is estimated that the human genome contains 10 million SNPs which contribute to the vast diversity we observe in the human population[21]. Of these 10 million SNPs, approximately 1.3 million result in an amino acid substitution in a coding region[22].

The second contributor to proteoform heterogeneity is alternative splicing. Transcribed RNA in eukaryotes consists of regions of introns and exons, and during splicing introns

are removed and exons are combined to create mature mRNA. In many cases splicing can be used to create a range of unique proteins by altering exon composition from the same mRNA[23]. Alternative splicing occurs in approximately 50% of eukaryotic proteins and contributes ~ 2.8 isoforms per protein[24–26]. While splicing was initially thought to be a potential explanation for the relative dearth of genes in the human genome, proteomic evidence suggests that most genes are only present in a single form[27,28]. Studies that have searched for alternative isoforms have mainly identified hundreds of alternative isoforms with one study observing 2,000[27–29]. This is especially surprising as transcriptomics has identified thousands of splice variants which continue to be unobserved at the protein level[?]. The purpose of these alternative transcripts that are not translated in observable quantities remains unanswered. Some may be only expressed in special circumstances, tissues, or may have a different purpose than generating a protein product[28,30]. While splicing clearly contributes to the complexity of the proteome, it does not lead to orders of magnitude increase in protein diversity.

Post-translational modifications layer another type of diversity to the proteome following translation. Phosphorylation, ubiquitination, acetylation, and glycosylation are some of the more well studied, although more than 300 different modifications have been observed[31,32]. These modifications are frequently used to activate a protein or otherwise alter its activity in the form of localization, protein-protein interactions, or a signal for degradation. A full 5% of the genome is dedicated to enzymes that post-translationally modify other proteins[33]. With these additions to the ~ 20,000 genes, the number of predicted proteoforms rapidly explodes

with estimates ranging between 100,000-1,000,000[24,34,35]. This leviathan of proteoforms requires our utmost technological prowess in order to effectively parse.

Clearly the proteome is an intricate collection with seemingly endless variation and complexity. Each proteoform in a system may then be expressed at a different abundance in the cell. In a human cell line, protein abundances vary between one and ten billion copies, or seven orders of magnitude[36,37]. Assuming that the most abundant proteins are the most likely to be identified, this complicates identification of the many low abundance proteins in a system. In certain cell types and tissues this problem is compounded further. For example, in plasma the variation spans ten orders of magnitude, and in red blood cells 92% of the protein dry weight is hemoglobin[38,39]. However, the low abundance proteins in these systems often pose the most interesting biological questions necessitating the development of new methods to manage complexity. For some tissues such as plasma, specialized depletion columns exist to remove the most abundant proteins from the sample[40]. Tackling these challenges that arise with difficult biological materials is necessary in order to reap the greatest benefits from recent technical advancements in mass spectrometry.

Protein Preparation for Mass Spectrometry

Many other factors can determine the functional activity of a system besides protein abundance. Proteins may be further modified with the addition of post-translational modifications (PTMs), such as with the addition of a phospho group to activate a kinase cascade. In this case, simply monitoring the protein abundance would fail to capture the important

change that is occurring. Other 'omic' tools such as phosphoproteomics, metabolomics, or lipidomics can measure these changes with the goal to directly monitor the relevant changes that are taking place. For these reasons when designing an experiment it is essential to carefully consider the question that is posed and choose the best tools and methods to address it.

In order to access the proteome for analysis, cells of the organism of interest must first be lysed. The goal of this process is to equally extract all protein classes, while minimizing protein degradation by non-specific proteases. This can be accomplished with the use of physical forces such as bead-beating, milling, sonication, or French press. Otherwise chemical based techniques can be used to disrupt the cell membrane, such as chaotropes or detergents. Care must be taken with the selection of detergents to solubilize proteins as many commonly used detergents are not compatible with mass spectrometry. Some such as digitonin, sodium deoxycholate, sodium dodecyl sulfate, and RapiGest are removable by precipitation or filter aided sample preparation (FASP)[41–43]. Protease and phosphatase inhibitors may be used to prevent off-target cleavage[44]. Following denaturing, reduction agents are added to reduce disulfide bonds as well as alkylating reagents such as iodoacetamide or chloroacetamide to prevent their reformation.

In order to profile the greatest number of proteins, most experiments opt to digest proteins into peptides. This is referred to as a 'bottom-up' approach in contrast to 'top-down' mass spectrometry in which intact proteins are analyzed. Top-down proteomics has the advantage of enabling the identification of all concomitant modifications present on a

proteoform[45]. However, bottom-up proteomics results in a more experimentally tractable analysis and is best suited for the comprehensive profiling of large numbers of proteins. Proteins in a sample are digested into an average of 61 peptides per protein significantly increasing the complexity of the mixture[46].

Trypsin which cleaves at the C-terminus of lysine and arginine residues, is one of the most commonly used enzymes . It has been modified to avoid self-degradation, is highly aggressive, and stable under a wide range of conditions[47,48]. The selection of an enzyme for digestion will determine what peptides are generated and consequently available for sequencing. Typically, peptides from 7-35 amino acids are able to be successfully sequenced by the mass spectrometer[49]. An *in silico* tryptic digest of the human proteome indicates peptides will have an average length of nine residues and at least one basic residue making them well suited to MS analysis[50]. If an experiment is directed to identify a particular PTM it is important to consider whether that site occurs on a tryptic peptide. The use of alternative proteases has been shown to increase sequence coverage and can access protein regions which may lack lysines and arginines[49,51]. Following protease digestion, salts are removed from the sample typically using solid phase extraction[52].

If the study is designed to investigate PTMs, enrichment techniques are often used at this point of the sample prep[53,54]. Many PTMs only exist in low abundance in the proteome and are thus not easily observed. Charge state, structure, or other physical properties may be used to enrich for modifications. For example, antibodies have been developed for many modifications such as ubiquitination, acetylation, and arginine methylation which rely

on immunoaffinity[55–57]. Phosphorylation enrichments often employ an interaction with the negative charge on the phospho group such as metal based techniques with titanium or iron[58]. The carbohydrate binding proteins lectins can be used to enrich for a range of glycosylated proteins[59]. Thanks to these and other enrichment techniques the range of PTMs accessible to analysis by mass spectrometry is rapidly blossoming. Future analyses of under-studied and low-abundance PTMs will rely on the development of new and improved enrichment methods.

The simultaneous analysis of all peptides in a sample would generate spectra with overwhelming complexity. For this reason it is necessary to employ a chromatographic separation prior to injection of peptides on the instrument. The coupling of reverse phase (RP) separation with MS forming an LC-MS system is currently one of the most widespread practices[60]. Peptides are bound to the column based on their hydrophobicity and can be eluted over a gradient by increasing the percentage of organic solvent. The quality of separations has recently come under increased scrutiny as a critical component for generating quality results[61,62]. While the ability to rapidly sequence peptides has previously been limited by the scan speed of the mass spectrometer, new leaps in instrumental technology mean this is no longer the constraining factor[63]. Improvements in peptide separations results in increased signal and decreased peak width allowing more peptides to be separated and sequenced by the instrument[64].

For many applications, injection of a single complex peptide mixture combined with LC-MS is adequate to achieve the desired proteomic depth and sequence coverage. When

improved protein identifications are desired, orthogonal separations can be employed to further reduce sample complexity. The sequential combination of RP and strong cation exchange (SCX) particles in multi-dimensional protein identification technology (MuDPIT) gave a dramatic boost to the number of peptides that could be identified in one experiment[65]. This was a significant improvement over 2D gel electrophoresis experiments in which protein containing bands were excised and analyzed[66,67]. Today, both on-line and off-line separation techniques are used to reduce sample complexity prior to injection on an LC-MS. SCX and high pH reverse phase separations are used to partition peptides into multiple samples or fractions which are then analyzed separately on the instrument[68–70]. These techniques reduce sample complexity and increase analysis time resulting in a significant increase in protein identifications.

The interface between the LC and the MS is an integral component in order to transition peptides from the liquid phase to the gas phase for analysis. Electrospray ionization and matrix assisted laser desorption/ionization (MALDI) are two of the most common ways to get ions into the gas phase. John Fenn and Koichi Tanaka were honored for these developments with the receipt of the Nobel Prize for the “development of soft ionization methods for the analysis of biological molecules" in 2002. These methods paved the way for analysis of biomolecules and allowed for the subsequent rapid development of the field of proteomics[71,72].

Mass Spectrometry Instrumentation

Mass spectrometry measures the mass to charge (m/z) of gas phase ions enabling their identification through accurate mass measurements. For this reason, almost any material that can be ionized can be analyzed with a mass spectrometer. Various mass analyzers can be implemented to measure peptides including the quadrupole mass filter, quadrupole linear ion trap (LTQ), time-of-flight (TOF), Orbitrap, and Fourier-transform ion cyclotron resonance (FT-ICR)[73]. Mass analyzers employ different physical properties of ions in order to identify the mass-to-charge. These analyzers can be combined in numerous combinations to create a variety of hybrid and tribrid instruments[74,75].

Typical proteomics experiments employ a two part tandem MS method. First, an initial survey scan is taken to identify the mass to charge of the peptide. Then, a peptide is selected for sequencing and isolated. It is then subjected to fragmentation, and analyzed with a second MS scan (MS2 or MS/MS). This scan will determine the mass to charge of fragment ions, and allow the identity of the peptide to be determined. The sequence of the peptide can be solved by the difference in mass to charge of the peaks in the MS2. These differences correspond to the mass of an amino acid and allow the peptide sequence to be read in the mass of the fragments.

Peptides can be fragmented using a variety of disassociation techniques depending on the type of analysis that is being performed. Collisional induced disassociation (CID) and higher-energy C-trap disassociation (HCD) generate b and y ions and are widely

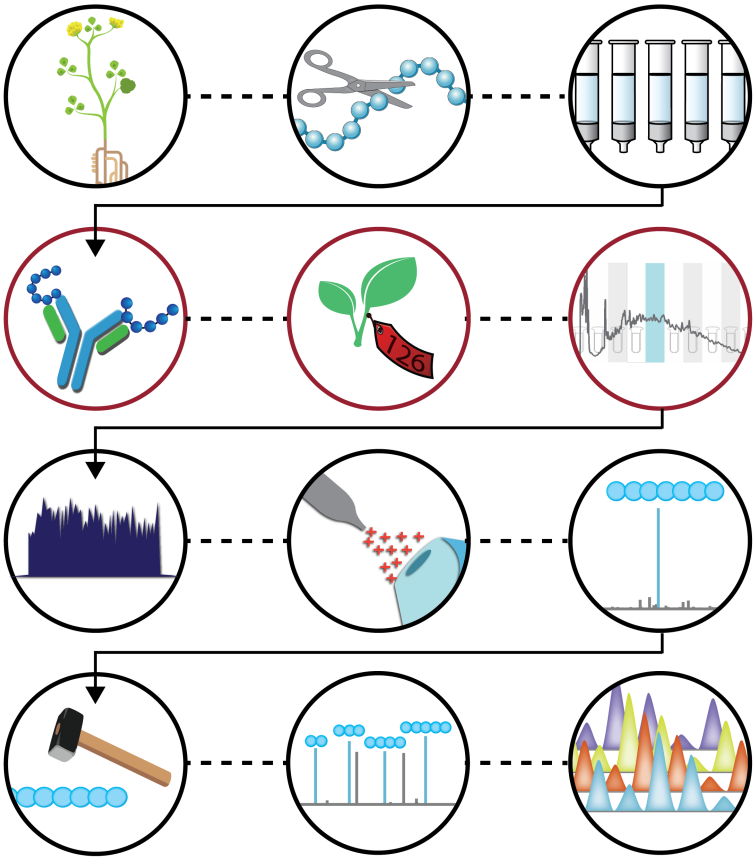


Figure 1.2: Peptide sample preparation for mass spectrometry. For a bottom-up workflow, tissues or cells are digested into peptides and de-salted. Optional steps depicted in red include PTM enrichment, labeling for quantification, and pre-fractionation. Peptides are then separated with chromatographic separations, ionized to the gas phase, and injected into the mass spectrometer. Survey scans are then collected, followed by peptide fragmentation, MS2 analysis, and quantification.

used for peptide sequencing (Michalski et al., 2012). Electron transfer disassociation (ETD) fragments peptides along the amide bond producing c and z ions[76]. ETD is also prevalent, especially in the study of post-translational modifications[77,78].

Once peptides have been fragmented, tandem MS/MS scans can be collected using data dependent acquisition (DDA) mode or data independent acquisition (DIA). Using DIA, successive mass to charge windows are isolated and all of the peptides contained are fragmented together. This allows the potential identification of more peptide species in a spectrum and has experienced a surge of popularity in recent years[79,80]. DDA intentionally selects peptides from a spectrum for sequencing, usually based on their relative abundance. Peptides that have been sequenced are then excluded from repeat analysis for a set time period. Otherwise preset mass to charge values can be analyzed for a targeted analysis[81]. Targeted methods afford a high level of reproducibility and sensitivity with data acquisition. Survey and tandem scans may be collected successively in the same mass analyzer or concurrently in separate analyzers to maximize collection speed. The acquisition of survey and tandem scans continues throughout the gradient elution which commonly ranges from one to three hours.

A routine peptide analysis can easily result in the acquisition of over 150,000 tandem MS spectra. The analysis of this wealth of data clearly requires the use of high throughput analysis and computation. These scans can be sequenced *de novo* where peptide sequences are determined directly from the mass spectra, however this is typically limited to niche applications[82]. Most often the spectra are searched against a database containing all potential

proteins that are expected to be found in the sample. This list of proteins is digested *in silico* with the appropriate enzyme to generate a list of peptides that the spectra are compared to. Each match is given a score and filtered to output a list of final peptide identifications[83]. To segregate true positive identifications from false positives, false discovery rate (FDR) correction is employed and allows the number of false positives to be limited to a known quantity[84,85]. A decoy database is usually constructed by reversing all the protein sequences creating false identifications which are also included when searching the data[86,87]. The scores of all peptides are then ranked and a threshold can be set which minimizes the number of false positives while retaining the maximum number of true identifications. Typically this threshold is set at 1% false positives. This method is effective because a false identification has an equal probability of matching to the forward database as the reverse database.

Peptide matches must then be assembled into proteins as the final output for the search. This can be complex as the sequence of many peptides may be present in multiple proteins making it impossible to determine from which protein they originated. To circumvent this, peptides are grouped according to the principle of parsimony, or Occam's razor[88]. Using this principle, the minimum number of proteins necessary to account for the observed peptides are assumed to be present in the sample. When insufficient data is present to distinguish two proteins, they are reported together as a single protein group.

Quantitation

To turn qualitative identifications into quantitative values numerous techniques can be employed depending on the type of experiment. It is important to note that all of these quantitation methods only provide relative quantitation between samples. Measuring absolute quantitative values is possible, but requires the addition of isotopically labeled internal standards and the use of a standard curve for high quality measurements[89]. For these reasons it is usually not done on the proteome scale.

Metabolic and isobaric labelling both rely on the addition of heavy isotopes into the sample[90]. One key difference is the time during sample preparation at which the labels are introduced. Isobaric or isotopic tags may be added to peptides once they are digested. When using metabolic labeling, the samples are grown with the addition of heavy label, often in the form of an amino acid, in the cell culture or diet. Metabolic and isotopic labelling both permit sample multiplexing[91]. This can increase reproducibility as numerous samples are analyzed together and also reduces instrument time.

Incorporating the label early in sample preparation allows the samples to be combined early and minimizes variation in sample preparation. Stable isotope labeling by amino acids in cell culture (SILAC) incorporates stable isotope containing amino acids in newly synthesized proteins[56]. When combined with 'light' cell populations, the peptides containing labeled amino acids remain distinguishable in the mass spectrometer and can be used for relative quantitation. This approach can also be applied as the stable isotope labeling of

mammals (SILAM), although creating a heavy labeled food source for mammals poses a greater technical challenge[92]. Additionally, obtaining complete labeling for some tissues in mouse requires multiple generations and thus can be almost prohibitively expensive[93].

When analyzing samples generated using a SILAC or SILAM method the generated mass spectra will appear significantly more complex. This is caused by the fact that two isotopic clusters are present for each peptide instead of one. Each copy of the labeled peptide can be potentially subjected to fragmentation and tandem MS introducing a layer of redundancy to data acquisition. It is likely that fewer total peptides and proteins will be identified due to superfluous sampling of isotopically labeled peptides[94]. Still, these methods are some of the most reliable at eliminating technical variation and provide accurate relative quantitation.

The other primary option for sample labeling is isotopic tagging. These methods may result in peptides with different masses as previously described, or the incorporated labels may be isobaric and chromatographically indistinguishable. When using isobaric tags, once the labeled peptides undergo fragmentation, reporter ions of unique masses are released allowing for quantitation. This circumvents the problem described above as each peptide appears as a single peak in the survey MS scan[90]. However, since the label is introduced following the digestion of protein into peptides, there is greater opportunity for variation to be introduced between samples. Options such as isobaric tags for relative and absolute quantitation (iTRAQ), tandem mass tags (TMT), and others allow for multiplexing up to 11 samples[95,96]. These techniques can be used in conjunction with other analytical methods,

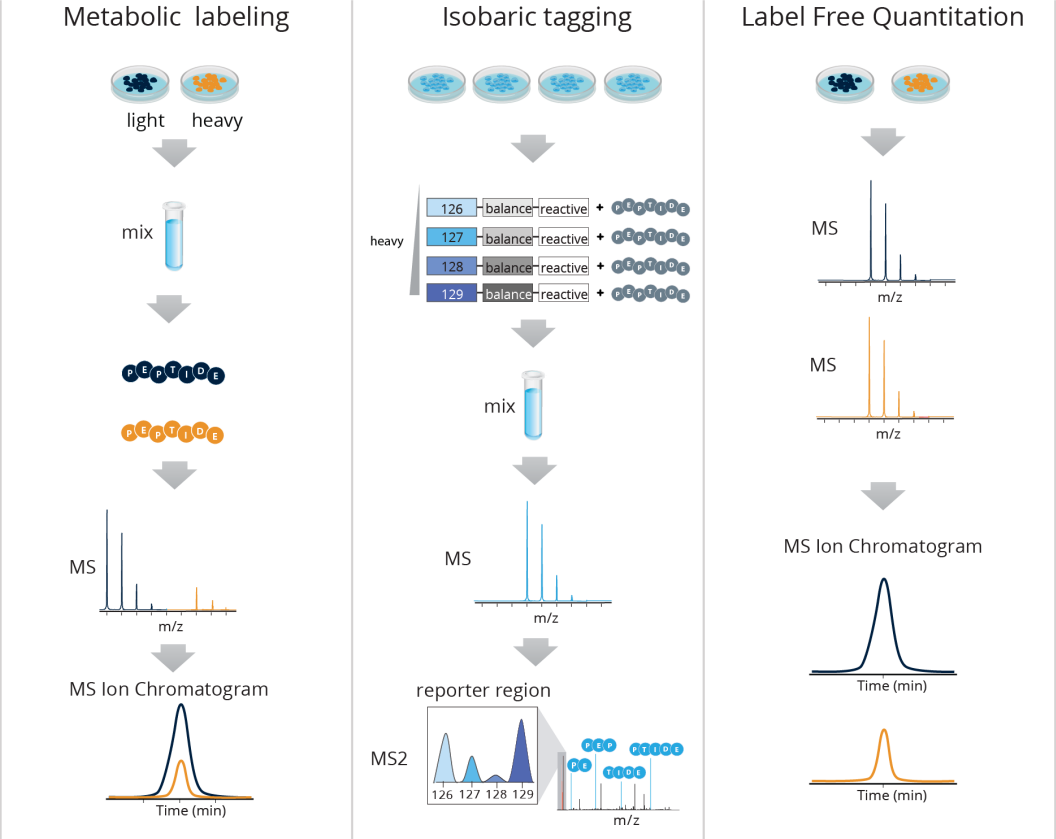


Figure 1.3: Methods for quantitative mass spectrometry. Metabolic, isobaric, and label free quantitation are three main options to obtain quantitative results using mass spectrometry.

such as enrichment for PTMs.

Experimenters can also opt to avoid sample labeling entirely instead choosing a label-free method for quantitation. In this method, peak area from the survey scans are integrated across retention time to generate peak intensities[97]. Label free data can also be acquired with the use of spectral counting or sequential window acquisition of all theoretical mass spectra (SWATH)[98,99]. These methods have the advantage of obvious ease of sample preparation as no additional labeling steps are required. Additionally with proper technique, excellent reproducibility can be still be achieved between samples[100]. This method has been successfully applied to very large scale experiments and with proper batch controls and replicates can facilitate broad comparison across many samples[101–103].

Data Integration and Analysis

Once quantitative proteomic data has been acquired, significant analysis remains in order to interpret and synthesize results. This task quickly grows in size when multiple types of 'omic' or phenotype data are available. Depending on the experiment, proteomic data may be integrated with genomic data for large scale quantitative trait loci (QTL) mapping studies, assembled into an interactome, or another analysis that requires additional computation[104,105]. Now that proteomics has far surpassed the analysis of hundreds of proteins, high throughput computing is a necessity to extract meaning from data sets. To meet these needs, statistical programming languages, such as R, are increasingly used for data analysis. As a result, numerous packages have been developed for public use that

focus on the analysis of data produced by mass spectrometry[106,107].

Another key point for collecting informative data is proper experimental design to address the questions that are being asked. Of the experimental methods that have been described heretofore, different choices would be made depending on the type of biology being investigated. Two broad classes of experiments that can be done with mass spectrometry are hypothesis driven and hypothesis generating. Proteomics is inherently adept at hypothesis generating as great quantities of data are produced in an unbiased fashion. These types of experiments encourage broadening research interests to proteins implicated in new biology. Historically, scientists have continually studied a surprisingly consistent set of proteins despite new proteins continually implicated in disease[108]. However, this benefit can easily become paralyzing as large quantities of data are difficult to parse, and forming informed hypotheses from data requires expertise in the relevant biology. For this reason, many successful projects are born from collaborations between mass spectrometry and biological groups. It is crucial that all members of the team are active in conceptualizing the experiment to ensure that both technical and biological controls are in place. Hypothesis driven experiments often benefit from more intentional experimental design, either in the use of a targeted method, or special considerations with sample preparation. Without these considerations, it is easy for valuable time to be wasted collecting data that cannot address the question. As the interdisciplinary nature of science continues to grow, clear and concise communication will always be needed to reap the benefits.

Overview of Projects

The chapters included herein demonstrate the versatility of mass spectrometry based proteomics to address diverse biological challenges. Often material for analysis poses additional complications due to its complexity or high dynamic range. **Chapter 2** outlines an example of this problem in which proteomic technologies are applied to clinical red blood samples collected from a cohort of twins. With this unique sample set, protein concentration heritabilities were calculated and integrated with metabolite concentration heritabilities. This data set was used to generate a model to understand blood metabolism in storage and predict its transfusion efficacy. **Chapter 3** focuses on the model organism *Saccharomyces cerevisiae* and specifically its response to a zinc deficient environment. Chapters **2** and **3** both implement label-free quantitation to compare relative protein abundances. In **Chapter 4** we turn our attention to a symbiotic system composed of the nitrogen fixing bacterium *Sinorhizobium meliloti* and its host, the model legume, *Medicago truncatula*. Post-translation modifications play an important role in symbiosis, and this chapter focuses on the role of ubiquitination. To quantify ubiquitination, we rely on enrichment strategies in conjunction with isobaric labeling. For **Chapter 5**, we employ heavy labeled amino acids in a SILAM analogous method to calculate protein half-lives on a large-scale in a mouse model. Nine mouse tissues were analyzed with a computational pipeline to measure half-lives. This dataset was subsequently examined to determine physical and biological factors which may contribute to protein half-life.

References

[1] J. J. Coon, A. S. Hebert, D. J. Bailey, A. Ulbrich, M. S. Westphall, E. E. Coughlin, and A. L. Richards, "The One Hour Yeast Proteome," *Molecular & Cellular Proteomics*, vol. 13, pp. 339–347, jan 2013.

[2] N. A. Kulak, G. Pichler, I. Paron, N. Nagaraj, and M. Mann, "Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells," *Nature Methods*, vol. 11, pp. 319–324, mar 2014.

[3] A. Schmidt, K. Kochanowski, S. Vedelaar, E. Ahrné, B. Volkmer, L. Callipo, K. Knoops, M. Bauer, R. Aebersold, and M. Heinemann, "The quantitative and condition-dependent Escherichia coli proteome," *Nature Biotechnology*, vol. 34, pp. 104–110, jan 2016.

[4] M. Frazier, R. A. Gibbs, D. M. Muzny, S. E. Scherer, J. B. Bouck, E. J. Sodergren, K. C. Worley, C. M. Rives, J. H. Gorrell, M. L. Metzker, W. FitzHugh, S. L. Naylor, R. S. Kucherlapati, D. L. Nelson, G. M. Weinstock, Y. Sakaki, A. Fujiyama, M. Hattori, T. Yada, A. Toyoda, T. Itoh, R. Funke, C. Kawagoe, H. Watanabe, Y. Totoki, T. Taylor, J. Weissenbach, R. Heilig, W. Saurin, F. Artiguenave, P. Brottier, T. Bruls, D. Gage, E. Pelletier, C. Robert, P. Wincker, D. R. Smith, L. Doucette-Stamm, M. Rubenfield, K. Weinstock, H. M. Lee, J. Dubois, A. Rosenthal, K. Harris, M. Platzer, G. Nyakatura, S. Taudien, A. Rump, H. Yang, J. Yu, J. Wang, G. Huang, J. Gu, L. Hood, A. Heaford, L. Rowen, A. Madan, S. Qin, R. W. Davis, N. A. Federspiel, A. P. Abola, M. J. Proctor, R. M. Myers, J. Schmutz, M. Dickson, J. Howland, J. Grimwood, D. R. Cox, M. V. Olson, R. Kaul, C. Raymond, N. Shimizu, K. Kawasaki, S. Minoshima, G. A. Evans, M. Athanasiou, L. Kann, R. Schultz, B. A. Roe, F. Chen, H. Pan, J. Ramser, H. Lehrach, R. Reinhardt, W. R. McCombie, M. de la Bastide, N. Dedhia, J. Lehoczky, H. BlÃ¶cker, K. Hornischer, G. Nordsiek, R. Agarwala, L. Aravind, J. A. Bailey, A. Bateman, S. Batzoglou, E. Birney, P. Bork, R. LeVine, D. G. Brown, C. B. Burge, L. Cerutti, H. C. Chen, D. Church, M. Clamp, R. R. Copley, T. Doerks, S. R. Eddy, E. E. Eichler, P. McEwan, E. S. Lander, T. S. Furey, J. Galagan, J. G. Gilbert, C. Harmon, Y. Hayashizaki, D. Haussler, H. Hermjakob, K. Hokamp, W. Jang, L. S. Johnson, K. McKernan, T. A. Jones, S. Kasif, A. Kaspryzk, S. Kennedy, W. J. Kent, P. Kitts, E. V. Koonin, I. Korf, D. Kulp, D. Lancet, J. Meldrim, T. M. Lowe, A. McLysaght, T. Mikkelsen, J. V. Moran, N. Mulder, V. J. Pollara, C. P. Ponting, G. Schuler, J. Schultz, G. Slater, J. P. Mesirov, A. F. Smit, E. Stupka, J. Szustakowski, D. Thierry-Mieg, J. Thierry-Mieg, L. Wagner, J. Wallis, R. Wheeler, A. Williams, Y. I. Wolf, C. Miranda, K. H. Wolfe, S. P. Yang, R. F. Yeh, F. Collins, M. S. Guyer, J. Peterson, A. Felsenfeld, K. A. Wetterstrand, A. Patrinos, M. J. Morgan, W. Morris, P. de Jong, J. J. Catanese, K. Osoegawa, H. Shizuya, S. Choi, Y. J. Chen, J. Szustakowki, J. Naylor, M. Rosetti, R. Santos, A. Sheridan, L. M. Linton, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman,

J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, B. Birren, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, C. Nusbaum, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, M. C. Zody, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. Baldwin, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chissoe, M. C. Wendl, K. D. Delehaunty, K. Devon, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, T. Hawkins, E. Branscomb, K. Dewar, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J. F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, and M. Doyle, "Initial sequencing and analysis of the human genome.," *Nature*, vol. 409, pp. 860–921, feb 2001.

[5] I. V. Grigoriev, S. Richards, M. M. Goldstein, P. S. Soltis, D. Haussler, J. Coddington, K. J. Hackett, J. C. Castilla-Rubio, G. Zhang, M. T. P. Gilbert, R. Durbin, A. Patrinos, H. Yang, W. E. Johnson, X. Xu, H. A. Lewin, E. D. Jarvis, S. V. Edwards, W. J. Kress, M.-A. van Sluys, K. A. Crandall, G. E. Robinson, F. Forest, and W. J. Baker, "Earth BioGenome Project: Sequencing life for the future of life," *Proceedings of the National Academy of Sciences*, vol. 115, pp. 4325–4333, apr 2018.

[6] J. Harrow, A. Frankish, J. M. Gonzalez, E. Tapanari, M. Diekhans, F. Kokocinski, B. L. Aken, D. Barrell, A. Zadissa, S. Searle, I. Barnes, A. Bignell, V. Boychenko, T. Hunt, M. Kay, G. Mukherjee, J. Rajan, G. Despacio-Reyes, G. Saunders, C. Steward, R. Harte, M. Lin, C. Howald, A. Tanzer, T. Derrien, J. Chrast, N. Walters, S. Balasubramanian, B. Pei, M. Tress, J. M. Rodriguez, I. Ezkurdia, J. Van Baren, M. Brent, D. Haussler, M. Kellis, A. Valencia, A. Reymond, M. Gerstein, R. Guigó, and T. J. Hubbard, "GENCODE: The reference human genome annotation for the ENCODE project," *Genome Research*, vol. 22, pp. 1760–1774, sep 2012.

[7] E. Pennisi, "ENCODE project writes eulogy for junk DNA," *Science*, vol. 337, pp. 1159–1161, sep 2012.

[8] E. Schad, P. Tompa, and H. Hegyi, "The relationship between proteome size, structural disorder and organism complexity," *Genome Biology*, vol. 12, p. R120, dec 2011.

[9] W. Makałowski, "The human genome structure and organization," *Acta Biochim. Pol.*, vol. 48, no. 3, pp. 587–598, 2001.

[10] G. Chen, T. G. Gharib, C.-C. Huang, J. M. G. Taylor, D. E. Misek, S. L. R. Kardia, T. J. Giordano, M. D. Iannettoni, M. B. Orringer, S. M. Hanash, and D. G. Beer, "Discordant Protein and mRNA Expression in Lung Adenocarcinomas," *Molecular & Cellular Proteomics*, vol. 1, pp. 304–313, apr 2002.

[11] P.-Z. Wen, V. A. Petyuk, L. Orozco, K. Weitz, C. C. Park, A. J. Lusis, R. Yordanova, R. Hagopian, J. Sinsheimer, P. Gargalovic, H. M. Kang, C. R. Farber, A. Ghazalpour, R. D. Smith, I. Neuhaus, I. N. Mungrue, D. J. Smith, C. Pan, N. Furlotte, T. Kirchgessner, B. Bennett, H. Brewer, C. Tilford, E. Eskin, N. Siemers, and D. G. Camp, "Comparative Analysis of Proteome and Transcriptome Variation in Mouse," *PLoS Genetics*, vol. 7, p. e1001393, jun 2011.

[12] S. P. Gygi, Y. Rochon, B. R. Franza, and R. Aebersold, "Correlation between Protein and mRNA Abundance in Yeast," *Molecular and Cellular Biology*, vol. 19, pp. 1720–1730, mar 2015.

[13] L. E. Pascal, L. D. True, D. S. Campbell, E. W. Deutsch, M. Risk, I. M. Coleman, L. J. Eichner, P. S. Nelson, and A. Y. Liu, "Correlation of mRNA and protein levels: Cell type-specific gene expression of cluster designation antigens in the prostate," *BMC Genomics*, vol. 9, p. 246, may 2008.

[14] E. S. Yeung, "Genome-wide correlation between mRNA and protein in a single cell," *Angewandte Chemie - International Edition*, vol. 50, pp. 583–585, jan 2011.

[15] K. V. Morris and J. S. Mattick, "The rise of regulatory RNA," *Nature Reviews Genetics*, vol. 15, pp. 423–437, jun 2014.

[16] N. Sonenberg and A. G. Hinnebusch, "Regulation of Translation Initiation in Eukaryotes: Mechanisms and Biological Targets," *Cell*, vol. 136, pp. 731–745, feb 2009.

[17] R. Sirdeshmukh, K. Mudgal, K. R. Murthy, D. S. Kelkar, C. A. Iacobuzio-Donahue, L. Balakrishnan, A. H. Patil, A. A. Khan, R. S. Nirujogi, S. Jain, S. M. Pinto, H. Gowda, R. Chaerkady, P. Leal-Rojas, P. G. Shaw, A. Radhakrishnan, A. Maitra, J. Zhong, M.-S. Kim, C. J. Mitchell, A. Sahu, P. Kumar, J. Advani, T.-C. Huang, T. S. K. Prasad, D. Freed, G. Dey, T. Subbannayya, S. S. K. Shankar, M. K. Halushka, S. Ahmad, P. Rajagopalan, S. D. Leach, J. T. Schroeder, R. H. Hruban, P. Satishchandra, A. Pandey, S. Renuse, K. K. Mukherjee, C. L. Kerr, X. Wu, A. Chatterjee, K. K. Datta, S. S. K. Shankar, R. Goel, B. George, L. D. N. Selvan, A. Marimuthu, N. A. Sahasrabuddhe, R. Isserlin, A. K. Madugundu, S. Jayaram, B. Muthusamy, H. Lam, G. D. Bader, R. Raju, D. Getnet, N. Syed, S. S. Manda, M. S. Zahari, S. K. Sreenivasamurthy, S. Chavan, J. K. Thomas, Y. Subbannayya, G. J. Sathe, S. Prasad, J. Sharma, A. Mahadevan, M. Kumar, V. Nanjappa, S. D. Yelamanchi, and C. G. Drake, "A draft map of the human proteome," *Nature*, vol. 509, pp. 575–581, may 2014.

[18] M. Wilhelm, J. Schlegl, H. Hahne, A. M. Gholami, M. Lieberenz, M. M. Savitski, E. Ziegler, L. Butzmann, S. Gessulat, H. Marx, T. Mathieson, S. Lemeer, K. Schnatbaum, U. Reimer, H. Wenschuh, M. Mollenhauer, J. Slotta-Huspenina, J. H. Boese, M. Bantscheff, A. Gerstmair, F. Faerber, and B. Kuster, "Mass-spectrometry-based draft of the human proteome," *Nature*, vol. 509, pp. 582–587, may 2014.

[19] L. M. Smith and N. L. Kelleher, "Proteoforms as the next proteomics currency," *Science*, vol. 359, pp. 1106–1107, mar 2018.

[20] L. M. Smith and N. L. Kelleher, "Proteoform: A single term describing protein complexity," *Nature Methods*, vol. 10, pp. 186–187, mar 2013.

[21] M. Cargill, D. Altshuler, J. Ireland, P. Sklar, K. Ardlie, N. Patil, C. R. Lane, E. P. Lim, N. Kalyanaraman, J. Nemesh, L. Ziaugra, L. Friedland, A. Rolfe, J. Warrington, R. Lipshutz, G. Q. Daley, and E. S. Lander, "Characterization of single-nucleotide polymorphisms in coding regions of human genes," *Nature Genetics*, vol. 22, pp. 231–238, jul 1999.

[22] C. Schaefer, A. Meier, B. Rost, and Y. Bromberg, "Snpdbe: Constructing an nsSnp functional impacts database," *Bioinformatics*, vol. 28, pp. 601–602, feb 2012.

[23] E. Park, Z. Pan, Z. Zhang, L. Lin, and Y. Xing, "The Expanding Landscape of Alternative Splicing Variation in Human Populations," *American Journal of Human Genetics*, vol. 102, pp. 11–26, jan 2018.

[24] A. T. Kopylov, V. G. Zgoda, E. V. Ilgisonis, E. A. Ponomarenko, A. V. Lisitsa, E. V. Poverennaya, M. A. Pyatnitskiy, and A. I. Archakov, "The Size of the Human Proteome: The Width and Depth," *International Journal of Analytical Chemistry*, vol. 2016, pp. 1–6, 2016.

[25] A. A. Sharov, D. B. Dudekula, and M. S. H. Ko, "Genome-wide assembly and analysis of alternative transcripts in mouse," *Genome Res.*, vol. 15, pp. 748–754, may 2005.

[26] J. Stetefeld and M. A. Ruegg, "Structural and functional diversity generated by alternative mRNA splicing," *Trends in Biochemical Sciences*, vol. 30, pp. 515–521, sep 2005.

[27] A. del Pozo, J. Harrow, M. L. Tress, A. Valencia, A. Frankish, J. M. Rodriguez, K. Ashman, and I. Ezkurdia, "Comparative Proteomics Reveals a Significant Bias Toward Alternative Protein Isoforms with Conserved Structure and Function," *Molecular Biology and Evolution*, vol. 29, pp. 2265–2283, sep 2012.

[28] M. L. Tress, F. Abascal, and A. Valencia, "Alternative Splicing May Not Be the Key to Proteome Complexity," *Trends in Biochemical Sciences*, vol. 42, no. 2, pp. 98–110, 2017.

[29] D. B. Bekker-Jensen, C. D. Kelstrup, T. S. Batth, S. C. Larsen, C. Haldrup, J. B. Bramsen, K. D. Sørensen, S. Høyer, T. F. Ørntoft, C. L. Andersen, M. L. Nielsen, and J. V. Olsen, "An Optimized Shotgun Strategy for the Rapid Generation of Comprehensive Human Proteomes.," *Cell systems*, vol. 4, pp. 587–599.e4, jun 2017.

[30] F. Mignone, C. Gissi, S. Liuni, and G. Pesole, "Untranslated regions of mRNAs.," *Genome biology*, vol. 3, no. 3, p. REVIEWS0004, 2002.

[31] J. Huang, F. Wang, M. Ye, and H. Zou, "Enrichment and separation techniques for large-scale proteomics analysis of the protein post-translational modifications," *Journal of Chromatography A*, vol. 1372, pp. 1–17, dec 2014.

[32] G. A. Khoury, R. C. Baliban, and C. A. Floudas, "Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database," *Scientific Reports*, vol. 1, p. 90, sep 2011.

[33] C. T. Walsh, S. Garneau-Tsodikova, and G. J. Gatto, "Protein posttranslational modifications: the chemistry of proteome diversifications," *Angew. Chem. Int. Ed. Engl.*, vol. 44, pp. 7342–7372, dec 2005.

[34] R. Aebersold, J. N. Agar, I. J. Amster, M. S. Baker, C. R. Bertozzi, E. S. Boja, C. E. Costello, B. F. Cravatt, C. Fenselau, B. A. Garcia, Y. Ge, J. Gunawardena, R. C. Hendrickson, P. J. Hergenrother, C. G. Huber, A. R. Ivanov, O. N. Jensen, M. C. Jewett, N. L. Kelleher, L. L. Kiessling, N. J. Krogan, M. R. Larsen, J. A. Loo, R. R. Ogorzalek Loo, E. Lundberg, M. J. MacCoss, P. Mallick, V. K. Mootha, M. Mrksich, T. W. Muir, S. M. Patrie, J. J. Pesavento, S. J. Pitteri, H. Rodriguez, A. Saghatelian, W. Sandoval, H. Schlüter, S. Sechi, S. A. Slavoff, L. M. Smith, M. P. Snyder, P. M. Thomas, M. Uhlén, J. E. Van Eyk, M. Vidal, D. R. Walt, F. M. White, E. R. Williams, T. Wohlschlager, V. H. Wysocki, N. A. Yates, N. L. Young, B. Zhang, R. R. O. Loo, E. Lundberg, M. J. MacCoss, P. Mallick, V. K. Mootha, M. Mrksich, T. W. Muir, S. M. Patrie, J. J. Pesavento, S. J. Pitteri, H. Rodriguez, A. Saghatelian, W. Sandoval, H. Schlüter, S. Sechi, S. A. Slavoff, L. M. Smith, M. P. Snyder, P. M. Thomas, M. Uhlén, J. E. Van Eyk, M. Vidal, D. R. Walt, F. M. White, E. R. Williams, T. Wohlschlager, V. H. Wysocki, N. A. Yates, N. L. Young, and B. Zhang, "How many human proteoforms are there?," *Nature Chemical Biology*, vol. 14, pp. 206–214, feb 2018.

[35] H. Berling, K. von Feilitzen, J. Nielsen, C. A.-K. Szigyarto, K. Edlund, B. M. Hallstrom, C. Kampf, E. Lundberg, M. Forsberg, F. Johansson, A. Asplund, P.-H. Edqvist, J. O. Takanen, L. Fagerberg, T. Alm, F. Ponten, P. Oksvold, A. Mardinoglu, E. Sjostedt, I. Olsson, J. Rockberg, G. von Heijne, J. Mulder, H. Tegel, L. Persson, M. Uhlen, M. Hamsten, M. Zwahlen, P. Nilsson, S. Hober, D. Djureinovic, J. Odeberg, A. Siverts-son, S. Navani, C. Lindskog, and J. M. Schwenk, "Tissue-based map of the human proteome," *Science*, vol. 347, pp. 1260419–1260419, jan 2015.

[36] M. Beck, A. Schmidt, J. Malmstroem, M. Claassen, A. Ori, A. Szymborska, F. Herzog, O. Rinner, J. Ellenberg, and R. Aebersold, "The quantitative proteome of a human cell line," *Mol. Syst. Biol.*, vol. 7, p. 549, nov 2011.

[37] R. A. Zubarev, "The challenge of the proteome dynamic range and its implications for in-depth proteomics," *PROTEOMICS*, vol. 13, pp. 723–726, mar 2013.

[38] D. Teupser, G. Pichler, N. A. Kulak, L. M. Holdt, M. Mann, and P. E. Geyer, "Plasma Proteome Profiling to Assess Human Health and Disease," *Cell Systems*, vol. 2, no. 3, pp. 185–195, 2016.

[39] A. M. Stadler, I. Digel, G. M. Artmann, J. P. Embs, G. Zaccai, and G. Büldt, "Hemoglobin dynamics in red blood cells: Correlation to body temperature," *Biophysical Journal*, vol. 95, pp. 5449–5461, dec 2008.

[40] P. A. Rudnick, R. J. C. Slebos, K. L. Cheek, M. Y. Martinez, C. Tu, S. E. Stein, and D. C. Liebler, "Depletion of Abundant Plasma Proteins and Limitations of Plasma Proteomics," *Journal of Proteome Research*, vol. 9, pp. 4982–4991, oct 2010.

[41] E. I. Chen, D. Cociorva, J. L. Norris, and J. R. Yates, "Optimization of mass spectrometry-compatible surfactants for shotgun proteomics," *Journal of Proteome Research*, vol. 6, pp. 2529–2538, jul 2007.

[42] Y. Lin, H. Lin, Z. Liu, K. Wang, and Y. Yan, "Improvement of a sample preparation method assisted by sodium deoxycholate for mass-spectrometry-based shotgun membrane proteomics," *Journal of Separation Science*, vol. 37, no. 22, pp. 3321–3329, 2014.

[43] J. R. Wiśniewski, A. Zougman, N. Nagaraj, and M. Mann, "Universal sample preparation method for proteome analysis," *Nature Methods*, vol. 6, pp. 359–362, may 2009.

[44] T. E. Thingholm, M. R. Larsen, C. R. Ingrell, M. Kassem, and O. N. Jensen, "TiO2-based phosphoproteomic analysis of the plasma membrane and the effects of phosphatase inhibitor treatment," *Journal of Proteome Research*, vol. 7, pp. 3304–3313, aug 2008.

[45] C. Boone and J. Adamec, "Top-Down Proteomics," *Proteomic Profiling and Analytical Chemistry: The Crossroads: Second Edition*, vol. 85, pp. 175–191, jul 2016.

[46] L. Martens, J. Vandekerckhove, and K. Gevaert, "DBToolkit: Processing protein databases for peptide-centric proteomics," *Bioinformatics*, vol. 21, pp. 3584–3585, jan 2005.

[47] J. V. Olsen, S.-E. Ong, and M. Mann, "Trypsin Cleaves Exclusively C-terminal to Arginine and Lysine Residues," *Molecular & Cellular Proteomics*, vol. 3, pp. 608–614, jun 2004.

[48] R. H. Rice, G. E. Means, and W. D. Brown, "Stabilization of bovine trypsin by reductive methylation," *BBA - Protein Structure*, vol. 492, pp. 316–321, jun 1977.

[49] D. L. Swaney, C. D. Wenger, and J. J. Coon, "Value of using multiple proteases for large-scale mass spectrometry-based proteomics," *Journal of Proteome Research*, vol. 9, pp. 1323–1329, mar 2010.

[50] E. Vandermarliere, M. Mueller, and L. Martens, "Getting intimate with trypsin, the leading protease in proteomics," *Mass Spectrometry Reviews*, vol. 32, no. 6, pp. 453–465, 2013.

[51] P. Giansanti, L. Tsiatsiani, T. Y. Low, and A. J. Heck, "Six alternative proteases for mass spectrometry-based proteomics beyond trypsin," *Nature Protocols*, vol. 11, pp. 993–1006, may 2016.

[52] J. Rappsilber, M. Mann, and Y. Ishihama, "Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips," *Nature Protocols*, vol. 2, no. 8, pp. 1896–1906, 2007.

[53] J. V. Olsen and M. Mann, "Status of Large-scale Analysis of Post-translational Modifications by Mass Spectrometry," *Molecular & Cellular Proteomics*, vol. 12, pp. 3444–3452, dec 2013.

[54] A. M. Silva, R. Vitorino, M. R. M. Domingues, C. M. Spickett, and P. Domingues, "Post-translational modifications and mass spectrometry detection," *Free Radical Biology and Medicine*, vol. 65, pp. 925–941, dec 2013.

[55] W. Kim, E. J. J. Bennett, E. L. L. Huttlin, A. Guo, J. Li, A. Possemato, M. E. E. Sowa, R. Rad, J. Rush, M. J. J. Comb, J. W. W. Harper, S. P. P. Gygi, A. Possemato, J. Rush, M. E. E. Sowa, E. L. L. Huttlin, W. Kim, J. Li, R. Rad, E. J. J. Bennett, A. Guo, J. W. W. Harper, S. P. P. Gygi, E. L. L. Huttlin, A. Guo, J. Li, A. Possemato, M. E. E. Sowa, R. Rad, J. Rush, M. J. J. Comb, J. W. W. Harper, and S. P. P. Gygi, "Systematic and quantitative assessment of the ubiquitin-modified proteome," *Molecular Cell*, vol. 44, pp. 325–340, oct 2011.

[56] S.-E. Ong and M. Mann, "A practical recipe for stable isotope labeling by amino acids in cell culture ({SILAC})," *Nature Protocols*, vol. 1, pp. 2650–2660, dec 2006.

[57] A. Kumar, B. T. Weinert, C. Choudhary, H.-S. Kim, C.-X. Deng, S. A. Wagner, and E. M. Sol, "Proteomic Investigations of Lysine Acetylation Identify Diverse Substrates of Mitochondrial Deacetylase Sirt3," *PLoS ONE*, vol. 7, p. e50545, dec 2012.

[58] N. M. Riley and J. J. Coon, "Phosphoproteomics in the Age of Rapid and Deep Proteome Profiling," *Analytical Chemistry*, vol. 88, pp. 74–94, jan 2016.

[59] Y. H. Ahn, J. Y. Kim, and J. S. Yoo, "Quantitative mass spectrometric analysis of glycoproteins combined with enrichment methods," *Mass Spectrometry Reviews*, vol. 34, pp. 148–165, apr 2015.

[60] J. J. Pitt, "Principles and {Applications} of {Liquid} {Chromatography}-{Mass} {Spectrometry} in {Clinical} {Biochemistry}," *Clin Biochem Rev*, vol. 30, pp. 19–34, feb 2009.

[61] F. Gritti and G. Guiochon, "Perspectives on the evolution of the column efficiency in liquid chromatography," *Analytical Chemistry*, vol. 85, pp. 3017–3035, mar 2013.

[62] J. J. Kirkland and J. J. DeStefano, "The art and science of forming packed analytical high-performance liquid chromatography columns," *Journal of Chromatography A*, vol. 1126, pp. 50–57, sep 2006.

[63] E. Shishkova, A. S. Hebert, and J. J. Coon, "Now, More Than Ever, Proteomics Needs Better Chromatography," *Cell Systems*, vol. 3, no. 4, pp. 321–324, 2016.

[64] E. Shishkova, A. S. Hebert, M. S. Westphall, and J. J. Coon, "Ultra-High Pressure (>30,000 psi) Packing of Capillary Columns Enhancing Depth of Shotgun Proteomic Analyses," *Analytical Chemistry*, vol. 90, pp. 11503–11508, oct 2018.

[65] M. P. Washburn, D. Wolters, and J. R. Yates, "Large-scale analysis of the yeast proteome by multidimensional protein identification technology," *Nature Biotechnology*, vol. 19, pp. 242–247, mar 2001.

[66] K. Bunai and K. Yamane, "Effectiveness and limitation of two-dimensional gel electrophoresis in bacterial membrane protein proteomics and perspectives," *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences*, vol. 815, pp. 227–236, feb 2005.

[67] E. Yaoita, S. Magdeldin, Y. Yoshida, I. Lokamani, S. Enany, B. Xu, T. Yamamoto, Z. Zureena, and Y. Zhang, "Basics and recent advances of two dimensional- polyacrylamide gel electrophoresis," *Clinical Proteomics*, vol. 11, p. 16, apr 2014.

[68] B. Manadas, V. M. Mendes, J. English, and M. J. Dunn, "Peptide fractionation in proteomics approaches," *Expert Review of Proteomics*, vol. 7, pp. 655–663, oct 2010.

[69] D. R. Stein, X. Hu, S. J. Mccorrister, G. R. Westmacott, F. A. Plummer, T. B. Ball, and M. S. Carpenter, "High pH reversed-phase chromatography as a superior fractionation scheme compared to off-gel isoelectric focusing for complex proteome analysis," *Proteomics*, vol. 13, pp. 2956–2966, oct 2013.

[70] F. Yang, Y. Shen, D. G. Camp, and R. D. Smith, "High-pH reversed-phase chromatography with fraction concatenation for 2D proteomic analysis," *Expert Review of Proteomics*, vol. 9, pp. 129–134, apr 2012.

[71] J. B. Fenn, M. Mann, C. K. Meng, S. F. Wong, and C. M. Whitehouse, "Electrospray ionization for mass spectrometry of large biomolecules," *Science*, vol. 246, pp. 64–71, oct 1989.

[72] K. Tanaka, H. Waki, Y. Ido, S. Akita, Y. Yoshida, T. Yoshida, and T. Matsuo, "Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry," *Rapid Communications in Mass Spectrometry*, vol. 2, no. 8, pp. 151–153, 1988.

[73] A. M. Haag, "Mass analyzers and mass spectrometers," in *Advances in Experimental Medicine and Biology* (H. Mirzaei and M. Carrasco, eds.), vol. 919 of *Advances in {Experimental} {Medicine} and {Biology}*, pp. 157–169, Cham: Springer International Publishing, 2016.

[74] G. Espadas, E. Borràs, C. Chiva, and E. Sabidó, "Evaluation of different peptide fragmentation types and mass analyzers in data-dependent methods using an Orbitrap Fusion Lumos Tribrid mass spectrometer," *Proteomics*, vol. 17, may 2017.

[75] G. L. Glish and D. J. Burinsky, "Hybrid Mass Spectrometers for Tandem Mass Spectrometry," *Journal of the American Society for Mass Spectrometry*, vol. 19, pp. 161–172, feb 2008.

[76] D. F. Hunt, J. J. Coon, M. J. Schroeder, J. E. P. Syka, and J. Shabanowitz, "Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry," *Proceedings of the National Academy of Sciences*, vol. 101, pp. 9528–9533, jun 2004.

[77] M. S. Kim and A. Pandey, "Electron transfer dissociation mass spectrometry in proteomics," *Proteomics*, vol. 12, pp. 530–542, feb 2012.

[78] L. M. Mikesh, B. Ueberheide, A. Chi, J. J. Coon, J. E. P. Syka, J. Shabanowitz, and D. F. Hunt, "The utility of ETD mass spectrometry in proteomic analysis," *Biochimica et Biophysica Acta - Proteins and Proteomics*, vol. 1764, pp. 1811–1822, dec 2006.

[79] A. Doerr, "DIA mass spectrometry," *Nature Methods*, vol. 12, pp. 35–35, dec 2014.

[80] A. Hu, W. S. Noble, and A. Wolf-Yadlin, "Technical advances in proteomics: new developments in data-independent acquisition," *F1000Research*, vol. 5, p. 419, mar 2016.

[81] V. Vidova and Z. Spacil, "A review on mass spectrometry-based quantitative proteomics: Targeted and data independent acquisition," *Analytica Chimica Acta*, vol. 964, pp. 7–23, apr 2017.

[82] K. F. Medzihradszky and R. J. Chalkley, "Lessons in de novo peptide sequencing by tandem mass spectrometry," *Mass Spectrometry Reviews*, vol. 34, pp. 43–63, feb 2015.

[83] J. R. Yates, J. K. Eng, A. L. McCormack, and D. Schieltz, "Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database.," *Analytical chemistry*, vol. 67, pp. 1426–36, apr 1995.

[84] S. Aggarwal and A. K. Yadav, "False discovery rate estimation in proteomics," *Methods in Molecular Biology*, vol. 1362, pp. 119–128, 2016.

[85] H. Choi and A. I. Nesvizhskii, "False discovery rates and related statistical concepts in mass spectrometry-based proteomics," *Journal of Proteome Research*, vol. 7, pp. 47–50, jan 2008.

[86] J. E. Elias and S. P. Gygi, "Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry," *Nature Methods*, vol. 4, pp. 207–214, mar 2007.

[87] L. Käll, J. D. Storey, M. J. MacCoss, and W. S. Noble, "Assigning significance to peptides identified by tandem mass spectrometry using decoy databases," *Journal of Proteome Research*, vol. 7, pp. 29–34, jan 2008.

[88] B. Zhang, M. C. Chambers, and D. L. Tabb, "Proteomic parsimony through bipartite graph analysis improves accuracy and transparency," *Journal of Proteome Research*, vol. 6, pp. 3549–3557, sep 2007.

[89] A. N. Kettenbach, J. Rush, and S. A. Gerber, "Absolute quantification of protein and post-translational modification abundance with stable isotope-labeled synthetic peptides," *Nature Protocols*, vol. 6, pp. 175–186, feb 2011.

[90] N. Rauniyar and J. R. Yates, "Isobaric labeling-based relative quantification in shotgun proteomics," *Journal of Proteome Research*, vol. 13, pp. 5293–5309, dec 2014.

[91] N. Pappireddi, L. Martin, and M. Wühr, "A Review on Quantitative Multiplexed Proteomics," *ChemBioChem*, vol. 0, no. ja, 2019.

[92] D. B. McClatchy and J. R. Yates, "Stable isotope labeling in mammals (SILAM)," *Methods in Molecular Biology*, vol. 1156, pp. 133–146, 2014.

[93] M. Krüger, M. Moser, S. Ussar, I. Thievessen, C. A. Luber, F. Forner, S. Schmidt, S. Zanivan, R. Fässler, and M. Mann, "SILAC Mouse for Quantitative Proteomics Uncovers Kindlin-3 as an Essential Factor for Red Blood Cell Function," *Cell*, vol. 134, pp. 353–364, jul 2008.

[94] X. Wang, Y. He, Y. Ye, X. Zhao, S. Deng, G. He, H. Zhu, N. Xu, and S. Liang, "SILAC-based quantitative MS approach for real-time recording protein-mediated cell-cell interactions," *Scientific Reports*, vol. 8, p. 8441, may 2018.

[95] A. Thompson, J. Schäfer, K. Kuhn, S. Kienle, J. Schwarz, G. Schmidt, T. Neumann, and C. Hamon, "Tandem mass tags: A novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS," *Analytical Chemistry*, vol. 75, pp. 1895–1904, apr 2003.

[96] S. Wiese, K. A. Reidegeld, H. E. Meyer, and B. Warscheid, "Protein labeling by iTRAQ: A new tool for quantitative mass spectrometry in proteome research," *Proteomics*, vol. 7, pp. 340–350, feb 2007.

[97] M. Y. Hein, C. A. Luber, I. Paron, J. Cox, N. Nagaraj, M. Mann, M. Y. Hein, C. A. Luber, I. Paron, N. Nagaraj, and M. Mann, "Accurate Proteome-wide Label-free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed MaxLFQ," *Molecular & Cellular Proteomics*, vol. 13, pp. 2513–2526, sep 2014.

[98] H. Liu, R. G. Sadygov, and J. R. Yates, "A Model for Random Sampling and Estimation of Relative Protein Abundance in Shotgun Proteomics," *Analytical Chemistry*, vol. 76, pp. 4193–4201, jul 2004.

[99] A. Mendoza, K. G. Pierce, J. R. Sevinsky, W. M. Old, N. G. Ahn, K. Meyer-Arendt, K. A. Resing, and L. Aveline-Wolf, "Comparison of Label-free Methods for Quantifying Human Proteins by Shotgun Proteomics," *Molecular & Cellular Proteomics*, vol. 4, pp. 1487–1502, oct 2005.

[100] D. L. Tabb, L. Vega-Montoto, P. A. Rudnick, A. M. Variyath, A.-J. L. Ham, D. M. Bunk, L. E. Kilpatrick, D. D. Billheimer, R. K. Blackman, H. L. Cardasis, S. A. Carr, K. R. Clauser, J. D. Jaffe, K. A. Kowalski, T. A. Neubert, F. E. Regnier, B. Schilling, T. J. Tegeler, M. Wang, P. Wang, J. R. Whiteaker, L. J. Zimmerman, S. J. Fisher, B. W. Gibson, C. R. Kinsinger, M. Mesri, H. Rodriguez, S. E. Stein, P. Tempst, A. G. Paulovich, D. C. Liebler, and C. Spiegelman, "Repeatability and Reproducibility in Proteomic Identifications by Liquid Chromatography Tandem Mass Spectrometry," *Journal of Proteome Research*, vol. 9, pp. 761–776, feb 2010.

[101] H. Marx, C. E. Minogue, D. Jayaraman, A. L. Richards, N. W. Kwiecien, A. F. Siahpirani, S. Rajasekar, J. Maeda, K. Garcia, A. R. Del Valle-Echevarria, J. D. Volkening, M. S. Westphall, S. Roy, M. R. Sussman, J. M. Ané, and J. J. Coon, "A proteomic atlas of the

legume Medicago truncatula and its nitrogen-fixing endosymbiont Sinorhizobium meliloti," *Nature Biotechnology*, vol. 34, pp. 1198–1205, nov 2016.

[102] J. A. Stefely, N. W. Kwiecien, E. C. Freiberger, A. L. Richards, A. Jochem, M. J. Rush, A. Ulbrich, K. P. Robinson, P. D. Hutchins, M. T. Veling, X. Guo, Z. A. Kemmerer, K. J. Connors, E. A. Trujillo, J. Sokol, H. Marx, M. S. Westphall, A. S. Hebert, D. J. Pagliarini, and J. J. Coon, "Mitochondrial protein functions elucidated by multi-omic mass spectrometry profiling," *Nature Biotechnology*, vol. 34, pp. 1191–1197, nov 2016.

[103] P. Jha, P. Blattmann, S. Dubuis, Y. Wu, T. Amariuta, J. Auwerx, C. A. Argmann, R. Aebersold, S. M. Houten, E. G. Williams, W. Wolski, and N. Zamboni, "Systems proteomics of liver mitochondria function," *Science*, vol. 352, pp. aad0189–aad0189, jun 2016.

[104] J. M. Chick, S. C. Munger, P. Simecek, E. L. Huttlin, K. Choi, D. M. Gatti, N. Raghupathy, K. L. Svenson, G. A. Churchill, and S. P. Gygi, "Defining the consequences of genetic variation on a proteome-wide scale," *Nature*, vol. 534, pp. 500–505, jun 2016.

[105] E. L. Huttlin, R. J. Bruckner, A. Ordureau, G. Zarraga, D. Kolippakkam, R. Rad, L. P. Vaites, S. Tam, B. K. Erickson, J. Szpyt, K. Baltier, J. Chick, M. P. Gygi, J. A. Paulo, M. E. Sowa, T. Harris, V. Guarani, F. Gebreab, M. Wühr, J. W. Harper, G. Colby, L. Ting, R. Dong, J. Mintseris, S. P. Gygi, R. A. Obar, S. Artavanis-Tsakonas, B. Zhai, and P. De Camilli, "The BioPlex Network: A Systematic Exploration of the Human Interactome," *Cell*, vol. 162, pp. 425–440, jul 2015.

[106] M. Choi, C. Y. Chang, T. Clough, D. Broudy, T. Killeen, B. MacLean, and O. Vitek, "MSstats: An R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments," *Bioinformatics*, vol. 30, pp. 2524–2526, sep 2014.

[107] E. Dogu, S. M. Taheri, R. Olivella, F. Marty, I. Lienert, L. Reiter, E. Sabido, and O. Vitek, "MSstatsQC 2.0: R/Bioconductor Package for Statistical Quality Control of Mass Spectrometry-Based Proteomics Experiments," *Journal of Proteome Research*, vol. 18, pp. 678–686, feb 2019.

[108] A. M. Edwards, R. Isserlin, G. D. Bader, S. V. Frye, T. M. Willson, and F. H. Yu, "Too many roads not taken," *Nature*, vol. 470, pp. 163–165, feb 2011.

Chapter 2

MULTI-OMICS EVIDENCE FOR INHERITANCE OF ENERGY PATHWAYS IN RED BLOOD CELLS

This chapter has been published and is reprinted with permission from:

Weisenhorn, EM, van't Erve, TJ, Westphall, MS, Hess, JR, Raife TJ, Coon JJ. *Multi-omics Evidence for Inheritance of Energy Pathways in Red Blood Cells.*
Molecular and Cellular Proteomics. **2016**, *15* (*12*), 3614-3623. doi: 10.1074/mcp.M116.062349.

Abstract

Each year over 90 million units of blood are transfused worldwide. Our dependence on this blood supply mandates optimized blood management and storage. During storage, red blood cells undergo degenerative processes resulting in altered metabolic characteristics which may make blood less viable for transfusion. However, not all stored blood spoils at the same rate, a difference that has been attributed to variable rates of energy usage and metabolism in red blood cells. Specific metabolite abundances are heritable traits; however, the link between heritability of energy metabolism and red blood cell storage profiles is unclear. Herein we performed a comprehensive metabolomics and proteomics study of red blood cells from 18 mono- and di-zygotic twin pairs to measure heritability and identify correlations with ATP and other molecular indices of energy metabolism. Without using affinity-based hemoglobin depletion, our work afforded the deepest multi-omic characterization of red blood cells to date (1,280 membrane proteins and 330 metabolites), with 119 membrane protein and 148 metabolite concentrations found to be over 30% heritable. We demonstrate a high degree of heritability in the concentration of energy metabolism metabolites, especially glycolytic metabolites. In addition to being heritable, proteins and metabolites involved in glycolysis and redox metabolism are highly correlated, suggesting that crucial energy metabolism pathways are inherited *en bloc* at distinct levels. We conclude that individuals can inherit a phenotype composed of higher or lower concentrations of these proteins together. This can result in vastly different red blood cells storage profiles

which may need to be considered to develop precise and individualized storage options. Beyond guiding proper blood storage, this intimate link in heritability between energy and redox metabolism pathways may someday prove useful in determining an individual's predisposition towards metabolic diseases.

Introduction

The potency of harvested red blood cells (RBCs) depends on their ability to survive and maintain function during storage. RBC viability primarily depends on their ability to resist programmed cell death-related fragmentation and phagocytosis by maintaining proper energetics and avoiding hemolysis, in which they break down into microvesicles and toxic byproducts including iron, heme, hemoglobin, and oxidized lipids. The released iron can feed bacterial infections and free hemoglobin can interfere with nitric oxide signaling[1,2]. A number of small and retrospective studies have suggested that prolonged RBC storage is associated with negative clinical outcomes; however, three larger randomized clinical trials showed no negative effects of longer-stored RBCs[3–6]. In short, the viability of stored RBCs is variable and not fully understood, but the accumulation of biophysical and metabolic changes known as storage lesions are linked to the ability to maintain flux through metabolic pathways during storage[7,8].

Post-storage RBC adenosine triphosphate (ATP) concentration is the single best predictor of RBC *in vivo* recovery in autologous blood transfusions[9–12]. Specifically, high ATP concentrations are correlated with low levels of hemolysis and other storage lesions in RBCs. Interestingly, post-storage ATP levels vary greatly between individuals but are consistent on repeat measure within an individual. This observation suggests that post-storage ATP, and thus stored RBC viability, may be influenced and/or determined by inheritance[13–15]. In prior analyses of these samples and in additional studies of mono- and di-zygotic twins,

some metabolite concentrations including glucose 6-phosphate, fructose 1,6-bisphosphate, glutathione, and glutathione disulfide were determined to be heritable in stored RBCs[13,14,16]. The metabolite concentrations of ribulose 5-phosphate, sorbitol, and xylulose 5-phosphate are heritable suggesting a genetic control of glucose metabolism[14–16].

Since RBCs eject all organelles, including the nucleus and mitochondria upon maturing, they have no ability to synthesize proteins in response to environmental stimuli. The lack of mitochondria in mature RBCs also leaves these cells unable to rely on oxidative phosphorylation; instead, RBCs are reliant on glycolysis for energy production. These unique metabolic attributes of RBCs provide a highly instructive model for unraveling how genetic regulation of metabolic pathways can impact blood storage viability.

Herein, we describe a multi-omics analysis of genetic and environmental factors dictating RBC variability. Our approach involved an extensive proteomic and metabolomics analysis of RBCs derived from a cohort of 18 mono- and di-zygotic twin-pairs.

The primary challenge of performing proteomic analyses on red blood cells is the wide dynamic range characterized by an abundance of hemoglobin. This was surmounted by focusing our analysis on the membrane fraction of red blood cells. While other studies have relied on time intensive affinity enrichment, utilizing the membrane fraction granted us the second greatest proteomic depth achieved in red blood cells which allowed us to process a multitude of clinical samples. Furthermore, much of the complexity and diversity in red blood cells is associated with the membrane including many energy metabolism components[17–19].

Despite the simplified composition of mature RBCs, i.e., no nucleus or mitochondria, detection and quantification of the RBC proteome presents a few challenges. First, RBCs must be purified from other blood cells. During this process, typically differential centrifugation, care must be taken to limit contamination, especially from the abundant plasma proteins. The next, and most significant, obstacle is the large dynamic range of protein abundance within the RBC[20,21]. Although the actual dynamic range of the RBC proteome is not yet known, the technical challenges are analogous to measuring the proteome of plasma, which has a dynamic range approaching twelve orders of magnitude[22]. Also similar to the plasma proteome, in which a single protein (albumin) constitutes 55% of the total protein content, hemoglobin comprises 97% , by mass, of the RBC proteome, making protein depletion a necessary consideration[23]. Of the remaining 3% , carbonic anhydrase accounts for 1/3, so that the remaining 2% of total protein mass is made up of several thousand different proteins. Identifying these low abundance proteins from the background presented by hemoglobin and carbonic anhydrase, is challenging[24,25]. Several methods attempt to counter this obstacle by use of various types of affinity or ion exchange separation techniques[26–28]. Even when employing these methods, most RBC proteome analyses yield detection of less than 1,000 proteins, with the exception of one which identified 1,578[29].

Most of these studies, especially those with the deepest coverage, require extensive protein and/or peptide fractionation which, in turn, yields considerable increases in analysis time – both sample preparation and instrument acquisition. Recent years have ushered in an era of proteomics where advances in peptide separation and mass spectrometer performance

has accelerated the rate and depth of proteome analysis[30]. We reasoned that application of this technology, combined with straightforward reversed-phase proteome fractionation, could expedite sample preparation and afford reasonably deep RBC proteomic analysis in short order, thus, affording the throughput for quantitative comparison of clinical RBC samples.

Using our method, we show that in RBCs the concentrations of components in crucial energy metabolism pathways are inherited *en bloc* at distinct levels. This results in different RBC storage phenotypes which can be used to further understanding of changes during storage and develop improved storage guidelines and methods. Furthermore, this rich dataset will provide a valuable resource for continuing studies of RBCs and the heritability of disease.

Materials and Methods

Twin Subject enrollment and sample collection. This report is a continuation of twin studies reported previously and utilized the same study subjects[14,16,31,32]. The study was approved by the Human Subjects office of The University of Iowa Carver College of Medicine. Written informed consent was obtained from all participating subjects. Subjects were qualified for participation by meeting criteria for autologous blood donation according to standard operating procedures of The University of Iowa DeGowin Blood Center. Twin pairs were not required to donate samples at the same time and each individual donated a single blood unit. Standard health history and demographic information were obtained

at the time of enrollment and informed consent. Reported height and weight were used to calculate body mass index (BMI). BMI was derived from the formula: BMI = weight (kg) / (height(m))^2. Each subject donated one unit of whole blood which were processed according to standard operating procedures into a leukocyte-reduced RBC unit in CP2D/AS-3 extended storage media (Haemonetics Corp, Braintree, MA). During processing, integral leukocyte reduction filters were retained for extraction of DNA.

Sample Preparation. Samples of AS-3 preserved RBC units were prepared from the main unit on each day of sampling. The AS-3 preserved RBCs were sampled by sterile docking of tubing to the RBC unit, back-filling the tubing with RBCs and sectioning into segments. This procedure was performed on the first day after donation (day 0), and every 14 days thereafter until day 56. This resulted in 5 time points at day 0, 14, 28, 42, and 56.

Segments were drained into 5 mL Eppendorf tubes; after mixing an aliquot is removed for complete blood count (CBC) testing using a hematology analyzer (Sysmex XE-2100™ Automated Hematology System, Sysmex Corp, Kobe, Japan). The remaining sample was centrifuged at 500 g for 5 min, after which the storage media (AS-3) was removed. Samples were further processed and used for measurement of ATP, GSH, and GSSG in RBCs as previously described[14,16].

Whole venous blood (EDTA, Vacutainer® purple top blood collection tube, 8 mL) collected from participants prior to blood donation was centrifuged at 500 g for 5 min, followed by removal of the plasma and buffy coat. RBCs were washed twice with cold iso-

tonic saline solution. After washing, a 30 μ L aliquot of the packed red blood cells (pRBCs) was removed for complete blood count (CBC) analysis (Sysmex XE-2100™ Automated Hematology System, Sysmex Corp, Kobe, Japan). A 100 μ L aliquot of pRBCs was lysed with 900 μ L of nanopure water. Samples were thoroughly mixed and stored at -80 ° C prior to proteomic and metabolomic analyses.

Zygosity testing DNA for zygosity testing was obtained from leukocyte reduction filters by rinsing filters with 15 mL DPBS. The rinse volume was centrifuged at 500 g for 10 min and the cell pellet was resuspended in 2 mL of DPBS. DNA was extracted from the cell pellet using a nucleic acid extraction instrument (AutoGen QuickGene 610L, AutoGen, Holliston, MA) and kit (Fuji QuickGene DNA Whole Blood Kit, AutoGen, Holliston, MA).

Genotyping was performed using a previously developed panel of 24 single nucleotide polymorphisms (SNPs) (10). SNP genotyping was performed using PCR assays (TaqMan, Applied Biosystems, Foster City, CA) on a Genotyping System (EP1 SNP, Fluidigm, San Francisco, CA) with a Dynamic Array Integrated Fluidic Circuit (GT48.48, Fluidigm, San Francisco, CA). Monozygotic (MZ) twins were identified by 90 % or greater genotype concordance; all other twin pairs were identified as dizygotic (DZ).

Global metabolomics profile analyses The untargeted metabolic profiling method employed for this analysis combined three independent platforms: ultrahigh performance liquid chromatography/tandem mass spectrometry (UHPLC/MS/MS) optimized for basic

species, UHPLC/MS/MS optimized for acidic species, and gas chromatography/mass spectrometry (GC/MS). Samples were analyzed using procedures described in van 't Erve *et al.* (14).

Sample Preparation for Proteomic Analysis Proteolytic Digestion: A 50 μ L aliquot of red blood cells lysed in 500 μ L DI water was centrifuged at 4° C for 30 minutes at 5 G. The supernatant was discarded and the pellet was re-suspended in 100 μ L lysis buffer (8 M Urea, 100 mM Tris, 10 mM TCEP, 40 mM chloroacetamide). The samples were then diluted with 50 mM Tris pH 7.5 until the pH reached 7.5 (~ 1 mL). Trypsin digestion was performed overnight at room temperature with trypsin (Promega, Madison, WI) added at a 1:50 (w/w) enzyme to protein ratio with an estimated protein quantity of 500 μ g. A second trypsin digestion was performed the following morning at 1:200 (w/w) enzyme to protein ratio for 1 h. Each digest was quenched by the addition of TFA and desalted over tC18 Sep-Pak cartridges (Waters, Milford, MA).

High pH fraction collection: Samples were fractionated using high pH reverse phase separation to increase proteomic depth. The solvent system consisted of mobile phase A (20 mM ammonium bicarbonate) and mobile phase B (20 mM ammonium bicarbonate 80% acetonitrile) which was run on an Ultimate 3000 UPLC system (Dionex Sunnyvale, CA) with a reverse phase C18 column. Gradient elution was performed at 400 μ L min^{-1} with the gradient increased from 0 to 6% B over 5 minutes followed by an increase to 80% B until 24 minutes and a wash at 100% B for 3 minutes. Eight fractions were collected from

each sample which were subsequently pooled resulting in four total fractions per sample.

nLC-MS/MS Analysis Samples were analyzed using a LC/MS instrument comprising an Orbitrap Elite hybrid mass spectrometer (Thermo Fisher Scientific). Reverse phase columns were prepared in house using a 75-360 μ m inner-outer diameter bare-fused silica capillary with laser pulled tip. The column was packed with 1.7 μ m diameter, 130 Åpore size, Bridged Ethylene Hybrid C18 particles (Waters) to a final length of 35 cm. The column was installed on a Dionex Ultimate 3000 UPLC system and heated to 60° C using an in house designed column heater for all runs[33,34]. Mobile phase buffer A was composed of water, 0.2% formic acid, and 5% DMSO. Mobile phase B was composed of acetonitrile, 0.2% formic acid, and 5% DMSO. 1 μ g of sample was injected as determined by quantitative colorimetric peptide assay (Pierce Rockford, IL). Gradient elution was performed at 300 nL min^{-1} with the gradient increased linearly from 0 to 60% B over 103 minutes followed by a linear increase to 100% B until 106 minutes and a wash at 100% B for 4 minutes. Survey scans of peptide precursors were collected from 300-1500 Th with an AGC target of 1,000,000 and a resolution of 60,000 followed by data dependent CID MS/MS scans of the 20 most intense peaks in the quadrupole linear ion trap mass analyzer. Precursors with charge states equal to 1 or unassigned were rejected and a 45 second dynamic exclusion was set to expedite identifications.

Data Analysis Label free quantification was performed using Maxquant software version 1.5.2.8[35] and the Andromeda search engine[36]. The results were searched against a *Homo sapiens* database containing 90,482 reviewed proteins plus isoforms downloaded from Uniprot on June 23, 2015. Enzyme specificity was set to fully tryptic with up to two missed cleavages and carbamidomethylation of cysteines as a fixed modification. Oxidation of methionines and protein N-terminal acetylation were set as variable modifications. The match between runs feature was utilized to decrease missing data values within the data set (35). Precursor mass tolerance was 20 ppm and product ions were searched at 0.5 Da tolerances. Peptides were filtered to a 1% FDR and combined to protein groups based on the rules of parsimony, with at least two peptides per protein. Pearson correlations were calculated between each protein and metabolite detected using Perseus software[37,38].

Experimental Design and Statistical Rationale Five di-zygotic twin pairs and thirteen mono-zygotic twin pairs were used in the study. No biological replicates were available since each individual was only required to donate blood at one time. Proteome and metabolome analyses were performed in randomized order to eliminate systematic biases.

Heritability calculations Heritability estimates were calculated for each protein and metabolite concentration measured, and for each measured time point when applicable. The first step to calculating heritability is using the one-way model of intraclass correlation coefficient (ICC) to determine the similarity of a measure in a twin pair: ICC = ($MS_{between}$

- MS_{within}) / ($MS_{between}$ + MS_{within}), where $MS_{between}$ is the estimate of the mean-square variance between all twin-pairs and MS_{within} is the estimate of the mean-square variance within the sets of pairs in that group (13). The ICC is used to compare the variation within specific pairs to that of the population as a whole, and falls on a scale of -1 to +1. Higher positive values indicate that there is less variation within the pairs of subjects than there would be within randomly paired subjects. Positive values approaching 0, as well as negative ICC values, indicate that the variation within pairs of subjects is similar to the variation expected within random pairs. A highly heritable trait between MZ twins would be expected to have an intraclass correlation coefficient near +1. Once ICC values were calculated, heritability was estimated using the method derived by Newman *et al.*, h^2 = (ICC_{MZ} - ICC_{DZ}) / (1 - ICC_{DZ})[39].

Results and Discussion

To rapidly remove hemoglobin from RBCs, we separated RBCs obtained from whole blood via differential centrifugation. Samples were then centrifuged again, to isolate the membrane fraction, which was kept for further proteomic analysis while the supernatant (containing predominantly hemoglobin) was discarded[40]. While enriching for the membrane fraction will bias our analyses toward the detection of membrane bound proteins, many proteins of interest in red blood cells are associated with the membrane including some glycolytic proteins[18,19,41]. Following this extraction, proteins were digested with trypsin, and the resultant peptides separated into eight fractionations via high pH reversed-phase liquid

chromatography (RPLC). These fractions were recombined, generating four fractions per sample. Each fraction was then analyzed using a 120 minute nanoLC-MS/MS method. In total, each sample required eight hours of mass spectrometer analysis, yielding an average of 3,678 peptide spectral matches (PSMs), 2,357 unique peptides, and 606 proteins per RBC sample.

Once the challenging problem of depleting hemoglobin had been overcome, we turned our attention to clinical RBC samples from 36 individuals including five di-zygotic and 13 mono-zygotic twin pairs. Twins were permitted to donate blood at separate times under separate conditions which serves to strengthen our confidence in heritability calculations. A 50 μ L aliquot of washed, lysed RBCs from each patient was analyzed as described above. Across all 36 patients we detected 1,280 proteins with an FDR less than 1% . Of these, 105 proteins were detected in all patients, however of the proteins in our model and those further analyzed in energy metabolism pathways all were present in at least 27 individuals. The large majority (92%) of these proteins were identified with at least two peptides uniquely mapping to their sequence. The mean protein sequence coverage was 19.4% . Our data shows significant overlap with previous results – among our 1,280 identified proteins, 941 have previously been observed (73%). When performing heritability calculations measurements were required to be present for five out of five dizygotic twin samples and 10 out of 13 monozygotic samples.

Metabolomic Analysis As a relative newcomer in the 'omics' era, metabolomics lags behind proteomics in the robust quantification of thousands of compounds. Discovery metabolomics aims to identify the entire metabolome present in cells; however, the greatest hurdle is the identification of unknown features. Using such discovery metabolomics assay ~ 300 unique metabolites were quantified from these same samples.

Briefly, the discovery metabolic profiling method combined three independent platforms: ultrahigh performance liquid chromatography/tandem mass spectrometry (UHPLC/MS/MS) optimized for basic species, UHPLC/MS/MS optimized for acidic species, and gas chromatography/mass spectrometry (GC/MS). This method resulted in the quantification of 328 metabolites including lipids, xenobiotics, dipeptides, and many metabolites from prominent energy pathways (Supplemental Table 4). Together,our proteomic and metabolomic datasets comprise the largest multi-omic dataset of red blood cells.

Correlation Analysis To identify potentially co-regulated proteins and metabolites Pearson correlation analysis and hierarchical clustering were performed between all proteins and metabolites yielding 58,000 correlations either greater than 0.75 or less than -0.75 which corresponds to 5% of the total correlations measured.

These clusters were found to contain unique protein groups containing proteosomal, fatty acid metabolism, or energy metabolism proteins. Of particular interest is a cluster containing numerous proteosomal proteins as well as those involved in glutathione metabolism and glycolysis (Figure 1). We also observe glycolytic and glutathione metabolism proteins

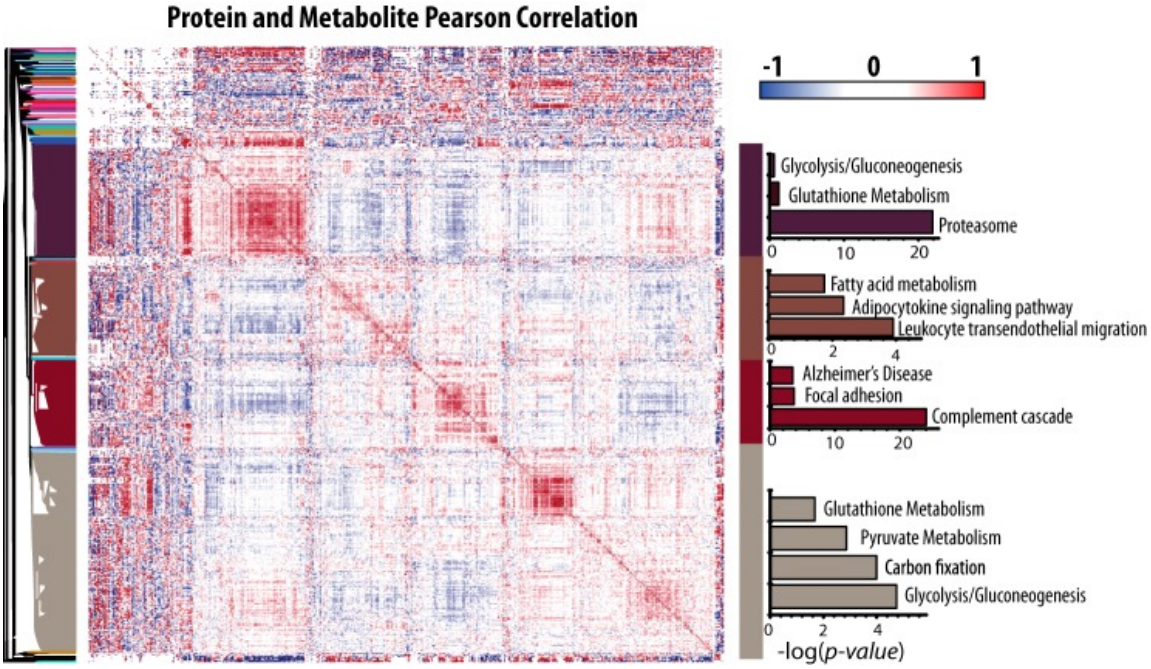


Figure 2.1: Red blood cell proteins and metabolites show clusters of high correlation. Pearson correlation values were calculated between every combination of proteins and metabolites and plotted using Perseus. KEGG pathway enrichment of various clusters was measured using enrichr and the negative log of the p-value for pathways of interest is reported in the bar charts to the right, where the color of the bars corresponds to the section of the dendrogram where the pathways are enriched.

clustering with pyruvate metabolism and carbon fixation. All of these pathways were significantly enriched Kyoto Encyclopedia of Genes and Genomes (KEGG) terms using the online software enrichr[42].

Glycolytic protein and metabolite levels were normalized using feature scaling to examine variation within glycolysis. By comparing normalized protein or metabolite levels in glycolysis we note that variation within the glycolytic pathway occurs *en bloc* at various levels (Figure 2). The distribution of variation within glycolytic proteins and metabolites indicates that metabolites are more tightly conserved than proteins (Figure 2 C, D). Glycolytic proteins and metabolites each cluster together and display a high number of positive correlations supporting our conclusion that variation in glycolysis occurs *en bloc*.

Heritability Among 18 twin pairs, zygosity testing identified 13 MZ and 5 DZ twin pairs. The means of age, weight, and BMI were not significantly different between MZ and DZ twin groups (**Table 1**). As previously reported, a high degree of estimated heritability for height (96 %), weight (97 %), and BMI (63 %) was observed in this study population (31). The similarity of these results to estimates in a previous report studying 30,111 twin pairs in 8 countries supports the validity of the sample population for determination of heritable traits[43].

Of the proteins and metabolites measured, 119 protein and 148 metabolite concentrations were found to be over 30% heritable, and 73 and 104 were greater than 50% heritable respectively (Figure 3) (Supplemental Table 3). Previous studies using this twin cohort

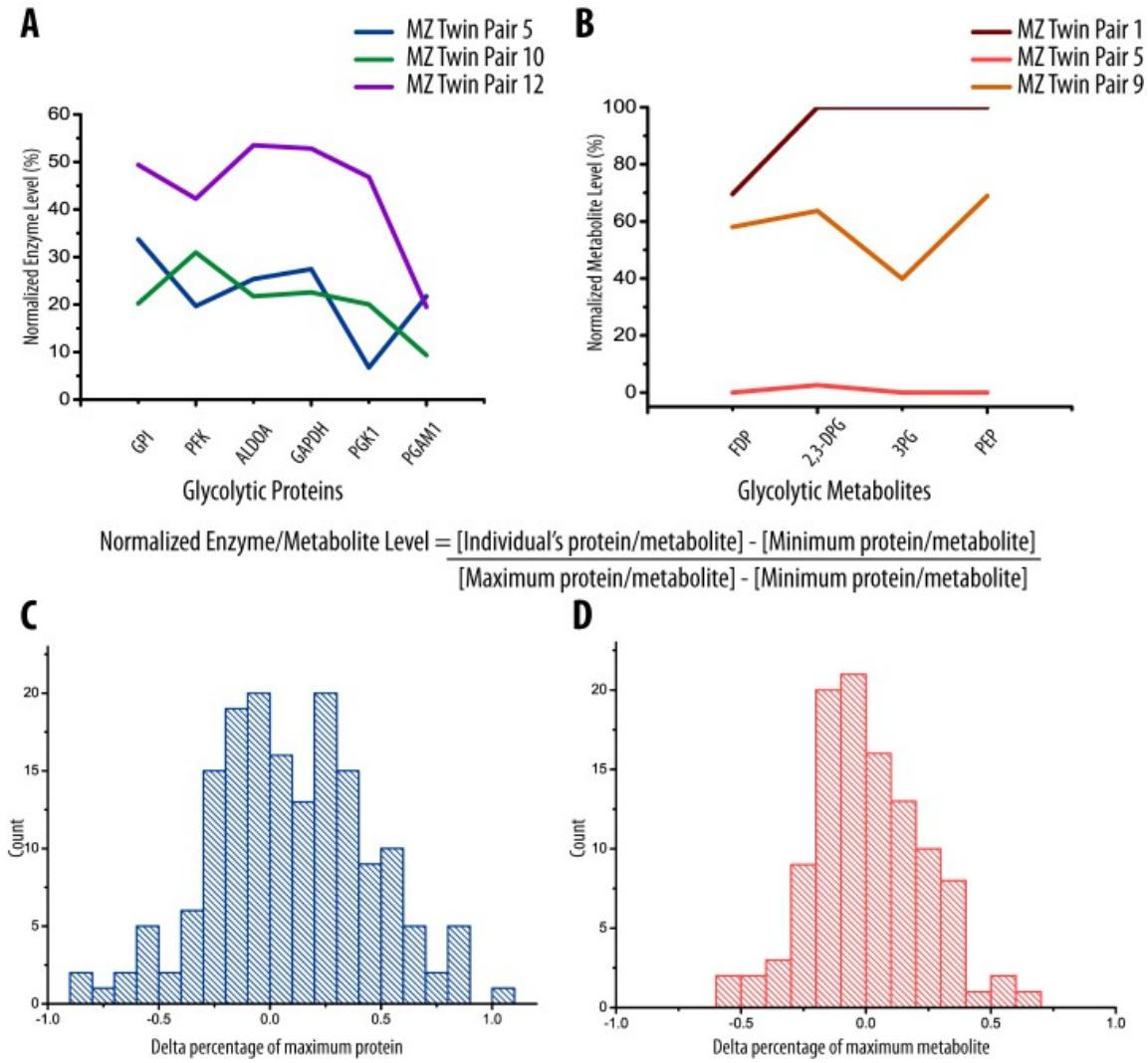
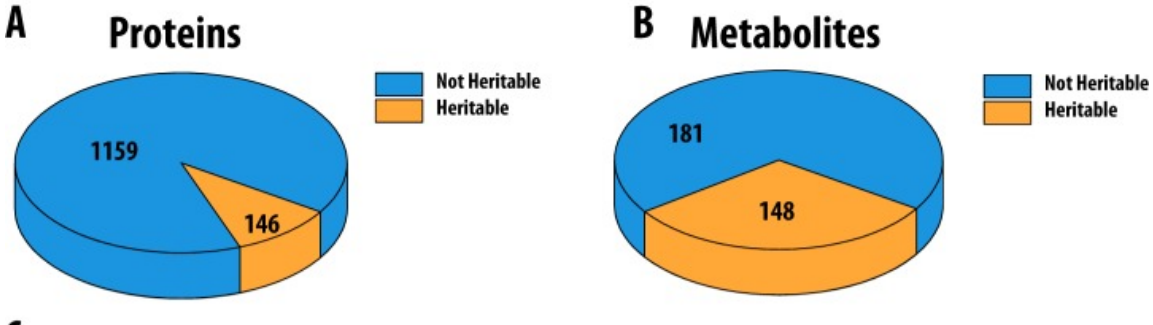


Figure 2.2: Relative abundance of glycolytic proteins and glycolytic metabolites is conserved among twin pairs. The average protein (A) and metabolite level (B) in three representative monozygotic twin pairs show variation in glycolytic activity within the population, indicating abundances of these proteins are influenced *en bloc*. All protein and metabolite levels were normalized to the percentage of maximum protein and metabolite level (see example calculation) using feature scaling. Pyruvate was excluded from the metabolites as it was found to not correlate with the other members. The difference between each protein (C) and metabolite (D) is reported in a histogram. Sequential proteins and metabolites in glycolysis were subtracted to give a scaled difference components of the pathway. The peak is centered at zero for proteins and metabolites indicating that individuals inherit high or low levels of glycolytic compounds together. Glycolytic metabolites appear to be more tightly conserved than protein.



C

Glycolysis

	Heritability
Phosphofructokinase	57%
Phosphoglycerate mutase	30%
Bisphosphoglycerate mutase	50%
Fructose diphosphate	71%
3-phosphoglycerate	75%
DHAP	79%
2,3-DPG	85%
Phosphoenolpyruvate	78%
Pyruvate	61%
Band 3	49%
Carbonic anhydrase1	84%
AMP	73%

Glutathione Metabolism

	Heritability
Glutamate	91%
Cysteinylglycine	85%
Glutathione peroxidase 4	75%
Glutathione disulfide	71%
Glutathione	56%
Hemoglobin delta	54%
Glutamate cysteine ligase	45%
Ribose-5-P	43%
Opthalmate	41%
Glutathione S-transferase	20%
Hemoglobin gamma	60%
Mannose-6-P	64%

Figure 2.3: A total of 1,280 proteins and 330 metabolites were detected in red blood cells. Of these, 119 and 148 were found to be over 30% heritable, respectively. To calculate heritability in proteins, measurements were required to be present in all three out of five dizygotic twin pairs and 10 out of 13 dizygotic twin pairs. (C) Proteins and metabolites greater than 30% heritable from glycolysis and glutathione metabolism are reported.

have used 30% heritability as a limit for consideration. In particular, we noted that the concentration of the key regulatory enzyme liver type phosphofructokinase (PFK) is 57% heritable as well as the concentration of phosphoglycerate mutase which is 29% heritable. Muscle and platelet PFK isoforms were also detected in our dataset, but were not detected in sufficient samples to determine heritability. Additionally the concentration of bisphosphoglycerate mutase (BPGM), a key enzyme for regulating the oxygen loading capacity of hemoglobin in red blood cells was 50% heritable. No other glycolytic proteins were found to be heritable; however, we were interested to observe that heritable proteins were found at important regulatory steps and branch points in the pathway. The heritability of glycolysis is further supported by high heritability estimates of the metabolites fructose 1,6-bisphosphate, 3-phosphoglycerate, DHAP, 2,3-DPG, phosphoenolpyruvate, and pyruvate. Within glutathione metabolism, GST, GCLC, GPx4, and several hemoglobin subunits were found to be heritable as well as many metabolites including glutamate, cysteinylglycine, GSSG, GSH, and ribose-5-phosphate (Figure 3).

High levels of heritability are similarly observed within glutathione metabolism and the pentose phosphate pathway. For example, concentrations of the proteins glutathione peroxidase, glutathione S-transferase, and glutamate cysteine ligase were heritable as well as the metabolite concentrations of ribose-5-phosphate, glutathione, and glutathione disulfide. These heritable metabolite concentrations are in accordance with those detected in previous red blood cell studies[14].

Many of the proteins and metabolites implicated in glutathione metabolism are also

correlated and cluster with those discussed previously in glycolysis. This is not surprising, as maintaining redox balance is of key importance to red blood cells and helps preserve sufficient NAD concentration to continue glycolysis. Glycolysis and glutathione metabolism are highly correlated and contain heritable concentrations of proteins and metabolites, implying they are inherited together at varying degrees (Figure 4).

The best marker for 42 days post-storage ATP concentration in glycolysis and glutathione metabolism is carbonic anhydrase 1 (CA1) which has a -0.56 correlation with ATP. This correlation strengthens with increased time in storage: -0.10 at day 0, -0.22 at day 14, -0.25 at day 28, and -0.56 at day 42. CA1 catalyzes the conversion of CO_2 and H_2O to produce carbonic acid, which is de-protonated at neutral pH, generating a proton and lowering the pH of stored blood.

The acidification of blood during storage is a well-characterized phenomenon and the resulting decreased pH inhibits PFK[44]. PFK inhibition caused by acidic conditions resulting from CA1 may explain the correlation we observe between higher CA1 concentrations and lower post-storage ATP. The newest blood storage solution, AS-7, buffers blood acidification with the addition of bicarbonate to increase post-storage ATP concentration and *in vivo* recovery[44–46]. Maintaining a high pH during storage is thus imperative for ATP production, and supported by the correlation we have shown between CA1 and post-storage ATP.

Model of post-storage ATP An important goal of transfusion medicine is to improve post-storage ATP levels in blood, as RBC ATP concentrations correlate positively with transfused

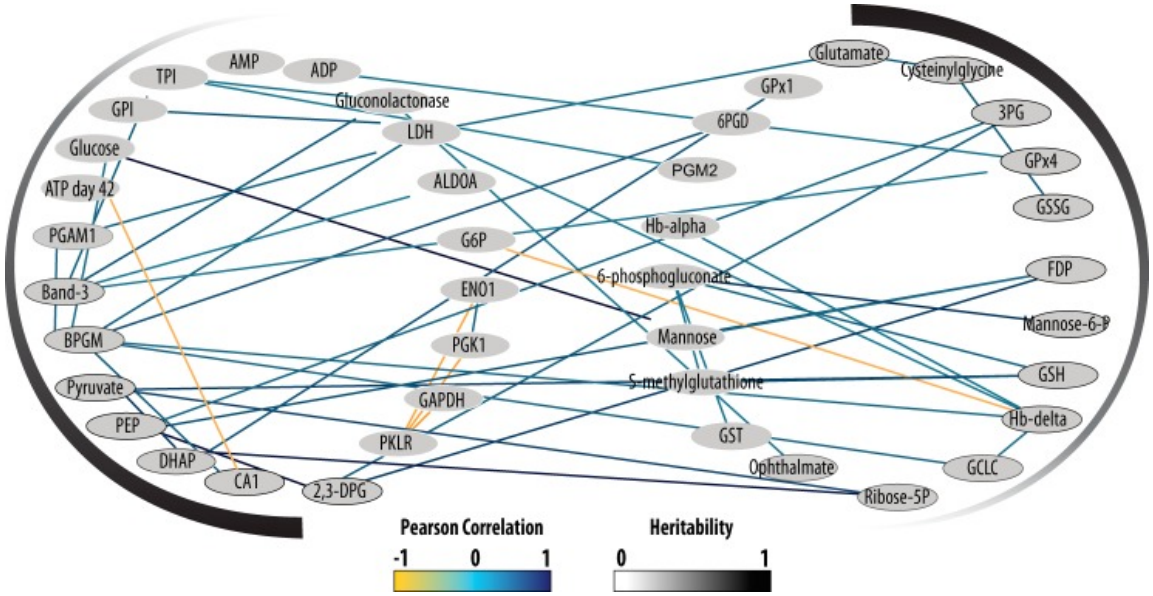


Figure 2.4: A high number of positive correlations are observed between both proteins and metabolites in the glycolytic and glutathione metabolism pathways . A Pearson correlation greater than 0.5 or less than -0.5 is required to show a connection. Heritability of these pathways can be observed in the shade of the node outline as well as by the gradient outside the network. This figure was created using Cytoscape™(58).

RBC recovery.[46,47]. Using our large scale data set we considered two potential phenotypes of high or low concentrations of ATP at day 42 of storage. The 'high' and 'low' post-storage ATP phenotypes can be correlated with proteins known to effect ATP concentrations and generate a model to understand post-storage ATP levels. We include five key parameters in this model including PFK, CA1, band 3, BPGM, and pH. Strikingly, concentrations of all protein components of this model were found to be at least 45% heritable. Band 3, BPGM, and CA1 correlate negatively with day 42 post storage ATP levels (-0.41, -0.39, -0.56) and together may shuttle flux away from glycolysis and ATP production. We also observe positive correlations between pH and ATP early in storage as discussed previously (Figure 5) which appears to weaken over time. Day 0 ATP correlates positively with pH at day 7, day 14, and day 28 (0.48, 0.80, 0.51) while day 42 ATP correlates positively with pH at day 7 and day 14(0.56, 0.57). No positive correlations between pH and ATP are observed at day 42 or 56. However, PFK concentrations correlate positively with pH later in storage at day 42 and 56 (0.45, 0.42) and are associated with increased ATP generation in glycolysis.

Conclusion

Our approach to RBC proteome characterization provided an expedient and robust analysis of low abundance RBC membrane proteins, and produced the most thorough analysis of the RBC membrane proteome to-date without the use of affinity based depletion strategies. We expect that further efforts to deplete hemoglobin would result in increased identification of low abundance proteins, but this would also increase processing time per sample. Using

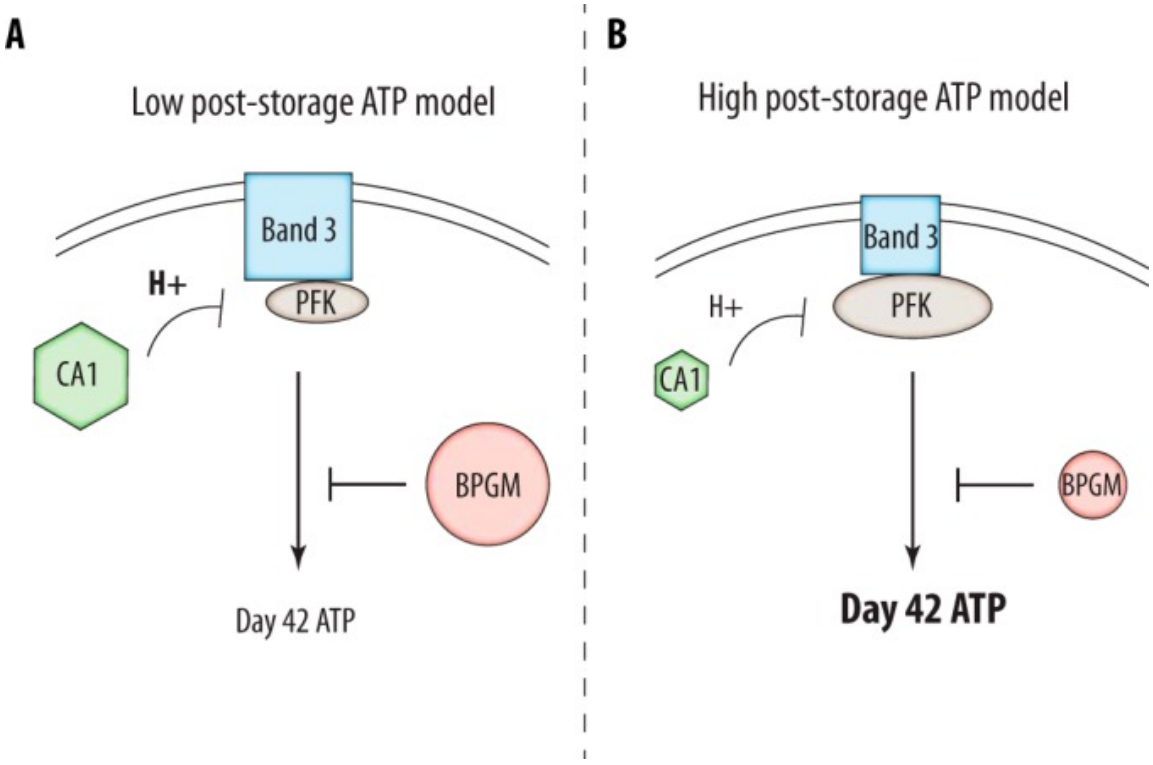


Figure 2.5: Post-storage ATP levels are determined by several key factors . (A) Low ATP levels following 42 days of storage are correlated with high levels of band 3, BPGM, and carbonic anhydrase. Band 3 binds glycolytic proteins decreasing flux through glycolysis while BPGM shunts intermediates to the luebering-rapoport pathway away from the generation of ATP. Similarly, high levels of carbonic anhydrase produce acidic conditions and subsequently inhibit PFK. In support of this we observe negative correlations between carbonic anhydrase and pH level during storage. Low post-storage ATP is additionally correlated with low pH. Correlation values of greater than 0.3 were required for consideration. (B)The opposite model leads to the generation of high ATP levels following 42 days of storage. The size of the protein in each case is representative of the concentration associated with each phenotype.

a unique data set of twin samples we determined heritability of over 700 proteins and metabolite concentrations.

Within our data set we took particular interest in the interactions between energy and glutathione metabolism. Forty-nine correlations greater than 0.5 are present between the proteins and metabolites of these pathways along with only five negative correlations. The negative correlations included three correlations between pyruvate kinase and other glycolytic proteins, between post-storage ATP and CA1, and between G6P and the delta subunit of hemoglobin. The hemoglobin subunits were included with glutathione metabolism because of their role in oxygen binding and ability to generate superoxide and hydrogen peroxide through hemoglobin[48]. Similarly, the anion transport protein band 3 was included with glycolytic proteins and metabolites because of its role binding glycolytic proteins in an oxygen dependent manner[19,49]. We thus propose that glutathione metabolism and glycolysis are highly connected pathways and may be linked by a similar regulatory mechanism.

Due to their role as oxygen carriers and thus the large quantity of oxyhemoglobin, red blood cells have an especially high burden of oxidizing species such as superoxide and hydrogen peroxide. Red blood cells limit the accumulation of these species by maintaining a large pool of reducing equivalents generated from glucose through the pentose phosphate pathway. In the lungs where O_2 levels are high, RBCs are exposed to higher levels of oxidative stress necessitating increased flux through the pentose phosphate pathway to generate reducing equivalents and supply the glutathione cycle. Meanwhile, in peripheral

tissues with low O_2 levels, erythrocytes must pass through narrow capillaries causing distortion from mechanical stress and cation leaks[19]. This causes an increased demand for ATP to restore intercellular ion balance. Band 3 specifically binds PFK, GAPDH, and ALDOA in the presence of oxyhemoglobin and diverts flux towards the pentose phosphate pathway to generate NADPH. As we would expect, band 3 correlates negatively with several glycolytic metabolites, including phosphoenolpyruvate and 2,3-DPG (-0.36 and -0.34 respectively). It also positively correlates with several glycolytic proteins, which is compatible with higher concentrations of band 3 being necessary to bind higher levels of glycolytic proteins.

Supporting this model we also observe high levels of correlation between glutathione metabolism and the pentose phosphate pathway. These are intimately linked pathways as the reductive equivalent NADPH generated by the pentose phosphate pathway is necessary to reduce glutathione disulfide generated during removal of hydrogen peroxide. Gluconolactonase is positively correlated with GPx1, GPx4, as well as glutamate cysteine ligase. Also, glutathione is correlated with 6-phosphogluconate, gluconolactonase, and 6-phosphogluconate dehydrogenase demonstrating that the pentose phosphate pathway is essential for the continuation of glutathione metabolism.

Many of the concentrations of proteins and metabolites in these pathways were also found to be heritable. Within glutathione metabolism, glutamate, GPx4, glutamate-cysteine ligase, glutathione, and glutathione disulfide concentrations were all found to be over 45% heritable. Furthermore, in glycolysis, concentrations of the regulatory enzyme PFK as

well as BPGM and the metabolites pyruvate, phosphoenolpyruvate, 3-phosphoglycerate, 1,3-bisphosphoglycerate, DHAP, and FDP were all found to be over 50% heritable (Figure 3). It is also noteworthy that one of the most heritable protein concentrations we observed was carbonic anhydrase, at 85% , which is an important regulator of pH and therefore PFK activity. Based on the strong correlation observed between these pathways, and the high levels of heritably throughout, we infer that glutathione metabolism, pentose phosphate pathway, and glycolysis are coupled pathways that can be inherited *en bloc* at various levels.

Together our results suggest a model in which inheritance of higher concentrations of band 3 and CA1 reduce flux through the glycolytic pathway by greater binding and inactivation of PFK and by allosteric inhibition of PFK through lower pH. In addition, inheritance of higher concentrations of BPGM may decrease ATP production by competing with PGK for an ATP-producing step in glycolysis. Higher inherited concentrations of PFK may increase flux through the glycolytic pathway. The combined effect of inheritance of these enzyme concentrations accounts both for the *en bloc* inheritance of glycolytic pathway intermediates, and the heritability of ATP concentration at day 42 and day 28 of storage. The heritable concentrations of many molecules in the pentose phosphate and glutathione pathways may also be predicted by this model.

Our model has implications in the management of blood storage as it confirms that energy metabolism and ATP concentrations are heritable traits. In the future we can imagine blood donors being tested once for levels of key heritable markers to determine a blood storage profile and optimum storage period. Based on our results, some donors

actually show an increase in ATP concentrations early in storage suggesting this blood could theoretically be stored longer than individuals that show a continual decrease in ATP. Individuals that have increasing ATP during storage also show correlations with higher PFK and lower BPGM. This could prevent potential ATP loss by the diversion of 1,3-bisphosphoglyverate to produce 2,3-bisphosphoglycerate. Furthermore, individuals that have decreasing ATP during storage have lower PFK and higher BPGM concentrations. These correlations are suggestive of heritable markers which could someday be used to predict post-storage ATP levels in blood donors.

The negative correlation identified between CA1 and post-storage ATP is significant in that it provides one reasonable explanation for ATP decreases that occur in storage. RBC units are stored in gas permeable bags allowing CO_2 to diffuse into the bag causing acidification and inhibiting PFK. This will decrease the natural rate of energy metabolism and subsequently ATP production. Atmospheric conditions have been shown capable of regulating RBC metabolism in storage previously in the case of oxygen saturation[49]. Our results suggest that the addition of CA1 inhibitors to stored blood, or the selection of donors known to have low inherited levels of CA1 are potential facile methods to increase the quality and lifetime of stored blood. The importance of pH modulation in blood storage was observed to be of key importance in the development of the newest blood storage solution, additive solution-7, which was primarily improved by increased buffering capacity[45].

RBC blood bank storage is not an activity that occurs in nature, so the strong genetic components suggests that it intersects with a deeper problem in evolutionary biology

such as the tradeoff involved in oxygen-based energetics with the risks of oxygen-induced biochemical damage. Keeping RBC glycolytic flux high is known to be advantageous in RBC storage and function and probably represents one of the poles of the deeper tradeoffs in cellular or whole body energetics. The implications for RBC storage are that it is possible to both identify markers to identify individual blood donors with better blood storage or to support identified aspects of metabolism in all donors that make all cells store better.

The heritability we observed in many pathways within erythrocytes may have ramifications in metabolic disease. Diseases such as Alzheimer's disease and cancer are known to involve aberrations of energy metabolism[50–55]. Current disease models suggest that a low glycolytic capacity may confer a risk of Alzheimer's disease but protect against cancer. Since we have determined many components of glycolysis to be heritable in erythrocytes, we hypothesize that other cell types are similarly affected, and individuals who inherit low levels of glycolytic proteins and metabolites may be more prone to developing Alzheimer's disease later in life. Similarly, people who inherit high levels of glycolysis may be inclined to develop cancer. Inverse comorbidity has been documented between these diseases, supporting our hypothesis that inheritance of energy metabolism along a spectrum may contribute to the incidence of cancer or Alzheimer's disease[56].

Supplemental Information Allraw files and annotated spectra for single peptide protein identifications from these experiments are available on Chorus (Project ID 1114). Annotated spectra of proteins identified by a single peptide can be viewed on MS Viewer with the key

jow7i89mmv.

Acknowledgements This publication was supported by the National Center for Advancing Translational Sciences, through Grant 2UL1TR000442, and National Institutes of Health Grants R01 GM073929, R01 CA169046, P41 GM108538, P42 ES013661, and P30 ES05605. Core facilities were supported in part by the Holden Comprehensive Cancer Center, P30 CA086862. TJvE thanks The University of Iowa Graduate College for support. NMR gratefully acknowledges support from a National Science Foundation Graduate Research Fellowship (DGE-1256259).The authors thank Allison Momany, and Dee A. Even, Jessica Nichol, and Jamie L'Heureux (The University of Iowa) for their technical expertise on twin studies and zygosity testing; the Widness lab (The University of Iowa) and the Sysmex Corp. (Kobe, Japan) for the use of the XE-2100 and XT-2000 automated hematology analyzers (P01 HL46925); the staff of The University of Iowa DeGowin Blood Center in recruiting subjects and obtaining the blood samples; and the ESR Facility for invaluable assistance. The authors also thank Brett A. Wagner, Sean M. Martin, C. Michael Knudson, Robyn Blendowski, Gary R. Buettner, Kelli K. Ryckman, Benjamin W. Darbro, Jeffrey C. Murray, and Emery Bresnick for acquisition of data and thoughtful conversation.

References

[1] J. T. Alexander, A. M. El-Ali, J. L. Newman, S. Karatela, B. L. Predmore, D. J. Lefer, R. L. Sutliff, and J. D. Roback, "Red blood cells stored for increasing periods produce progressive impairments in nitric oxide-mediated vasodilation," *Transfusion*, vol. 53, pp. 2619–2628, nov 2013.

[2] S. L. Spitalnik, "Stored red blood cell transfusions: Iron, inflammation, immunity, and infection," *Transfusion*, vol. 54, pp. 2365–2371, jun 2014.

[3] Mary Keller, Jean Raymond, Wayne LaMorte, Frederick Millham, and Erwin Hirsch, "Effects of Age of Transfused Blood on Length of Stay in Trauma Patients: A Preliminary Report.," *The Journal of Trauma: Injury, Infection, and Critical Care*, vol. 53, no. 5, pp. 1023–1025, 2002.

[4] P. J. Offner, E. E. Moore, W. L. Biffl, J. L. Johnson, and C. C. Silliman, "Increased rate of infection associated with transfusion of old blood after severe injury," *Archives of Surgery*, vol. 137, pp. 711–717, jun 2002.

[5] R. L. Sparrow, "Time to revisit red blood cell additive solutions and storage conditions: A role for "omics" analyses," *Blood Transfusion*, vol. 10, pp. s7—-s11, may 2012.

[6] G. Zallen, P. J. Offner, E. E. Moore, J. Blackwell, D. J. Ciesla, J. Gabriel, C. Denny, and C. C. Silliman, "Age of transfused blood is an independent risk factor for postinjury multiple organ failure," *American Journal of Surgery*, vol. 178, pp. 570–572, dec 1999.

[7] J. R. Hess, "Red cell changes during storage," *Transfusion and Apheresis Science*, vol. 43, pp. 51–59, aug 2010.

[8] J. R. Hess, "Scientific problems in the regulation of red blood cell products," *Transfusion*, vol. 52, pp. 1827–1835, aug 2012.

[9] J. R. Hess, "Measures of stored red blood cell quality," *Vox Sanguinis*, vol. 107, pp. 1–9, jul 2014.

[10] M. Luten, B. Roerdinkholder-Stoelwinder, N. P. Schaap, W. J. De Grip, H. J. Bos, and G. J. Bosman, "Survival of red blood cells after transfusion: A comparison between red cells concentrates of different storage periods," *Transfusion*, vol. 48, pp. 1478–1485, jul 2008.

[11] K. Nakao, T. Wada, T. Kamiyama, M. Nakao, and K. Nagano, "A direct relationship between adenosine triphosphate-level and in vivo viability of erythrocytes," *Nature*, vol. 194, pp. 877–878, jun 1962.

[12] T. J. Reid, J. G. Babcock, C. P. Derse-Anthony, H. R. Hill, L. E. Lippert, and J. R. Hess, "The viability of autologous human red cells stored in additive solution 5 and exposed to 25 °C for 24 hours," *Transfusion*, vol. 39, pp. 991–997, sep 1999.

[13] T. E. Gilroy, G. J. Brewer, and C. F. Sing, "Genetic control of glycolysis in human erythrocytes," *Genetics*, vol. 94, pp. 719–732, mar 1980.

[14] T. J. Van'T Erve, C. M. Doskey, B. A. Wagner, J. R. Hess, B. W. Darbro, K. K. Ryckman, J. C. Murray, T. J. Raife, and G. R. Buettner, "Heritability of glutathione and related metabolites in stored red blood cells," *Free Radical Biology and Medicine*, vol. 76, pp. 107–113, nov 2014.

[15] S. Chakraborty and A. B. D. Chaudhuri, "Heritability of Some Important Parameters of the Antioxidant Defense System Like Glucose-6-Phosphate Dehydrogenase, Catalase, Glutathione Peroxidase and Lipid Peroxidation in Red Blood Cells by Twin Study," vol. 1, no. 1, pp. 1–4, 2001.

[16] T. J. Van 'T Erve, B. A. Wagner, S. M. Martin, C. M. Knudson, R. Blendowski, M. Keaton, T. Holt, J. R. Hess, G. R. Buettner, K. K. Ryckman, B. W. Darbro, J. C. Murray, and T. J. Raife, "The heritability of metabolite concentrations in stored human red blood cells," *Transfusion*, vol. 54, pp. 2055–2063, aug 2014.

[17] N. Mohandas and P. G. Gallagher, "Red cell membrane: Past, present, and future," *Blood*, vol. 112, pp. 3939–3948, nov 2008.

[18] R. Almizraq, J. D. R. Tchir, J. L. Holovati, and J. P. Acker, "Storage of red blood cells affects membrane composition, microvesiculation, and in vitro quality," *Transfusion*, vol. 53, pp. 2258–2267, oct 2013.

[19] I. A. Lewis, M. E. Campanella, J. L. Markley, and P. S. Low, "Role of band 3 in regulating metabolic flux of red blood cells.," *Pnas*, vol. 106, pp. 18515–18520, nov 2009.

[20] E. M. Pasini, M. Kirkegaard, D. Salerno, P. Mortensen, M. Mann, and A. W. Thomas, "Deep Coverage Mouse Red Blood Cell Proteome: A First Comparison with the Human Red Blood Cell," *Mol Cell Proteomics*, vol. 7, pp. 1317–1330, jul 2008.

[21] E. M. Pasini, H. U. Lutz, M. Mann, and A. W. Thomas, "Red blood cell (RBC) membrane proteomics - Part I: Proteomics and RBC physiology," *Journal of Proteomics*, vol. 73, pp. 403–420, jan 2010.

[22] A. Z. Amirkulova, G. Romanazzi, A. S. Utarbayeva, G. O. Rvaydarova, and O. V. Chebonenko, "Participation of antioxidant enzymes in defense reactions of wheat to the stressing environments," *Research Journal of Pharmaceutical, Biological and Chemical Sciences*, vol. 7, pp. 2651–2658, nov 2016.

[23] B. Barasa and M. Slijper, "Challenges for red blood cell biomarker discovery through proteomics," *Biochimica et Biophysica Acta - Proteins and Proteomics*, vol. 1844, pp. 1003–1010, may 2014.

[24] R. A. Zubarev, "The challenge of the proteome dynamic range and its implications for in-depth proteomics," *Proteomics*, vol. 13, pp. 723–726, mar 2013.

[25] S. R. Goodman, O. Daescu, D. G. Kakhniashvili, and M. Zivanic, "The proteomics and interactomics of human erythrocytes," *Experimental Biology and Medicine*, vol. 238, pp. 509–518, may 2013.

[26] J. H. Ringrose, W. W. Van Solinge, S. Mohammed, M. C. O'Flaherty, R. Van Wijk, A. J. Heck, and M. Slijper, "Highly efficient depletion strategy for the two most abundant erythrocyte soluble proteins improves proteome coverage dramatically," *Journal of Proteome Research*, vol. 7, pp. 3060–3063, jul 2008.

[27] G. M. D'Amici, S. Rinalducci, and L. Zolla, "Depletion of hemoglobin and carbonic anhydrase from erythrocyte cytosolic samples by preparative clear native electrophoresis," *Nature Protocols*, vol. 7, pp. 36–44, jan 2012.

[28] K. Walpurgis, M. Kohler, A. Thomas, F. Wenzel, H. Geyer, W. Schänzer, and M. Thevis, "Validated hemoglobin-depletion approach for red blood cell lysate proteome analysis by means of 2D PAGE and Orbitrap MS," *Electrophoresis*, vol. 33, pp. 2537–2545, aug 2012.

[29] F. Roux-Dalvai, A. Gonzalez de Peredo, C. Simó, L. Guerrier, D. Bouyssié, A. Zanella, A. Citterio, O. Burlet-Schiltz, E. Boschetti, P. G. Righetti, and B. Monsarrat, "Extensive Analysis of the Cytoplasmic Proteome of Human Erythrocytes Using the Peptide Ligand Library Technology and Advanced Mass Spectrometry," *Molecular & Cellular Proteomics*, vol. 7, pp. 2254–2269, nov 2008.

[30] N. M. Riley, A. S. Hebert, and J. J. Coon, "Proteomics Moves into the Fast Lane," *Cell Systems*, vol. 2, pp. 142–143, mar 2016.

[31] T. J. Van 'T Erve, B. A. Wagner, K. K. Ryckman, T. J. Raife, and G. R. Buettner, "The concentration of glutathione in human erythrocytes is a heritable trait," *Free Radical Biology and Medicine*, vol. 65, pp. 742–749, dec 2013.

[32] T. J. Van't Erve, B. A. Wagner, S. M. Martin, C. M. Knudson, R. Blendowski, M. Keaton, T. Holt, J. R. Hess, G. R. Buettner, K. K. Ryckman, B. W. Darbro, J. C. Murray, and T. J. Raife, "The heritability of hemolysis in stored human red blood cells," *Transfusion*, vol. 55, pp. 1178–1185, jun 2015.

[33] A. S. Hebert, A. L. Richards, D. J. Bailey, A. Ulbrich, E. E. Coughlin, M. S. Westphall, and J. J. Coon, "The One Hour Yeast Proteome," *Molecular & Cellular Proteomics*, vol. 13, pp. 339–347, jan 2014.

[34] M. Tvrdon, "The business cycle and unemployment: Empirical evidence from the visegrad group countries," *International Journal of Mathematical Models and Methods in Applied Sciences*, vol. 5, pp. 679–687, may 2011.

[35] J. Cox and M. Mann, "MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification," *Nature Biotechnology*, vol. 26, pp. 1367–1372, dec 2008.

[36] J. Cox, N. Neuhauser, A. Michalski, R. A. Scheltema, J. V. Olsen, and M. Mann, "Andromeda: A peptide search engine integrated into the MaxQuant environment," *Journal of Proteome Research*, vol. 10, pp. 1794–1805, apr 2011.

[37] J. Cox and M. Mann, "1D and 2D annotation enrichment: a statistical method integrating quantitative proteomics with complementary high-throughput data.," *BMC bioinformatics*, vol. 13 Suppl 16, p. S12, nov 2012.

[38] S. Tyanova, T. Temu, P. Sinitcyn, A. Carlson, M. Y. Hein, T. Geiger, M. Mann, and J. Cox, "The Perseus computational platform for comprehensive analysis of (prote)omics data," *Nature Methods*, vol. 13, pp. 731–740, jun 2016.

[39] P. Sandiford, "Twins; a Study of Heredity and Environment.," *Journal of Educational Psychology*, vol. 28, pp. 709–711, apr 1937.

[40] E. N. Pesciotta, S. Sriswasdi, H. Y. Tang, P. J. Mason, M. Bessler, and D. W. Speicher, "A label-free proteome analysis strategy for identifying quantitative changes in erythrocyte membranes induced by red cell disorders," *Journal of Proteomics*, vol. 76, pp. 194–202, dec 2012.

[41] M. E. Campanella, H. Chu, and P. S. Low, "Assembly and regulation of a glycolytic enzyme complex on the human erythrocyte membrane," *Proceedings of the National Academy of Sciences*, vol. 102, pp. 2402–2407, feb 2005.

[42] E. Y. Chen, C. M. Tan, Y. Kou, Q. Duan, Z. Wang, G. V. Meirelles, N. R. Clark, and A. Ma'ayan, "Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool," *BMC Bioinformatics*, vol. 14, p. 128, 2013.

[43] K. Silventoinen, S. Sammalisto, M. Perola, D. I. Boomsma, B. K. Cornes, C. Davis, L. Dunkel, M. De Lange, J. R. Harris, J. V. Hjelmborg, M. Luciano, N. G. Martin, J. Mortensen, L. Nisticò, N. L. Pedersen, A. Skytthe, T. D. Spector, M. A. Stazi, G. Willemsen, and J. Kaprio, "Heritability of Adult Body Height: A Comparative Study of Twin Cohorts in Eight Countries," *Twin Research*, vol. 6, pp. 399–408, oct 2003.

[44] J. R. Hess, N. Rugg, A. D. Knapp, J. F. Gormas, H. R. Hill, C. K. Oliver, L. E. Lippert, and T. J. Greenwalt, "The role of electrolytes and pH in RBC ASs," *Transfusion*, vol. 41, pp. 1045–1051, aug 2001.

[45] J. A. Cancelas, L. J. Dumont, L. A. Maes, N. Rugg, L. Herschel, P. H. Whitley, Z. M. Szczepiokowski, A. H. Siegel, J. R. Hess, and M. Zia, "Additive solution-7 reduces the red blood cell cold storage lesion," *Transfusion*, vol. 55, pp. 491–498, mar 2015.

[46] J. R. Hess, "An update on solutions for red cell storage," *Vox Sanguinis*, vol. 91, pp. 13–19, jul 2006.

[47] R. J. Dern and J. J. Wiorkowski, "Studies on the preservation of human blood. {IV}. {The} hereditary component of pre- and poststorage erythrocyte adenosine triphosphate levels.," *Journal of Laboratory and Clinical Medicine*, vol. 73, pp. 1019–1029, mar 1969.

[48] R. M. Johnson, G. Goyette, Y. Ravindranath, and Y. S. Ho, "Hemoglobin autoxidation and regulation of endogenous H2O2levels in erythrocytes," *Free Radical Biology and Medicine*, vol. 39, pp. 1407–1417, dec 2005.

[49] M. L. Harrison, P. Rathinavelu, P. Arese, R. L. Geahlen, and P. S. Low, "Role of band 3 tyrosine phosphorylation in the regulation of erythrocyte glycolysis," *Journal of Biological Chemistry*, vol. 266, pp. 4106–4111, mar 1991.

[50] C. Reitz, "Genetic diagnosis and prognosis of Alzheimer's disease: Challenges and opportunities," *Expert Review of Molecular Diagnostics*, vol. 15, pp. 339–348, mar 2015.

[51] L. A. Demetrius and D. K. Simon, "An inverse-Warburg effect and the origin of Alzheimer's disease," *Biogerontology*, vol. 13, pp. 583–594, dec 2012.

[52] L. A. Demetrius, P. J. Magistretti, and L. Pellerin, "Alzheimer's disease: The amyloid hypothesis and the Inverse Warburg effect," *Frontiers in Physiology*, vol. 6, jan 2015.

[53] P. P. Hsu and D. M. Sabatini, "Cancer cell metabolism: Warburg and beyond," *Cell*, vol. 134, pp. 703–707, sep 2008.

[54] E. M. Palsson-Mcdermott and L. A. J. O'Neill, "The Warburg effect then and now: From cancer to inflammatory diseases," *BioEssays*, vol. 35, pp. 965–973, nov 2013.

[55] L. A. Demetrius and J. Driver, "Alzheimer's as a metabolic disease," *Biogerontology*, vol. 14, pp. 641–649, nov 2013.

[56] J. A. Driver, "Inverse association between cancer and neurodegenerative disease: review of the epidemiologic and biological evidence," *Biogerontology*, vol. 15, pp. 547–557, dec 2014.

Chapter 3

THE CELLULAR ECONOMY OF THE *saccharomyces cerevisiae* ZINC PROTEOME

This chapter has been published and is reprinted with permission from:

Wang, Y., **Weisenhorn, E.M.**, MacDiarmid, C., Andreini, C., Bucci, M., Taggart, J., Banci, L., Russell, J., Coon JJ., Eide, D. *The cellular economy of the Saccharomyces cerevisiae proteome.* Metallomics. **2018**, *12*, doi: 10.1039/C8MT00269J.

Abstract

Zinc is an essential cofactor for many proteins. A key mechanism of zinc homeostasis during deficiency is “zinc sparing" in which specific zinc-binding proteins are repressed to reduce the cellular requirement. In this report, we evaluated zinc sparing across the zinc proteome of *Saccharomyces cerevisiae*. The yeast zinc proteome of 582 known or potential zinc-binding proteins was identified using a bioinformatics analysis that combined global domain searches with local motif searches. Protein abundance was determined by mass spectrometry. In zinc-replete cells, we detected over 2500 proteins among which 229 were zinc proteins. Based on copy number estimates and binding stoichiometries, a replete cell contains ~ 9 million zinc-binding sites on proteins. During zinc deficiency, many zinc proteins decreased in abundance and the zinc-binding requirement decreased to ~ 5 million zinc atoms per cell. Many of these effects were due at least in part to changes in mRNA levels rather than simply protein degradation. Measurements of cellular zinc content showed that the level of zinc atoms per cell dropped from over 20 million in replete cells to only 1.7 million in deficient cells. These results confirmed the ability of replete cells to store excess zinc and suggested that the majority of zinc-binding sites on proteins in deficient cells are either unmetalated or mismetalated. Our analysis of two abundant zinc proteins, Fba1 aldolase and Met6 methionine synthetase, supported that hypothesis. Thus, we have discovered widespread zinc sparing mechanisms and obtained evidence of a high accumulation of zinc proteins that lack their cofactor during deficiency.

Introduction

Zinc is an essential catalytic and/or structural cofactor for many proteins. Approximately 9% of genes in eukaryotic organisms and ~ 5% of prokaryotic genes encode proteins that bind zinc to become functional[1]. The abundance and importance of zinc-dependent proteins is reflected in the concept of the “zinc quota" . The zinc quota is defined as the amount of zinc in a cell grown under a given condition[2]. The “minimum zinc quota" is the lowest amount of zinc per cell that allows for optimal growth. The minimum zinc quota varies widely for different organisms and has been experimentally determined to be ~ 10^5 atoms of zinc per cell in *E. coli*, ~ 10^7atoms in yeast, and ~ 10^8 atoms in mammalian cells[2–5]. Many studies have indicated that the level of labile or exchangeable zinc in cells is very low and the great majority is tightly bound by the proteins that require this metal for function[6–8].Therefore, the minimum zinc quota is likely dictated by the number of zinc-binding sites on proteins that require the metal for optimal cellular physiology.

Organisms have evolved with many mechanisms of zinc homeostasis. During times of excess zinc exposure, these mechanisms limit uptake and promote efflux to maintain the intracellular metal content at tolerable levels[9,10]. They also control the generation of intracellular zinc stores in the form of either organellar or buffered cytosolic (*e.g.* metallothionein) pools that are available for later use[2,11]. During zinc deficiency, homeostatic mechanisms work to maintain zinc levels at the minimum zinc quota[12]. These mechanisms increase zinc uptake, decrease zinc efflux, and mobilize the release of zinc from intracellular

storage sites. An additional mechanism of zinc homeostasis during deficiency has been called “zinc sparing" , *i.e.* reducing the levels of specific zinc-binding proteins to decrease the cellular zinc requirement[13]. In some cases, reduced production of a zinc-dependent protein is compensated by increased synthesis of a zinc-independent paralog. In many bacterial species, for example, several zinc-dependent ribosomal subunits are repressed during deficiency and corresponding zinc-independent subunits are upregulated to take their place[14–16]. In this way, cells can reduce their total requirement for zinc and prioritize the distribution of the limited supply of this nutrient to more critical sites and functions.

In *Saccharomyces cerevisiae*, the Zap1 transcription factor is the central regulator of zinc homeostasis[12,17]. Zap1 is a transcriptional activator protein whose activity is low in zinc-replete cells and high in deficient cells. Zap1 increases the expression of many genes including those that encode zinc uptake transporters in the plasma membrane. In addition, Zap1 increases expression of organellar transporters that control the levels of zinc in intracellular compartments such as the vacuole and the endoplasmic reticulum. In addition to maintaining zinc homeostasis, Zap1 also regulates genes involved in adapting cellular processes to the challenges of zinc deficiency. For example, the *CKI1* and *EKI1* genes are induced by Zap1 to maintain phospholipid synthesis[18,19]. The *CTT1* catalase gene is also induced by Zap1 and this response is likely to eliminate the oxidative stress that arises during zinc deficiency[20,21].

Among the adaptive responses to zinc deficiency, Zap1 induces expression of the *TSA1* gene. Tsa1 is a dual function protein that acts as a peroxidase and as a “holdase" -type

protein chaperone[22]. Tsa1 function is essential for growth of zinc-deficient cells. Our initial studies indicated a role of the Tsa1 peroxidase activity in protecting zinc-deficient cells against an elevated level of reactive oxygen species[21]. Subsequently, we found that the Tsa1 protein chaperone activity was more important than the peroxidase function for zinc-deficient growth[23]. Zinc-deficient cells lacking Tsa1 chaperone activity accumulated elevated levels of stress-responsive protein chaperones, suggesting elevated unfolded protein accumulation and a corresponding heat shock response. Also consistent with this hypothesis, zinc-deficient *tsa1*Δ cells accumulated distinct cytoplasmic foci marked by the Hsp104 disaggregase chaperone. These foci resembled the “IPOD" compartment that accumulates in cells accumulating abundant unfolded proteins[24]. Our findings for Tsa1 suggested that zinc-deficient cells accumulate unfolded zinc-dependent proteins because they lack their metal cofactor needed for folding and stability. Under these conditions, the holdase function of the Tsa1 chaperone may stabilize zinc apoproteins and shield them from misfolding and aggregation until zinc supplies increase.

Zap1 also controls an important zinc sparing mechanism involving the abundant zinc-binding alcohol dehydrogenases Adh1 and Adh3[25]. Under zinc-replete conditions, the *ADH1* gene is expressed and its protein product accumulates to high levels. Under deficient conditions, the *ADH1* promoter is repressed by an intergenic noncoding regulatory RNA under the control of Zap1. This mechanism allows for a transcriptional activator to repress expression of a target gene. The less abundant Adh3 alcohol dehydrogenase is regulated in a similar manner. At the same time, Zap1 induces expression of the *ADH4* gene. *ADH4*

encodes an alternative alcohol dehydrogenase that accumulates to lower levels than Adh1 and whose activity is dependent on zinc or possibly iron[26,27]. While Adh1 and Adh3 require two zinc atoms per monomer for function, Adh4 requires only one metal ion and is not as highly expressed as Adh1, thereby providing for a net reduction in the zinc requirement of the cell during deficiency regardless of which metal it uses.

Zap1-independent mechanisms of zinc sparing in *S. cerevisiae* have also been discovered. For example, RNA polymerase I (RNAPI) is targeted for degradation in zinc-deficient cells[28]. In replete cells, RNAPI large subunit Rpa190 is ubiquitinated and retained in the nucleus and nucleolus where it transcribes ribosomal RNA genes. Under zinc-deficient conditions, Rpa190 becomes deubiquitinated, RNAPI is exported from the nucleus and taken up into the vacuole where it is degraded by vacuolar proteases. Similarly, the zinc-binding vacuolar alkaline phosphatase Pho8 is also targeted for degradation in zinc-deficient cells through a mechanism that is independent of Zap1[29]. The zinc released into the vacuole by these mechanisms is likely transported back to the cytosol where it is used by other proteins for their function.

In this study, we performed an analysis of the yeast proteome in replete cells and cells transitioning to conditions of severe zinc deficiency. A major goal of this study was to identify additional examples of zinc sparing in the yeast zinc proteome. A second goal was to test the hypothesis raised by our studies of Tsa1 that zinc-deficient cells accumulate high levels of zinc apoproteins that require protein chaperones for their stability in the absence of their metal cofactor. We describe an extensive catalog of predicted zinc proteins in the

yeast proteome and their abundance and distribution in zinc-replete conditions. During the transition to deficiency, we found that decreased accumulation of zinc-binding proteins is widespread and many effects are mediated at least in part by changes in mRNA levels rather than simply degradation of apoproteins. In addition, we present evidence that the accumulation of apoproteins is high during deficiency and that the majority of zinc sites in a cell are not occupied by zinc under these conditions. Thus, we provide unique insights into the economy of zinc in a eukaryotic cell.

Materials and Methods

Strains and growth conditions Yeast were grown in rich (YPD), synthetic defined (SD), or low zinc medium (LZM), as previously described[30]. LZM contains 20 mM citrate and 1 mM EDTA to buffer pH and zinc availability. Glucose (2%) was the carbon source for all experiments. LZM + 1 μ M $ZnCl_2$ was routinely used as the zinc-deficient condition, and LZM + 100 μ M $ZnCl_2$ as the replete condition. The yeast strains used in this work were BY4743 (*MATa/MATα his3/his3 leu2/leu2 ura3/ura3 lys2/LYS2 met15/MET15*), BY4742 (*MATa his3 leu2 ura3 lys2*), BY4741 (*MATα his3 leu2 ura3 met15*), and BY4741 *fba1DAmP* (Thermo Fisher Scientific)[31].

Cataloging the zinc proteome of yeast Using the approach described in Valasatava *et al.*, we created two libraries of Hidden Markov Model profiles: a library of zinc-binding Pfam domains, and a library of zinc-binding structural motifs[32–34]. The Pfam domain library was

created by merging two lists: first, a list of Pfam domains with known 3D structure that contain a zinc-binding site extracted from MetalPDB in which each of these domains could be associated with the residues responsible for zinc binding and with their positions within the domain sequence and second, a list of Pfam domains without a known 3D structure but annotated as zinc-binding obtained by text mining of the annotations in the Pfam database[32]. The procedure resulted in a set of 573 Pfam domains: 541 with an associated zinc-containing 3D structure, and an additional 32 annotated as zinc-binding domains. The library of zinc-binding structural motifs was created by splitting into fragments the zinc-binding sites stored in MetalPDB as of June 2017, as described in Rosato *et al.*[35]. Only one representative was kept for zinc-binding sites that, though found in different PDB structures, fall in the same position of the same protein domain. Sites that are not physiologically relevant based on literature inspection (*e.g.*, zinc-substituted structures, spurious sites) were manually removed from the dataset. This procedure resulted in a library of 6450 zinc-binding motifs derived from 2651 zinc-binding sites. An additional library of 339 zinc-binding motifs was compiled separately because the native metal ion of their corresponding proteins is still under investigation.

The zinc proteome of yeast was obtained by using the hmmscan tool to search each yeast sequence for the profiles contained in the two libraries[33]. A yeast sequence was identified as a potential zinc-binding site if at least one of the following conditions was verified: (A) the profiles of all the fragments of a given site matched the sequence with an *e*-value lower than 10^{-3} and the corresponding ligands are conserved in the sequence. (B) The profile of a

domain with associated ligands matched the sequence with an *e*-value lower than 10^{-5} and the ligands are conserved in the sequence. (C) The profile of a domain with no associated ligands matched the sequence with an *e*-value lower than 10^{-7}. These predictions were integrated by adding the proteins annotated as zinc-binding in the UniProt database[36]. In total, 571 yeast proteins were identified as zinc binding using the 6450 motifs. An additional set of 45 potential zinc-binding proteins were identified with the 339 input sites as able to bind zinc but for which the identity of the native metal ion is still unknown. For each of these proteins, the subcellular location and enzyme EC number were retrieved from UniProt. For each site contained in these proteins, the information on zinc function, number of zinc ions per monomer, and structural zinc-site classification was imported from the site in the library that yielded the best match in the search, and then manually checked[37].

Mass spectrometry analysis and protein copy number estimations Mass spectrometry for proteomics analysis was performed on four biological replicates each for times 0, 4, 8, and 12 h of zinc deficiency and three biological replicates for 16 h of deficiency. Frozen cell pellets of various cell counts were lysed in 300 µl cold methanol. Precipitated proteins were separated by centrifugation for 10 minutes at 13,400 *g* and 4 ° C. The protein pellet was resuspended in 200 µl lysis buffer (8 M urea, 20 mM TCEP, 80 mM chloroacetamide, 100 mM Tris pH 8) and diluted with 1 ml 100 mM Tris pH 8. Protein digestion was performed overnight with trypsin (4 µg) before centrifuging for 5 minutes at 10,000 *g*. The resulting supernatant was de-salted with Strata C18 solid phase extraction cartridges and quantified

using Pierce Quantitative Colorimetric Peptide Assay (Thermo Fisher Scientific). Peptides were then dried in a vacuum centrifuge before resuspension in 0.2% formic acid.

Samples were analyzed using a LC/MS instrument comprising an Orbitrap Fusion Lumos tribrid mass spectrometer (Thermo Fisher Scientific). Mobile phase A consisted of 0.2% formic acid in water and mobile phase B consisted of 0.2% formic acid in acetonitrile. A 75 minutes gradient ranging from 0% to 50% B was used spanning a total runtime of 90 minutes. Analytes were injected onto a 1.7-micron C18 column packed in-house to a length of 35 cm and heated to 60 ° C. Survey scans of peptide precursors were collected from 350–1350 Th with an AGC target of 1,000,000 and a resolution of 60,000 in the orbitrap followed by HCD MS/MS turbo scans taken in the ion trap.

The resulting LC-MS proteomic data were processed using Maxquant software version 1.5.2.8 and searched against a *Saccharomyces cerevisiae* database downloaded from Uniprot on 8/10/16. The digestion enzyme was set to trypsin with up to two missed cleavages, carbamidomethylation of cysteines as a fixed modification, and oxidation of methionines and protein N-terminal acetylation as variable modifications. The match between runs feature was utilized to decrease missing data values within the data set. Precursor mass tolerance was 20 ppm and product ions were searched at 4.5 ppm tolerances. Peptides were filtered to a 1% FDR and combined to protein groups based on the rules of parsimony. Protein copy number calculations were performed in Perseus using the Proteomic Ruler plugin[38]. This method uses the peak intensities of histone proteins, which are proportional to DNA content, to estimate protein abundance on a per cell basis. Statistically significant effects

were defined as those proteins with 1.5-fold or greater changes and a *p*-value of <0.05 after Benjamini–Hochberg correction. The proteomics data was deposited in CHORUS under Project ID number 1530. Hierarchical cluster was performed using Euclidean distance with Perseus[39]. Gene Ontology analysis of four clusters produced from hierarchical clustering was performed using DAVID[40].

Quantitative RT-PCR analysis Quantitative analysis of gene expression by RT-PCR was performed as previously described[41]. Assays were performed on three biological replicates. Gene expression was calculated relative to the average C_t values for three control genes (*TAF10*, *ACT1* and *CMD1*) selected from multiple candidate genes tested for their highly stable expression under the conditions used in our experiments (data not shown).

ICP-AES analysis Culture aliquots were washed three times in cold water, pelleted, and frozen in liquid nitrogen. Total zinc was determined using inductively-coupled plasma atomic emission spectrometry (ICP-AES) on six biological replicates for each timepoint. Cell pellets were desiccated by incubation at 60 ° C for 12–18 h and subsequent dry weights recorded. The dried yeast pellets were acid digested in 250 µ l OmniTrace 70% HNO_3 (EMD Chemicals) at 60 ° C for 12–18 h with 150–200 rpm orbital shaking. The acid lysates were then diluted to 5% HNO_3 with OmniTrace water (EMD Chemicals) and analyzed by ICP-AES (5100 SVDV, Agilent Technologies). The ICP-AES instrument was calibrated using National Institute of Standards and Technology (NIST)-traceable elemental standards

and validated using NIST-traceable 1577b bovine liver reference material. The detection range for zinc was 0.005–5 parts per million with inter-assay precision <10% . Cesium (50 ppm) was used for ionization suppression and yttrium (5 ppm) was used as an internal standard for all samples. All reagents and plasticware were certified or routinely tested for trace metal work. Zinc content was determined using native software (ICPExpert) and normalized to the cell number in each sample.

Protein extraction and immunoblotting Yeast protein extracts were prepared for immunoblotting with a TCA extraction protocol as previously described[42]. SDS-PAGE and immunoblotting was conducted using a Li-Cor Odyssey infrared dye detection system as previously described[42]. Antibodies used were anti-HA (12CA5, Roche product 11583816001), anti-Fba1 (a gift from Dr Magdalena Boguta), and anti-Pgk1 (Abcam product 22C5D8, lot # GR166098)[43]. IR 680 dye-labeled secondary anti-mouse antibody (product 680LT, lot # C30605-02) was obtained from Li-Cor.

Assay of Fba1 aldolase activity Fba1 enzyme activity was determined using either permeabilized yeast cells or cell lysates. An assay developed by Freire was modified for use with 96-well plates[44]. After harvest, cells were washed once with equal volume of ice-cold deionized water containing 1 mM EDTA, twice with ice-cold deionized water, and resuspended in 0.1 M MES-NaOH pH 6.5 buffer. Cell densities of each sample were normalized to A_{595} = 1.5, and 0.01% digitonin (w/v) was added to permeabilize the cells. After incubating at

30 ° C in a shaking water bath for 10 minutes, cells were placed on ice and then washed twice, resuspended with ice-cold 25 mM KH_2PO_4/K_2HPO_4 pH 7.4 buffer to a density of $A_{595} = 0.5$. A standard assay of aldolase activity contained 10 units of triose-phosphate isomerase (TPI), 4 units of glyceraldehyde 3-phosphate dehydrogenase (GAPDH) in a 25 mM KH_2PO_4/K_2HPO_4 pH 7.4 buffer containing 5 mM sodium arsenate, and 5 mM of NAD^+ and 20 µ l of permeabilized cells. The assay was started by adding 5 mM of the substrate, fructose 1,6-bisphosphate (FBP), and incubated at 30 ° C in a temperature-controlled plate reader. The absorbance at 340 mm of NADH generated by the assay was recorded at 1 minute intervals for up to 20 minutes. The recorded A_{340} values were graphed and the linear portion of the graph was used to calculate the rates of Fba1 aldolase activity using an NADH standard curve following normalization to the cell density (A_{600}) in each reaction. Alternatively, some assays were conducted using 1–10 µ g of protein lysate per reaction and activity was normalized to total protein level. Cell pellets were resuspended in 0.5 ml of 25 mM KH_2PO_4/K_2HPO_4 pH 7.4 buffer containing 1 mM PMSF and 1 × EDTA-free protease inhibitor mix (Roche) and transferred to 1.5 ml tubes. A 0.2 ml volume of glass beads was added and the cells disrupted by vortexing for 5 minutes at 4 ° C. The homogenate was centrifuged at 16,000 × g for 10 minutes at 4 ° C and the supernatant stored on ice. The protein concentration was determined using a Bio-Rad DC kit against a BSA standard. To determine specific activity of Fba1 (in nmol NADH per minute per µ g Fba1 protein), A_{340} absorbance values were converted to NADH concentration using a standard curve and resulting activity values (nmol NADH per minute per µ g total protein) were normalized to

Fba1-HA protein level as determined by immunoblotting, taking into account the value for Fba1 abundance in zinc replete cells that we determined by mass spectrometry (*i.e.* 1% of total protein). This normalization was not performed in cases where Fba1 abundance was known to be stable during the experiment (*e.g.*). For some experiments, zinc was stripped from Fba1 by treating permeabilized cells with 50 mM EDTA and 5 mM FBP in a 30 ° C shaking water bath for 30 minutes. The cells were then washed five times with ice-cold 25 mM KH_2PO_4/K_2HPO_4 buffer containing 1 μ M EDTA to prevent any zinc remetalation. Zinc was added back to aliquots of stripped cells by treating with 20 μ M $ZnCl_2$ in 30 ° C shaking water bath. The cells were then washed three times with ice-cold 25 mM KH_2PO_4/K_2HPO_4 buffer containing 1 μ M EDTA and resuspended at $A_{595} = 0.5$ for the assay.

Plasmid constructions All plasmids were constructed by homologous recombination in yeast. Plasmids expressing Fba1 with a triple repeat of the hemagglutinin antigen (HA) epitope fused to the C-terminus were constructed by PCR-amplification of wild-type or mutant *FBA1* promoter and coding sequence and fused to the HA tags and terminator of the low copy episomal plasmid YCp-ZRC1-HA digested with *Bam*HI and *Eco*RI. To generate a wild-type tagged plasmid, a fragment was amplified with the oligonucleotides Fbacds-ha (5'-TAGCCCGCATAGTCAGGAACATCGTATGGGTATAAAGTGTTAGTGGTACGGAAAGT-3') and Fba1pr5'(5'-CACGACGTTGTAAAACGACGGCCAGTGAATTCTGACGCAAGCCCTAAGAA-3') and used to gap repair YCp-ZRC1-HA. To introduce the H265A mutation, a 5'-

most fragment was generated using the oligonucleotides H-Arev (5′-TGGAATTCTTGGACA GTAGAACCGGAACCACCGGCGAAGACCAAGA-3′) and Fba1pr5′, and a 3′-most fragment with H-Acomp (5′-GGTGGTTCCGGTTCTACTGT-3′) and Fbacds-ha. The two fragments were combined with gapped YCp-ZRC1-HA for yeast transformation. To construct the H111A and E183A mutants, a PCR fragment was amplified from the corresponding mutant versions of pFL44L-FBA1 using the oligonucleotides Fba1pr5′ and Fbacds-ha prior to gap repair. All plasmids were fully sequenced to verify the mutations. Functionality of the wild-type epitope-tagged plasmid was verified by complementation of aldolase enzyme activity when expressed in an *fba1DAmP* mutant strain (data not shown). Fba1-3xHA accumulated to a similar level as untagged Fba1 when both forms were coexpressed and detected with anti-Fba1, and the tagged protein had similar specific activity to untagged (data not shown), indicating that the epitope tags had no major effect on stability or function. The Met6-HA construct and inactivated mutant versions were constructed similarly to Fba1-HA using gap repair. The wild-type Met6 plasmid (YCpMet6-HA) was constructed by amplifying Met6 from BY4742 genomic DNA using met6cds3′ (5′-CACGACGTTGTAAAACGACGGCCAGTGAATTCCACAGCCATTCAACTCAG-3′) and Met6cds3′ (5-TAGCCCGCATAGTCAGGAACATCGTATGGGTAATTCTTGTATTGTTCACGGA-3′) primers, followed by gap repair of YCp-ZRC1-HA digested with *Eco*RI and *Bam*HI. To construct the H655A mutant, a fragment containing the mutation was generated by PCR amplification of YCpMet6-HA using the oligonucleotides met65′-1 (5′-GGCTGACAAGGATTCTCT-3′) and met6H-A3′ (5′-CATCAGCATCCAAAGCCTTGATATGGTTTGGATCCAAGTCAGAGTA

ACAGAAAGCAGAGTGTATTTGAGTCTTGTT3′). This fragment was used to gap repair YCpMet6-HA digested with *Sac*I and *Bam*HI. To construct the D612A mutant, a fragment was amplified from YCpMet6-HA with the met65′-1 and met6DA3′ (CTTCTCTTAAAGCTG-GTTCAgCAACTTGGATAACCTTGATACCGGCAGCT) oligonucleotides, and a second fragment was amplified with the met6da5′ (TATCCAAGTTGcTGAACCAGCTTTAAGA-GAAGGTTTACCATTGAGAGAAGGTA) and met6cds3′ oligonucleotides. The two fragments were combined and used to gap repair YCpMet6-HA digested with *Sac*I and *Bam*HI. Wild-type and mutant plasmids were fully sequenced, and the effect of the mutations on activity was verified by complementation of the methionine auxotrophy of a *met6* deletion mutant.

NEM/PEG maleimide analysis To identify cysteine residues showing zinc-dependent reactivity with *N*-ethylmaleimide (NEM) *in vivo*, 5 ml cultures of yeast were grown in zinc-replete (LZM + 100 μ M $ZnCl_2$) or deficient (LZM + 1 μ M $ZnCl_2$) medium to log phase (A_{595} 0.3–0.4) and harvested by centrifugation. Cells were washed twice with ice-cold 1 × PBS + 1 mM EDTA and resuspended in 5 ml PBS + 1 mM EDTA. A solution of 1 M NEM in 100% ethanol was added to 5 ml of cells to give a final concentration of 5 mM NEM. For a negative control, the same volume of 100% ethanol was added to an identical aliquot of cells. After 30 min incubation at 30° with shaking, the cultures were harvested by centrifugation and washed twice with 1 × PBS. Protein was extracted using the TCA method and redissolved in buffer A (200 mM Tris base, 1% SDS, 1 mM EDTA)[42]. After

measurement of protein concentration (DC protein assay, Bio-Rad), aliquots of protein were processed to modify cysteines with mPEG-5 kDa (Sigma) as previously described[42]. Briefly, aliquots of 500 μ g protein were treated with 20 mM DTT for 10 min at 65 ° C to reduce disulfide bonds, then reprecipitated by adding 1/10 volume 100% TCA. Precipitated samples were centrifuged and washed twice with acetone to remove TCA, then redissolved in buffer B (100 mM Tris-Cl pH 7.4, 2% SDS, 1 mM EDTA) + 5 mM PEG-maleimide (mPEG). After overnight incubation at 30 ° C, 10–30 μ g of each protein sample was analyzed by SDS-PAGE and immunoblotting to determine the degree of mPEG modification of cysteine residues. To determine the degree to which cysteines were normally oxidized *in vivo* (and thus unavailable for reaction with NEM), some control samples were not treated with DTT prior to mPEG treatment.

Results and Discussion

Cataloging the zinc proteome of *S. cerevisiae* To determine the effects of zinc status on the abundance of zinc-binding proteins, we first cataloged the proteins that coordinate that ion. Proteins encoded by the yeast genome that are predicted to bind zinc were identified by a bioinformatics analysis that combined global domain searches with local motif searches. Based on this analysis, we identified 571 known or likely zinc-binding proteins in yeast. To this group, we added several transporter proteins (Zrt1, Zrt2, Fet4, Pho84, Zrt3, Zrc1, Cot1, Zrg17, Msc2) that have been implicated to transport zinc either exclusively or among known substrates. Also included was Zps1, an accessory protein for zinc uptake that is the

S. cerevisiae ortholog of *C. albicans* “zincophore" Pra1[45]. The yeast genome also encodes two metallothioneins, Cup1 and Crs5, that confer copper resistance. While both also bind zinc *in vitro*, only Crs5 contributes to zinc resistance when metal levels are high[46,47]. Therefore, Crs5 was included in the zinc proteome and Cup1 was not. These additions raised the total number of predicted zinc-binding proteins in yeast to 582. We refer to this catalog of known or likely zinc-binding proteins as the “zinc proteome" . Their number represents ~ 10% of the total yeast proteome and this is similar to the prevalence of zinc-binding proteins encoded by other eukaryotic genomes[1]. In addition, 45 other proteins were identified that may bind zinc based on structural data but these proteins lacked published references to support that zinc is the native metal ion. Therefore, we were unable to assess the functional relevance of zinc binding by members of this group.

The catalog of yeast zinc proteins is a useful tool to study the role of zinc in a eukaryotic cell. We examined the zinc proteome of yeast from a number of perspectives to characterize the diverse functions, types of zinc binding, and subcellular distributions of these proteins (Figure 1 A-D). summarizes classifications of the predicted zinc proteome based on the number of genes in each category that encode those proteins. When classified by the general function of the zinc cofactor, the most abundant group of these genes encode proteins with structural zinc sites (70%) followed by proteins that bind zinc as a catalytic cofactor (18%) (Figure 1A). Some genes, specifically members of the alcohol dehydrogenase family, encode proteins that bind two zincs with one serving a structural role and the other acting in catalysis; these comprise 2% of the total. The remainder encode zinc transport proteins

(2%) and genes whose protein products met our search criteria but the function of their metal cofactor has not yet been determined (8%).

Focusing on the 116 genes encoding proteins for which zinc plays a catalytic role (including those having both catalytic and structural zinc sites), enzymes in all six general enzyme classes were found (Figure 1B). The majority of these genes encode hydrolases with smaller numbers of the other classes represented. Similar distributions of genes in these enzyme classes has been found for the predicted zinc proteomes of other organisms[48]. For proteins in which zinc plays structural roles, several commonly shared motifs were observed (Figure 1C). These included the C_2H_2-like zinc fingers and Zn_2Cys_6 zinc fingers most often found in DNA-binding transcription factors[37]. Additional structural zinc-binding motifs observed included zinc ribbons, treble clef motifs, and zinc necklace domains[34]. Some proteins bind multiple zinc atoms and have more than one type of site.

While it has been long recognized that zinc-binding proteins are present in many different compartments, the distribution of specific proteins has not been determined for an entire zinc proteome. Because previous studies have determined the subcellular distribution of the majority of yeast proteins, we could assign subcellular locations to all but 8% of the predicted zinc proteome[49,50]. The majority of zinc proteome genes (74%) encode nuclear and/or cytosolic proteins (Figure 1D). Less common were genes encoding proteins found in the secretory pathway (ER, Golgi, endosome), mitochondria, vacuole, peroxisomes, plasma membrane, and cell wall.

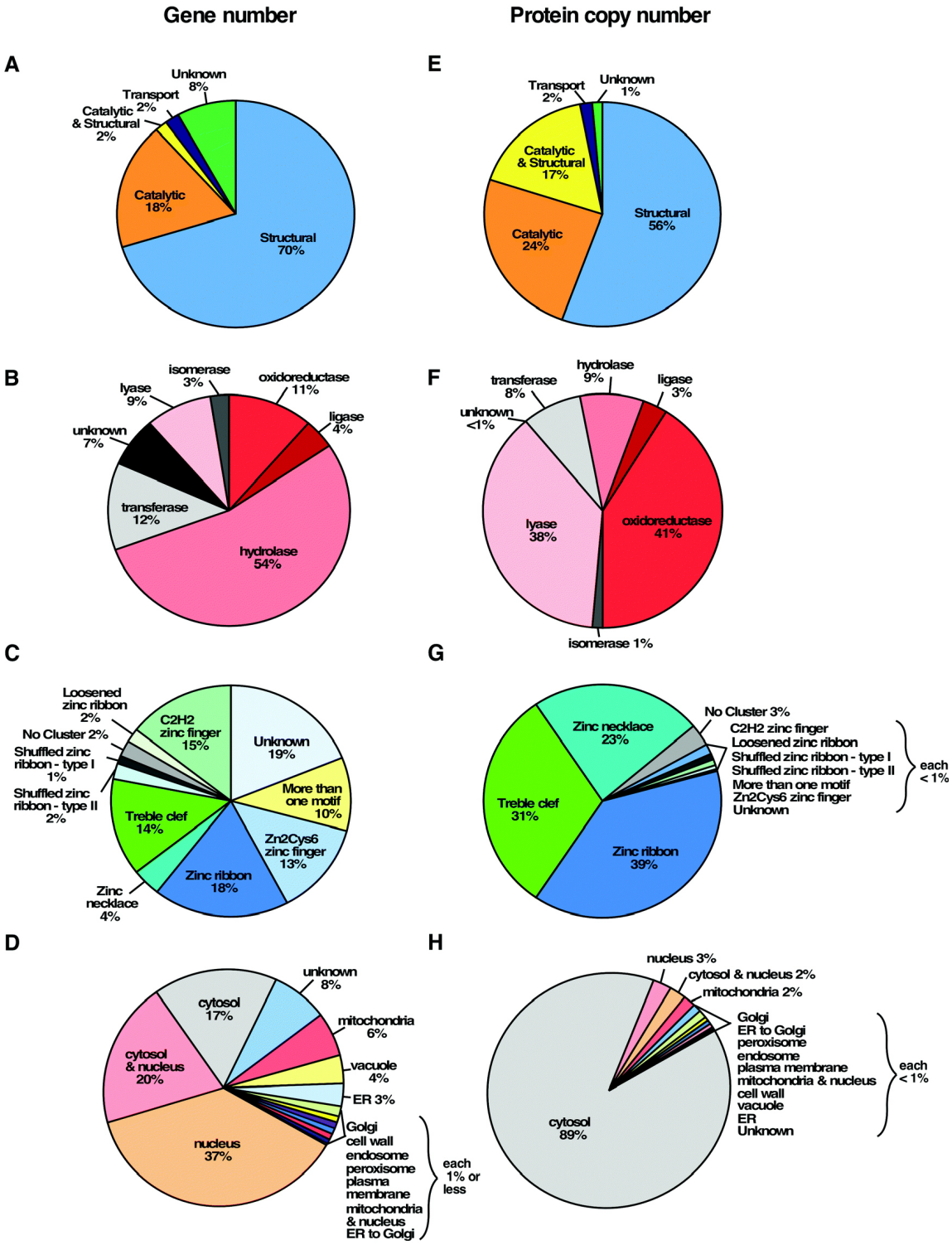


Figure 3.1: The zinc proteome of *S. cerevisiae*. The zinc proteome was classified based on general cofactor function (A and E), enzyme classification (B and F), structural zinc-binding site (C and G), and subcellular localization (D and H). In panels A–D, proportions were determined based on the number of genes encoding proteins in each group. In panels E–H, proportions were determined based on protein copy number in each group in replete cells.

Mass spectrometry analysis of the predicted zinc proteome in replete cells To estimate the absolute abundance of zinc-binding proteins in replete cells, quantitative label-free mass spectrometry was performed on total protein samples from cells grown to exponential phase under zinc-replete conditions (*i.e.* LZM supplemented with 100 μ M $ZnCl_2$)[51,52]. The resulting peak intensities were converted to protein copy number per cell using the "Proteomic Ruler" method[53]. Of the ~ 6000 total genes in yeast, we obtained copy number estimates of 2582 gene products in zinc-replete cells. The median coefficient of variation was 12% for each set of replicates demonstrating the high quality of the mass spectrometry data (Fig. S1, ESI). From these data, we estimated the total number of proteins per cell to be $\sim 9.2 \times 10^7$, which is similar to estimates made using other methods[54,55]. Of the 582 proteins in the predicted zinc proteome, we estimated the copy number of 229 (39%). These proteins added up to 7.6×10^6 zinc-binding proteins per cell or ~ 8% of the total protein number. Those zinc proteins not detected in this analysis likely represent proteins of low abundance and therefore would not contribute greatly to this estimate of total zinc protein number.

In zinc-replete cells, the abundance of zinc proteins delineated by general functional classifications was roughly similar to the distribution based on gene number (Figure 1E). The majority of zinc protein molecules in a cell had structural sites followed by enzymes that use zinc for catalysis. The proteins with both catalytic and structural sites made up a much larger fraction of the total zinc protein abundance (17%) than reflected by their gene number (2%) due to the high level of alcohol dehydrogenases (Adh1, Adh3, *etc.*). An

impact of the high abundance of alcohol dehydrogenases was also observed when grouping by enzyme class with oxidoreductases (including the alcohol dehydrogenases) making up a much larger proportion of protein copy number (41%) than was observed based on gene number (11%) (Figure 1F). Similarly, lyase protein copy number was more abundant (38%) than their gene number (9%) largely due to the high abundance of the glycolytic enzyme aldolase (Fba1). While hydrolases are very numerous at the gene level, they were consistently of low abundance and accounted for only 9% of detectable proteins that use zinc as a catalytic cofactor.

Considering structural motifs, zinc ribbons were a large fraction primarily due to the high number of ribosomal subunits with this motif (Figure 1G). The high abundance of alcohol dehydrogenases also had a large impact on the fraction of proteins with zinc necklace motifs. Also indicated by the data is the low relative abundance at the protein level of many other classes of zinc motifs. For example, many of the DNA-binding C_2H_2 and Zn_2Cys_6 zinc finger proteins were not detected by mass spectrometry suggesting their abundance is very low. Finally, when considering the subcellular distribution of zinc-binding proteins, the vast majority are found in the cytosol (Figure 1H). While cytosolic zinc proteins represent only 17% of genes, 89% of zinc protein abundance is cytosolic (not including the 2% of proteins with both nuclear and cytosolic distributions). The nucleus, mitochondria, and other compartments contain far lower levels of zinc-binding proteins.

The detectable zinc-binding proteins ranged from fewer than 10 to over 10^6 copies per cell. Most of these proteins were of relatively low abundance (<10,000 copies per

cell) and only a few were of very high abundance (>100,000 copies per cell) (Figure 2A). When these protein copy numbers were translated into number of zinc-binding sites based on their predicted stoichiometries of metal binding, it was clear that a small number of highly abundant proteins dominate the total zinc requirement of a replete cell (Figure 2B). In fact, the twenty most abundant zinc proteins accounted for almost 90% of the total zinc requirement of the cell (Figure 2C). These included Adh1 alcohol dehydrogenase, Fba1 aldolase, several zinc-binding ribosomal subunits, Sod1 superoxide dismutase, and Met6 methionine synthetase. Based on our analysis, the total number of zinc-binding sites in a replete cell was calculated to be 9×10^6. Notably, this number is very close to the experimentally determined minimum zinc quota of a yeast cell. Note that transporter proteins were not included in this estimate of zinc-binding sites because they only transiently interact with zinc. Unsurprisingly, neither Crs5 nor Cup1 metallothioneins were detected; these proteins are induced by copper treatment and are expressed at very low levels in untreated cells[56].

The response of the yeast proteome to zinc deficiency To determine the effects of zinc deficiency on the total proteome and specifically the zinc proteome of yeast, label-free quantitative proteomics was performed on total protein samples from cells transitioning from growth in zinc-replete to deficient conditions. Cells grown to exponential phase in a zinc-replete medium (LZM + 100 μ M $ZnCl_2$, *i.e.* the same cells sampled above) were transferred to a zinc-deficient medium (LZM + 1 μ M $ZnCl_2$) and cells were harvested

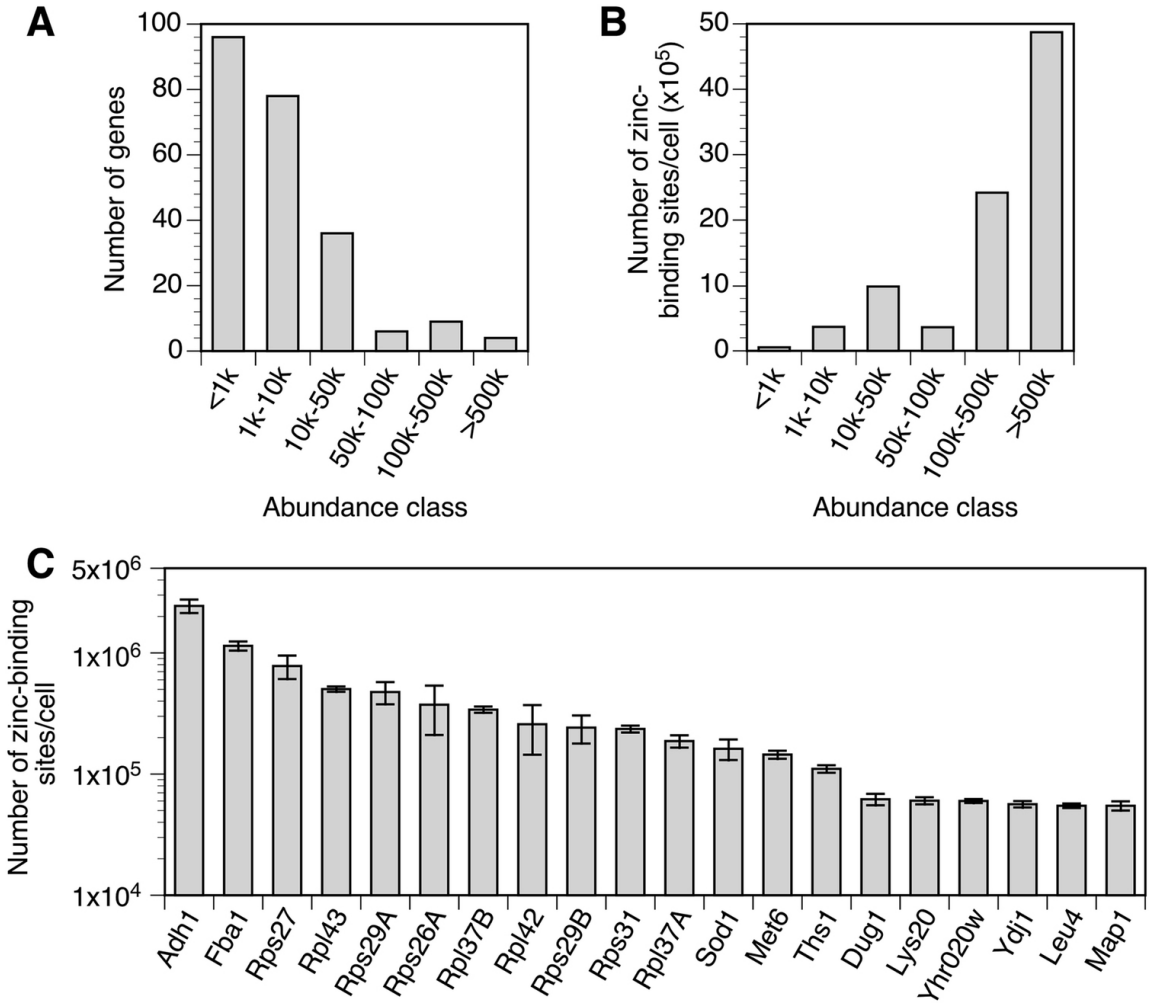


Figure 3.2: Abundance classes of the zinc proteome in replete cells.. (A) The number of genes encoding proteins in each abundance class is plotted. (B) The number of predicted zinc atoms bound by proteins in each abundance class is plotted. (C) The 20 proteins most abundant proteins in the zinc proteome are plotted with their predicted number of bound zinc atoms. Some ribosomal subunits are encoded by gene pairs, *e.g. RPS27A* and *RPS27B*, whose proteins are not distinguishable by mass spectrometry; these are labeled without specific reference to the paralogous genes. The error bars represent $\pm$ 1 S.D. calculated from 4 biological replicates

after 4, 8, 12, and 16 hours after that transition. These times correspond to approximately 2, 3, 4 and 5 generations of growth. LC-MS peak intensities were converted to protein copy number using the Proteomics Ruler method.

The effect of zinc deficiency on the total proteome is plotted in Fig. 3. Cluster analysis was performed to identify cohorts of similarly affected proteins and four clusters were identified. Proteins not included in these four clusters did not share significantly enriched gene ontology (GO) terms. Clusters 1 and 2 consisted of proteins that decreased in abundance during zinc deficiency with cluster 1 showing a larger fold decrease. Cluster 3 increased during zinc deficiency while cluster 4 showed a small increase or no change. Fig. 3 displays some of the significant GO terms that were found for these clusters. A complete list of GO terms is available in Table S4 (ESI). In cluster 1, down-regulated proteins were related to ribosomal proteins, ribosome biogenesis, and cytoplasmic translation. Cluster 2 included many terms found in cluster 1 including protein biosynthesis, ribosome biogenesis, translation, and also unfolded protein binding and chaperone proteins. These results suggest that zinc deficiency causes a systemic decrease in protein synthesis and protein chaperone capacity. Cluster 3 contains up-regulated proteins related to oxidation–reduction processes, mitochondrial function, and vesicle-mediated transport. Up-regulated proteins in cluster 4 are related to proteasome activity, actin binding, and glycolysis. Many of these changes may help mitigate the deficit of zinc.

Many of the effects of zinc deficiency on the total proteome are likely due to indirect consequences of metal status. To focus on responses that are directly controlled by zinc

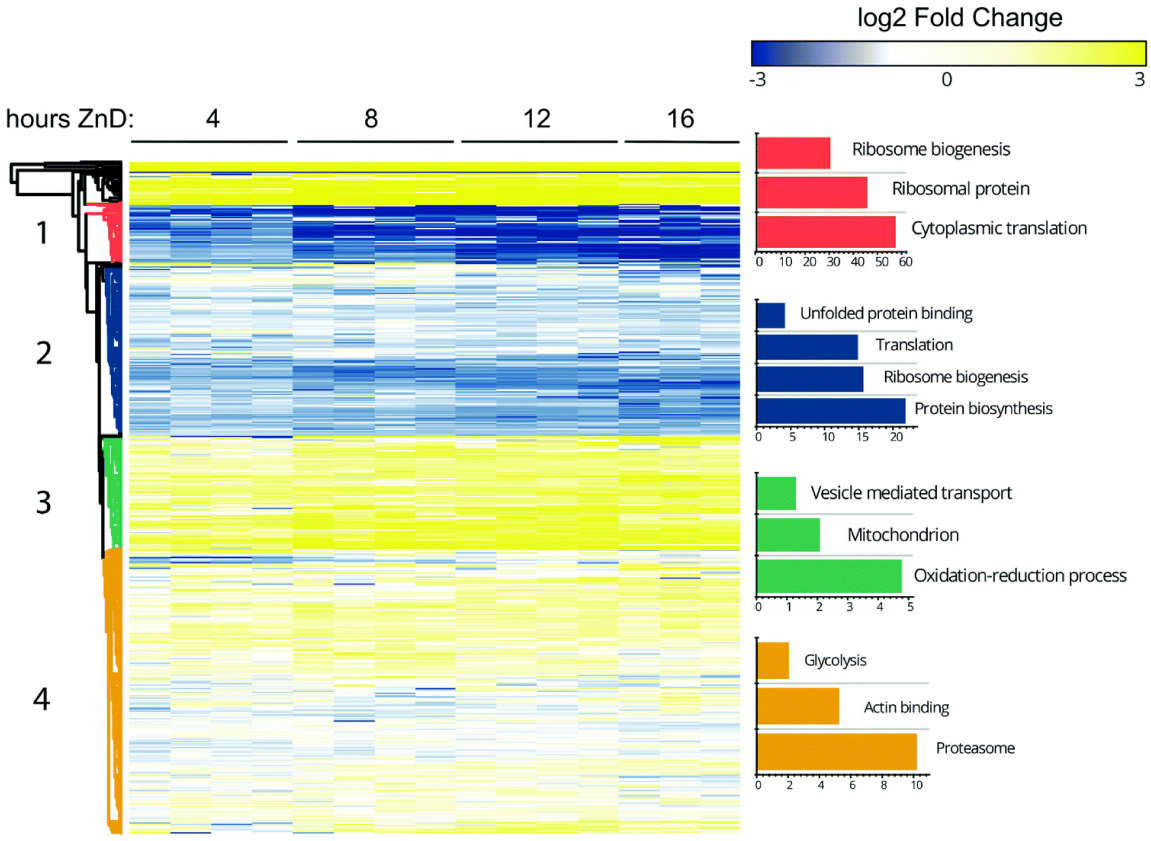


Figure 3.3: The response of the total proteome to the transition from zinc-replete to deficient conditions.. Zinc-replete cells were transferred to a zinc-deficient medium and harvested for proteomics analysis after 4, 8, 12, and 16 h. The resulting changes in protein abundance relative to replete conditions for 2119 proteins are plotted as a heat map. Cluster analysis was performed and four clusters (1–4) were identified. Gene ontology (GO) analysis of each cluster was then performed and significant terms observed are plotted with the $-$ log of their p-values in each histogram color-coded for their corresponding cluster.

status, we examined the abundance of proteins known or hypothesized to be regulated by the Zap1 transcription factor. Based on several published studies, Zap1 is thought to regulate transcription of 87 genes in response to zinc deficiency[20,57–62]. For the majority of those genes, including the zinc transporters *ZRT1*, *ZRT3*, and *FET4*, Zap1 activation increases gene expression and protein abundance during zinc deficiency. For a small number of other target genes, Zap1 represses gene expression and thereby reduces protein accumulation during deficiency. Of the 87 known or predicted Zap1-regulated genes, protein abundance was measurable for 39 of their encoded proteins and the effect of zinc deficiency on their levels is depicted in Fig. 4. For most Zap1 target genes, the abundance of their protein products increased in zinc deficiency, consistent with the action of Zap1 as a transcriptional activator. For some proteins, *e.g.* Ald3, Ctt1, Hsp26, and Adh4, the response to deficiency was immediate and strong. For other proteins, the response was much slower and/or of less dynamic range (*e.g.* Pep4 and Zrc1). In contrast, four proteins (Adh1, Adh3, Rad27, Hnt1) showed decreased protein abundance during the transition to deficiency. Adh1 and Adh3 were previously shown to be controlled by Zap1-regulated non-coding RNAs that repress promoter function[25]. The mechanisms controlling Rad27 and Hnt1 are currently being studied. These varied results demonstrate the diversity of responses of the proteins encoded by genes regulated by Zap1.

The response of the predicted zinc proteome to zinc deficiency Of the 229 zinc-binding proteins that were measured in replete cells, 199 of those had sufficient data (measurements

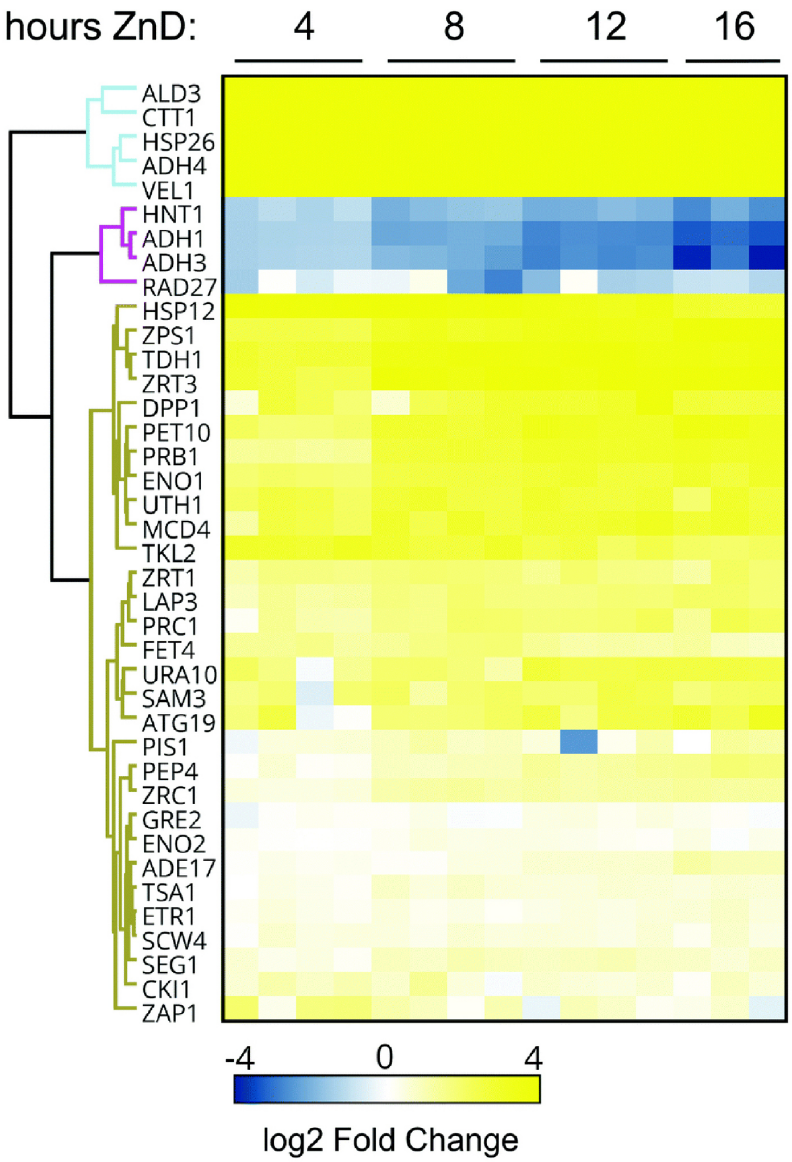


Figure 3.4: The response of proteins encoded by the Zap1 regulon to the transition from zinc-replete to deficient conditions.. During the transition to zinc-deficient conditions, the resulting changes in protein abundance relative to replete conditions for 39 proteins are plotted as a heat map. Proteins affected similarly were identified by cluster analysis.

at T_0 and at least two subsequent time points) to assess the effects of zinc deficiency on their abundance (Table S3, ESI). In addition, there were 13 zinc-binding proteins that were not detected at T_0 but were detected at two or more subsequent timepoints. Similarly, there were 30 zinc-binding proteins detectable at T_0 that were not detected at later time points suggesting a marked loss of protein abundance. As shown in Fig. 5, of the proteins in the predicted zinc proteome that were detected by mass spectrometry, many more proteins decreased in abundance during the transition from zinc-replete to deficient conditions than increased. Using a fold change of 1.5 or greater and a false discovery rate of 0.05, we found 31 proteins that increased in abundance and more than twice that number of proteins that decreased in response to zinc deficiency[63]. Among those proteins that increased in abundance were several zinc transport proteins (Zrt1, Fet4, Zrt3, Zrc1, Zps1) that are regulated by Zap1. The level of the Msc2 zinc transporter of the endoplasmic reticulum, while not Zap1 regulated, also increased in zinc-deficient cells[64]. Also among the increasing proteins was the Adh4 alcohol dehydrogenase, which is activated by Zap1 to likely replace the lost activity of the repressed Adh1 and Adh3 proteins. In addition, many proteins involved in chromatin modification (*e.g.* Rpd3, Hda1, Hst2, Set3, Gis1, Rsc3), components of the basal transcription machinery (*e.g.* Sua7, Brf1, Tfa1, Rpb9), and numerous proteases (*e.g.* Ape2, Ape3, Dpp3, Rpn11) increased in abundance in response to zinc deficiency. Furthermore, the gag-pol fusion proteins of transposable elements Ty1 and Ty2 increased. Increased levels of these Ty-encoded proteins have been observed under other stress conditions[65–67].

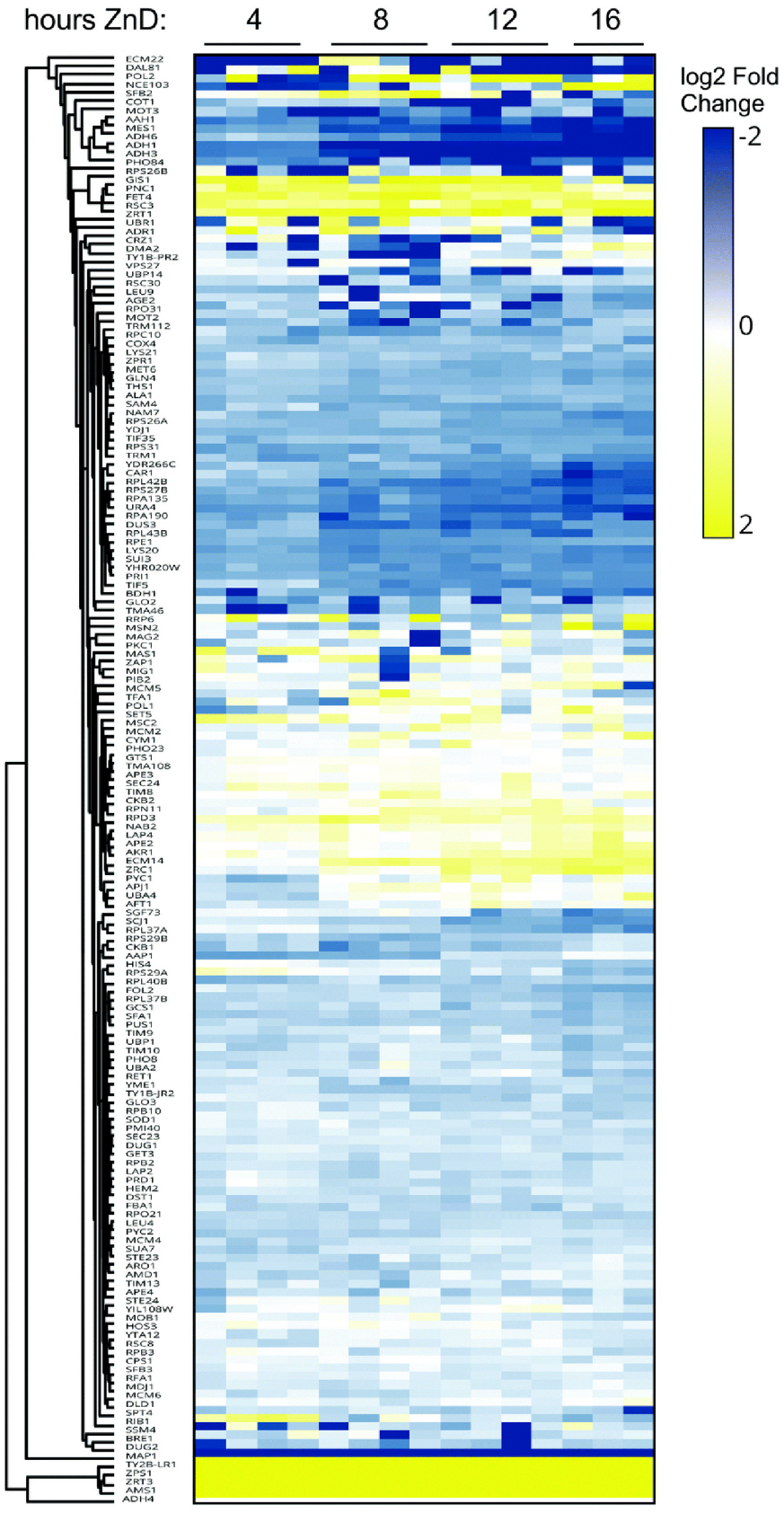


Figure 3.5: The response of proteins in the zinc proteome to the transition from zinc-replete to deficient conditions.. During the transition to zinc-deficient conditions, the resulting changes in protein abundance relative to replete conditions for 159 proteins are plotted as a heat map. Proteins affected similarly were identified by cluster analysis.

Many more zinc-binding proteins that we detected decreased in abundance during deficiency. These included zinc-binding ribosomal subunits and several ribosome biogenesis proteins (*e.g.* Bud20, Nmd3, Reh1), general translation factors (*e.g.* Tif5, Tif35, Sui3), tRNA synthetases and tRNA modifying proteins (*e.g.* Ths1, Mes1, Dus3), several RNA polymerase subunits and general transcription factors (*e.g.* Rpa12, Rpa135, Rpa190, Pri1, Taf1) and six alcohol dehydrogenases (Adh1, Adh3, Adh5, Adh6, Bdh1, Sfa1). The effect of zinc deficiency on the abundance of several specific examples is shown in Fig. 6. To determine if any of the observed changes in protein abundance were the result, at least in part, of altered transcription, we performed quantitative RT-PCR analysis of mRNA levels of 28 of the affected proteins in replete cells and in cells after 8 and 16 h of zinc deficiency. Those data are shown for specific examples in Fig. 6 and the full results are reported in. For 16 of the tested proteins, their decreased abundance during deficiency had some element of transcriptional control ($p < 0.05$). For example, the effect of Zap1 on *ADH1* and *ADH3* expression in zinc deficiency were observed in the RT-PCR results (Figure 6). Our analysis also suggested that the decreases of Adh5, Adh6, and Bdh1 alcohol dehydrogenases also occur at least in part at the transcriptional level.

In contrast, the mRNA encoding 12 of the proteins that decrease in zinc-deficient cells did not decrease during this transition. For example, *RPA190* mRNA was not reduced despite a large decrease in protein. This result is consistent with previous studies showing that Rpa190 is targeted for degradation specifically under zinc-deficient conditions[28]. Other examples of post-transcriptional effects newly identified in this analysis are Map1, Bud20,

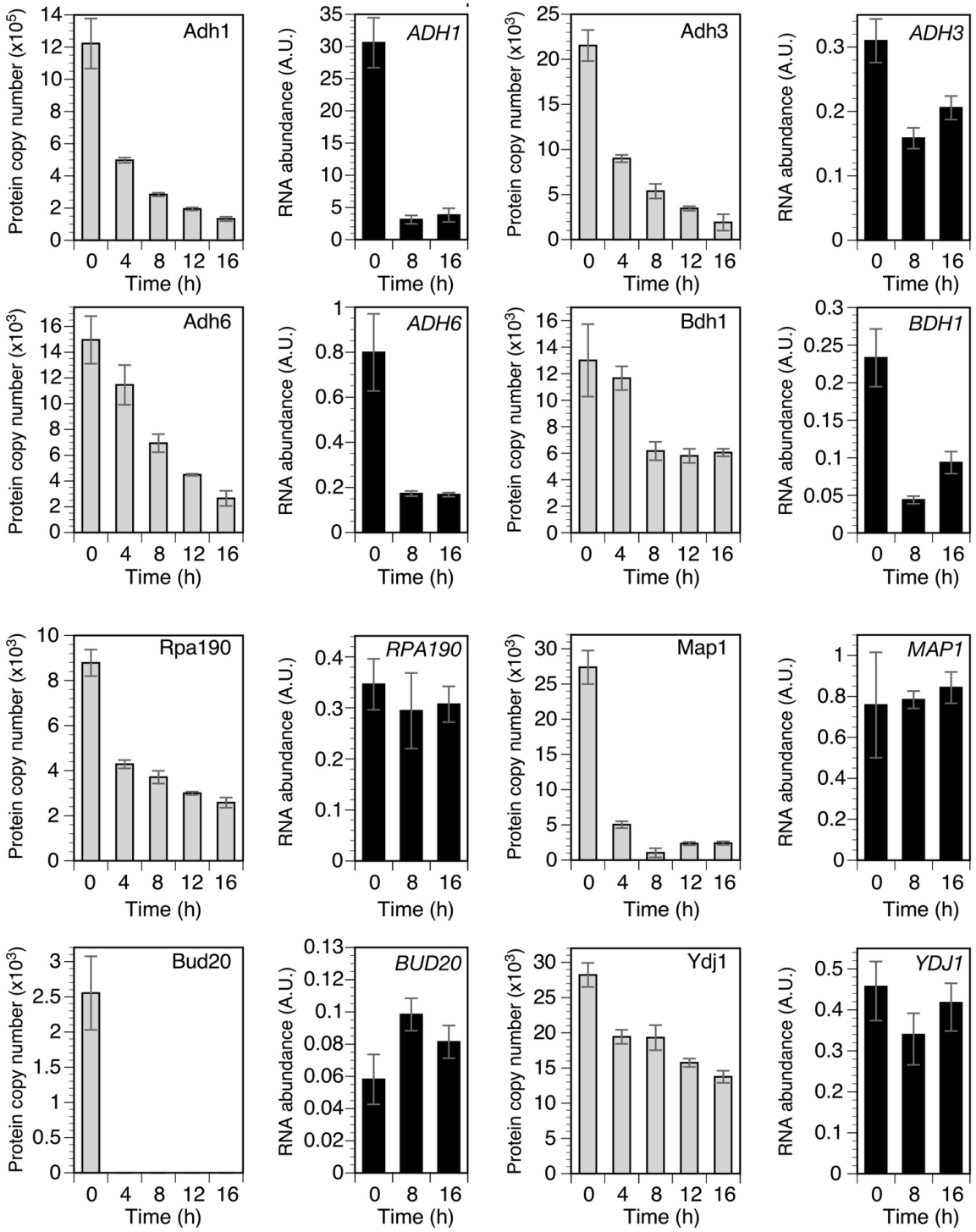


Figure 3.6: The effect of zinc deficiency on the abundance of example zinc proteins and their mRNA.. For each, the effect on the transition from replete to deficient conditions is plotted relative to protein copy number (left panels, gray columns) and mRNA abundance (right panels, filled columns, a.u. = arbitrary units). Protein levels are from 3–4 replicates per timepoint and the mRNA levels were determined by quantitative RT-PCR ($n = 3$). The error bars indicate $\pm$ 1 S.D.

	Protein copy number (per cell)					RNA level				
Protein/gene	0 h	8 h	16 h	Ratio of T_{16}/T_0	p-value	0 h	8 h	16 h	Ratio of T_{16}/T_0	p-value
Zrt1	29 467	104 388	99 222	3.36	0.001	1.444	7.064	11.028	7.64	0.00005
Zrt3	627	7351	8247	13.20	0.001	0.520	5.261	5.422	10.43	0.000009
Adh1	1 222 338	283 955	132 595	0.11	0.000006	30.58	3.11	3.81	0.12	0.04
Adh3	21 531	5370	1905	0.09	0.00009	0.31	0.16	0.21	0.66	0.002
Adh5	2659	0	0	N/A	N/A	0.53	0.24	0.36	0.67	0.01
Adh6	14 959	6939	2649	0.18	0.00005	0.80	0.17	0.17	0.21	0.0003
Bdh1	13 002	6165	6047	0.47	0.01	0.23	0.04	0.09	0.40	0.0005
Aah1	3344	1206	801	0.24	0.01	0.29	0.29	0.23	0.77	0.09
Bud20	2552	0	0	N/A	N/A	0.06	0.10	0.08	1.40	0.05
Car1	12 831	8155	4461	0.35	0.006	0.05	0.03	0.03	0.65	0.03
Dus3	3684	1471	1200	0.33	0.02	0.04	0.04	0.03	0.78	0.3
Fcy1	15 308	0	0	N/A	N/A	0.83	0.38	0.52	0.63	0.02
Map1	27 378	1034	2405	0.09	0.00007	0.76	0.78	0.84	1.11	0.6
Mes1	28 274	8847	6347	0.22	0.001	0.59	0.12	0.14	0.23	0.002
Met6	144 826	101 595	86 610	0.60	0.003	1.18	0.55	0.88	0.75	0.1
Nmd3	3559	1398	1720	0.48	0.01	0.51	0.25	0.26	0.50	0.003
Reh1	356	141	200	0.56	N/A	0.23	0.11	0.23	1.03	0.6
Rpa12	3567	0	0	N/A	N/A	0.72	0.28	0.36	0.51	0.001
Rpa135	5595	2543	1844	0.33	0.003	0.24	0.22	0.18	0.76	0.2
Rpa190	8784	3709	2580	0.29	0.0003	0.35	0.29	0.31	0.89	0.2
Rpe1	13 710	6477	6076	0.44	0.01	0.14	0.12	0.12	0.81	0.06
Sui3	27 585	12 387	11 510	0.42	0.009	1.26	0.38	0.51	0.40	0.0002
Ths1	55 203	36 381	31 764	0.58	0.005	1.40	0.49	0.82	0.59	0.0001
Tif35	24 715	14 869	12 549	0.51	0.006	3.65	1.34	1.76	0.48	0.0003
Tif5	35 778	18 271	17 712	0.50	0.005	0.43	0.23	0.26	0.61	0.03
Ura4	23 647	11 292	8225	0.35	0.006	0.83	0.13	0.22	0.26	0.0007
Ycr087c-A	2431	0	0	N/A	N/A	0.30	0.15	0.12	0.40	0.000008
Ydj1	28 211	19 306	13 749	0.49	0.002	0.46	0.34	0.42	0.91	0.3
Yhr020w	59 964	29 010	26 517	0.44	0.002	0.36	0.22	0.29	0.80	0.2
Zpr1	5310	4095	3285	0.62	0.02	0.20	0.21	0.16	0.77	0.1

Figure 3.7: Protein copy number and RNA abundance for example zinc proteome members under zinc-replete (0 h) and deficient (8 and 16 h) conditions.. Protein copy numbers are from mass spectrometry analysis and mRNA abundance was determined by quantitative RT-PCR (n = 3). The ratios of levels at 16 h and 0 h are reported and the p-values were calculated using Student's t-test

and Ydj1. These data indicated that there is a transcriptional component to the decreased accumulation of many, but not all, zinc proteins during the transition to zinc deficiency.

Estimates of *in vivo* zinc-binding site number and zinc content suggest significant deficits in zinc metalation during deficiency One purpose of the widespread decrease in zinc-binding proteins during deficiency may be zinc sparing, *i.e.* the regulated decrease in the zinc requirement of the cell (Figure 7A). shows the relative zinc-sparing effects for zinc-binding proteins grouped by their general function. Decreases in ribosome subunits and ribosomal biogenesis factors and decreases in alcohol dehydrogenase abundance reduce the zinc requirement to the greatest degree while other functional groups contribute to a lesser extent. The hypothesis of targeted zinc sparing suggested that the decrease in zinc proteins is significantly greater than the effects observed for the total proteome. During the transition to zinc deficiency, the total proteome decreased 24% from 9.2×10^7 copies per cell to 7.0×10^7copies (Figure 7B). Similarly, the copy number of non-zinc proteins also decreased 24% from 8.4×10^7 copies per cell to 6.5×10^7. These decreases may reflect the decreased translation capacity of zinc-deficient cells (Fig 3) and/or the results of autophagic degradation, which was previously shown to be induced by zinc deficiency[68–70]. In contrast, the effect of zinc deficiency on the zinc proteome was even more striking with protein copy number dropping from 7.5×10^6 to 4.5×10^6 copies per cell (39%). While regulation of specific zinc proteins during the transition to deficiency may occur for many reasons, these data suggest that widespread zinc-sparing mechanisms are at work to specifically decrease

the zinc demand of a cell during deficiency.

Factoring in their predicted stoichiometries of zinc binding, the abundance of zinc proteins in replete cells translates into 9×10^6 zinc-binding sites on proteins per cell (Figure 7C). By 16 h after the transition to zinc deficiency, the number of zinc sites decreased to 4.7×10^6 sites. Thus, the zinc requirement was found to drop after 16 h of zinc deficiency by over 4×10^6 zinc sites per cell or $\sim$ 45% of the sites found in replete cells. To determine the amount of zinc available in deficient cells to meet this demand, zinc abundance was determined using ICP-AES. Total zinc in replete cells grown under these conditions was measured to be 2.3×10^7 atoms per cell (Figure 7C). This value is approximately twice the estimated number of zinc-binding sites on proteins. By 16 h of zinc deficiency, zinc levels dropped to only 1.7×10^6 atoms per cell, *i.e.* $\sim$ 30% of the estimated number of zinc-binding sites on proteins. These results suggested that despite the large decrease in zinc protein number described above, a zinc-deficient cell grown under these conditions accumulates a very high level of apoproteins or, alternatively, zinc proteins that are mismetalated with some other cation.

Analyses of zinc-binding proteins indicate reduced zinc metalation during deficiency

To assess to what degree zinc metalation is disrupted during deficiency, we focused our analysis on the Fba1 aldolase protein. This glycolytic enzyme is the second most abundant protein in the zinc proteome of replete cells and the most abundant zinc protein in deficient cells (Figure 2). Fba1 levels were similar in replete and deficient cells with approximately

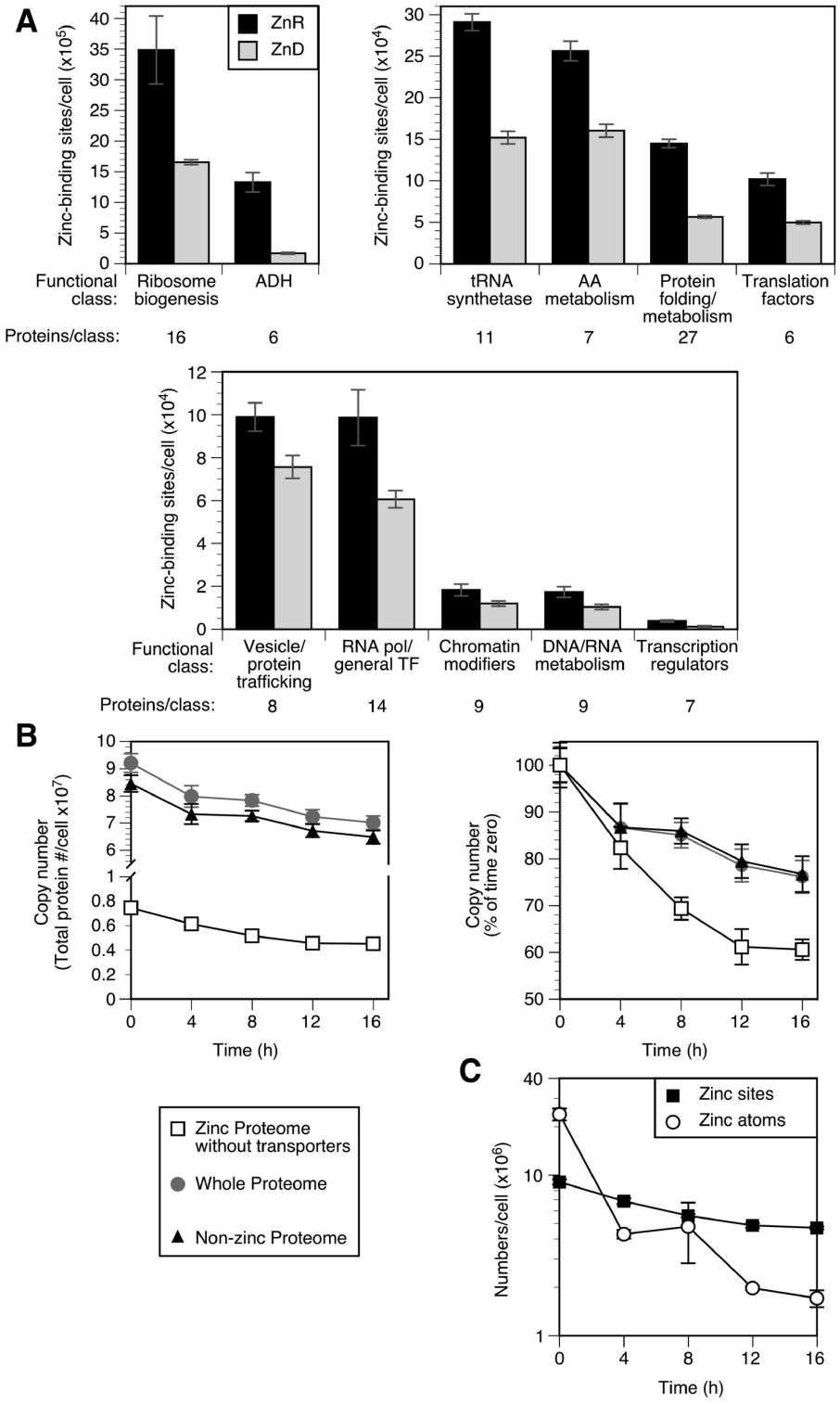


Figure 3.8: Zinc sparing is widespread in the zinc proteome.. (A) 120 members of the zinc proteome that decrease during zinc deficiency were grouped by general function and plotted as the number of zinc-binding sites in each group under replete (filled columns) and deficient (*i.e.* 16 h) (gray columns) conditions. (B) The effect of zinc deficiency on the copy number per cell of the total proteome (circles), the total proteome not including the zinc proteome (triangles), and the zinc proteome (squares) is plotted. (C) The estimated number of zinc atoms bound by the zinc proteome (squares) is plotted relative to the number of zinc atoms per cell as determined by ICP-AES (circles). The error bars indicate $\pm$ 1 S.D. calculated from 3–4 biological replicates.

1.1×10^6 copies per cell. Thus, the zinc requirement of this enzyme alone is over half of the entire zinc content of a deficient cell. Given these factors, we hypothesized that most Fba1 molecules are either unmetalated or mismetalated during zinc deficiency. To test this hypothesis, we used an enzymatic assay of aldolase activity in permeabilized cells. We confirmed the specificity of the assay for Fba1 aldolase activity using zinc-replete wild-type cells and *fba1DAmP* mutant cells that have lower *FBA1*expression[31]. Immunoblots indicated that the *fba1DAmP* strain accumulated ~ 25% of the wild-type level of protein and the measured enzyme activity showed a similar decrease (Figure 8A). *FBA1* is essential for viability so a null allele could not be used for this experiment.

To determine the effect of zinc status on Fba1 function, aldolase specific activity normalized to Fba1 protein levels as determined by immunoblotting, was measured in cells after growth in zinc-replete and deficient conditions (Figure 8B). Aldolase specific activity was high in zinc-replete cells and greatly reduced in deficient cells. These results were consistent with our hypothesis of reduced zinc metalation during deficiency. To further test this hypothesis, we determined whether zinc added back *in vitro* restored aldolase activity. EDTA treatment of permeabilized zinc-replete cells reduced aldolase activity and zinc subsequently added back restored activity to the full activity of untreated cells (Figure 8C). In contrast, zinc addition to EDTA-stripped deficient cells for 5, 15, or 30 minutes did not restore activity to replete levels. These data suggest that zinc-deficient cells contain a pool of inactive aldolase that cannot be metalated efficiently with zinc. This conclusion was further supported by analysis of Fba1 activity following zinc addition *in vivo*. When

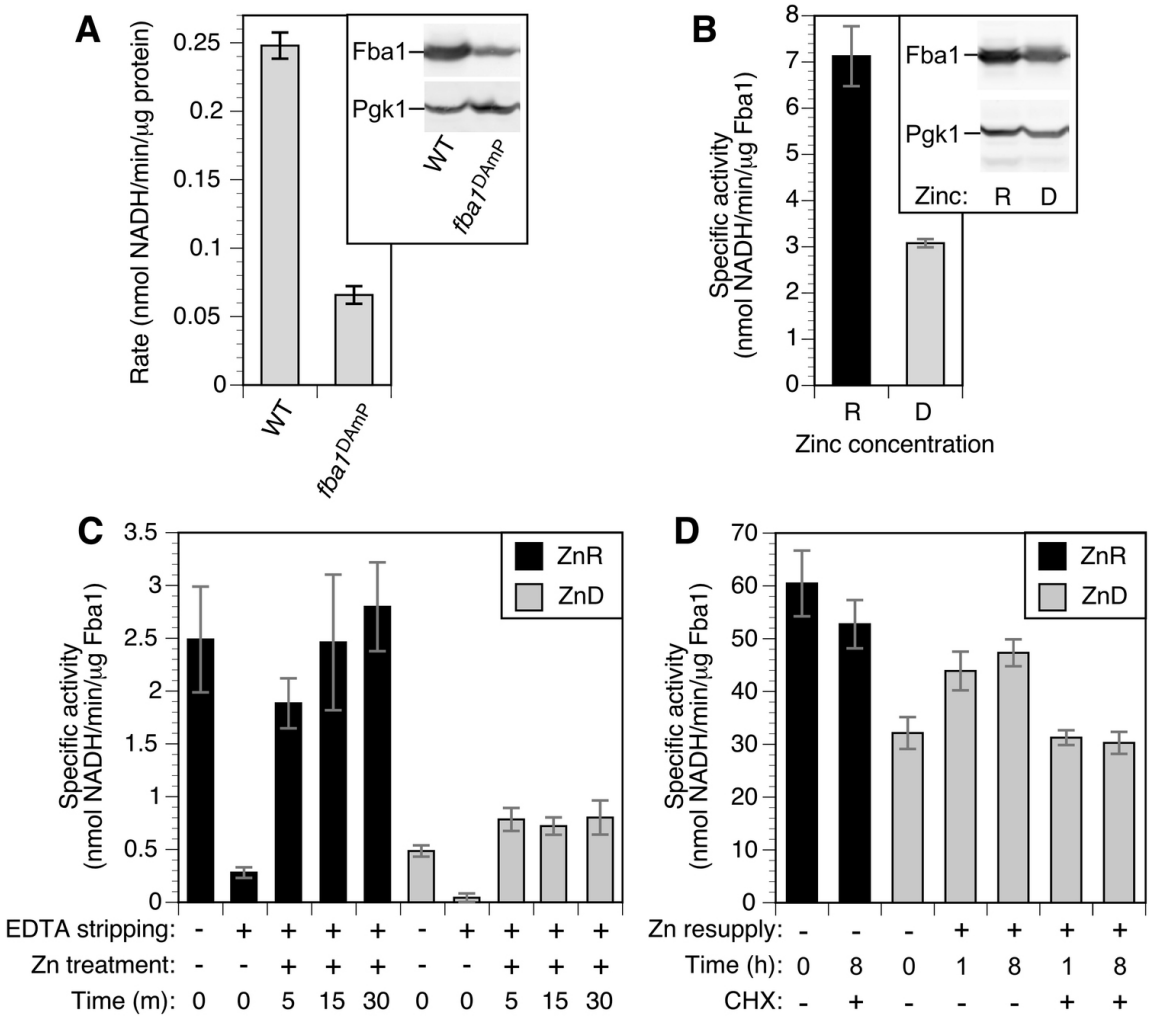


Figure 3.9: Effects of zinc status on Fba1 aldolase activity.. (A) Aldolase activity in wild-type (BY4741) and *fba1DAmP* cells normalized to total protein. The inset shows an immunoblot of Fba1 protein level detected with anti-Fba1 and anti-Pgk1 in 20 μ g of total protein. (B) Aldolase specific activity in zinc-replete (R) and deficient (D) BY4741 cells normalized to Fba1 protein levels. The inset shows an immunoblot of Fba1 protein level detected as described for panel A. (C) Aldolase specific activity in permeabilized replete (ZnR) and deficient (ZnD) cells before and after *in vitro* EDTA stripping and zinc resupply for the times indicated. (D) Aldolase specific activity in zinc-replete and deficient cells following zinc resupply *in vivo* with and without cycloheximide treatment (100 μ g ml^{-1}). The error bars indicate $\pm$ 1 S.D. ($n = 3$).

zinc-deficient cells were supplemented with zinc *in vivo* prior to permeabilization, much of the missing Fba1-normalized activity was recovered after 1 h of zinc treatment and ~ 80% of full activity was observed after 8 h (Figure 8D). However, this restoration of activity was completely blocked when the translation inhibitor cycloheximide was added prior to zinc addition. These results suggested that the inactive Fba1 aldolase in zinc-deficient cells cannot be activated with zinc supplementation and new protein synthesis is required to restore full aldolase activity.

To assess the metalation state of aldolase *in vivo*, we adapted a method previously used to probe zinc binding by proteins *in vitro*[71–73].*N*-Ethylmaleimide (NEM) is a cell-permeable reagent that alkylates free cysteine thiol groups in proteins. Reactivity of thiols with NEM is greatly reduced by metal binding to those residues or, for non-ligand cysteines, when the thiol group is inaccessible to solvent. Thus, reactivity to NEM can be used to assess both zinc binding to cysteine ligands and zinc-dependent conformation changes that affect the accessibility of non-ligand cysteines. In our modified procedure, cells growing in culture are treated with NEM to modify reactive cysteine thiols *in vivo*, lysed, proteins denatured, and then treated with PEG-maleimide, which modifies unreacted thiols that had been protected from *in vivo* NEM modification. Thiols alkylated by NEM are no longer reactive to PEG-maleimide whereas sites that were not alkylated become modified with the PEG reagent. While NEM modification only slightly alters protein molecular mass and is not detectable by immunoblotting, PEG-maleimide modification increases the apparent molecular mass on SDS-PAGE by ~ 15 kDa per moiety and is therefore readily detected by

immunoblotting. Thiols that are oxidized *in vivo* are also not modified by NEM; these can be identified as being protected from NEM modification *in vivo*, reactive to PEG-maleimide *in vitro* following DTT treatment, but not reactive with PEG-maleimide without pre-treatment with reductant.

Fba1 contains five cysteine residues. While none of these are directly involved in zinc binding, we found that their accessibility to modification by NEM is altered by zinc status. Without modification, Fba1 tagged with the hemagglutinin antigen epitope (Fba1-3xHA) migrates at ~ 40 kDa on SDS-PAGE (Figure 9A). When treated with PEG-maleimide alone following denaturation, the Fba1-HA protein shifts in mobility to ~ 115 kDa, corresponding to the addition of five PEG moieties. When zinc-replete cells were treated with NEM *in vivo* prior to PEG-maleimide treatment *in vitro*, 1–4 cysteines were protected from NEM reaction to varying degrees with most copies having two NEM-resistant cysteines. In zinc-deficient cells, NEM sensitivity increased as indicated by the decreased abundance of 2–4 × PEG-modified forms and increased abundance of proteins with 0–1 × PEG moieties added. This result indicates a zinc-dependent conformational change *in vivo* that changes the sensitivity of one or more cysteines to NEM. The effect on NEM sensitivity was observed both with and without DTT treatment confirming that protection *in vivo* was not due to thiol oxidation. To quantify these effects, band intensities were measured and the ratio of 0–1 × PEG-modified forms *vs.* 2–4 × PEG-modified forms was determined (Figure 9B). This analysis demonstrated higher levels of PEG modification (*i.e.* NEM resistance) in replete cells and lower levels (*i.e.* NEM sensitivity) in deficient cells. In contrast, no effect

of zinc was observed on the NEM sensitivity of the single cysteine in Pgk1; this cysteine is protected from NEM modification *in vivo* due to solvent inaccessibility but is modified with PEG-maleimide following protein denaturation (Figure 9A). Modification of Pgk1 with PEG-maleimide is highly efficient and provides a positive control for this reaction.

To assess whether the conformational change observed for the wild-type Fba1-3xHA protein was due to zinc binding *in vivo*, we mutated the histidine zinc ligand at position 111 to alanine (H111A). No zinc-responsive changes in NEM reactivity were then observed (Figure 9C). A similar analysis of a mutation in glutamate 183 (E183A) supported this conclusion. The E183A mutation disrupts enzyme activity without affecting metal binding. Despite its lack of activity, the same zinc-dependent conformational change observed for the wild type protein was seen with this mutant (Figure 9D). To further confirm that the observed conformational changes were due to the presence or absence of zinc binding, we performed the experiment in permeabilized replete cells using stripped and re-metalated Fba1-3xHA. Without EDTA stripping, PEG-maleimide reactivity was high indicating NEM resistance (Figure 9E). When Fba1-3xHA was stripped of metal with EDTA treatment, NEM resistance was reduced and re-addition of zinc restored protection from NEM modification. These effects are quantitated in Figure 9F and the reciprocal effect of EDTA and zinc on Fba1 enzyme activity is shown. Again, no effect of EDTA or zinc treatment was observed on the PEG sensitivity of Pgk1. These results support the conclusion that the change in NEM sensitivity is directly due to zinc binding in the Fba1 active site.

To test whether the *in vivo* dependence of Fba1 reactivation on protein synthesis was

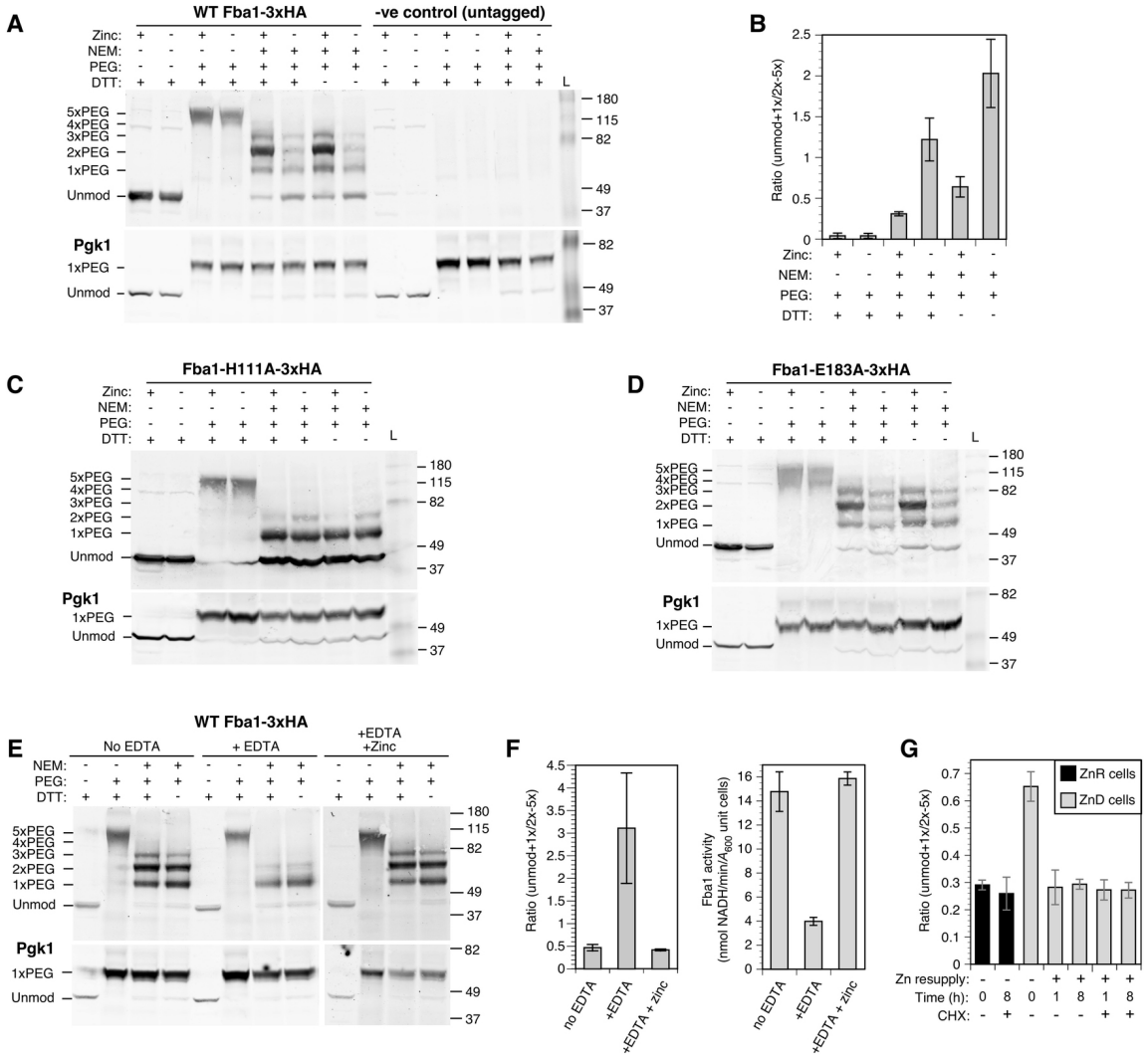


Figure 3.10: ***In vivo*** **analysis of zinc binding by Fba1..** (A) NEM/PEG-maleimide analysis of wild-type (BY4742) cells expressing Fba1-3xHA or an untagged Fba1 allele. Zinc-replete (+) or deficient (−) cells were treated with and without NEM, proteins harvested, and then treated with or without DTT prior to PEG-maleimide treatment. The positions of molecular mass markers (kDa) and of unmodified and PEG-modified forms of Fba1-3xHA are indicated. Pgk1, which has a single cysteine that is insensitive to zinc supply, was used as a control for the efficiency of PEG labeling. (B) Quantitation of the results in panel A. The ratios of unmodified + 1 × PEG-modified to 2× –5× PEG-modified forms are shown and the error bars indicate ± 1 S.D. ($n = 3$). (C and D) NEM/PEG-maleimide analysis of wild-type (BY4742) cells expressing Fba1-H111A-3xHA or Fba1-E183A-3xHA as described for panel A. (E) NEM/PEG-maleimide analysis of permeabilized wild-type (BY4742) cells expressing Fba1-3xHA following stripping with EDTA and reloading with zinc. (F) Quantitation of the results in panel (E). The ratios of unmodified + 1× PEG-modified to 2× –5× PEG-modified forms are shown (left panel) is compared with Fba1 activity determined from the same samples (right panel). The error bars indicate ± 1 S.D. ($n = 3$). (G) NEM/PEG analysis of wild-type (BY4743) cells expressing Fba1-3xHA grown in zinc-replete (ZnR) or deficient (ZnD) and following resupply with zinc (100 μ M in LZM) *in vivo* for 1 and 8 h, with and without cycloheximide treatment (100 μ g ml^{-1}). The ratios of unmodified + 1× PEG-modified to 2× –5× PEG-modified forms are shown and the error bars indicate ± 1 S.D. ($n = 3$).

due to slow re-metalation with zinc, we used the NEM/PEG-maleimide method. Cells were grown in zinc-replete and deficient conditions. The deficient cells were then resupplied with zinc for 1 or 8 h in the presence or absence of cycloheximide. *In vivo* NEM treatment was carried out prior to zinc addition and at subsequent timepoints and the cells were then processed for PEG-maleimide modification. In the absence of cycloheximide, full re-metalation was apparent after 1 h of zinc treatment (Figure 9G). Surprisingly, apparent re-metalation was also detected in cycloheximide-treated cells after 1 h and 8 h despite the absence of increased activity. These results suggest that while a substantial fraction of Fba1 protein is not metalated in zinc-deficient cells, it is re-metalated rapidly *in vivo* and some other factor limits its activity. We hypothesize that the protein may be damaged or modified in some way.

Our analysis of Fba1 identified a zinc-responsive conformational change that affects the accessibility of non-ligand cysteines to NEM modification. To test for effects of zinc status specifically on zinc-binding cysteine ligands in a protein, we tested the *in vivo* NEM sensitivity of Met6 methionine synthetase. Met6 accumulates to ~ 145,000 molecules per cell in zinc-replete cells and ~ 87,000 in deficient cells. Fungal methionine synthetases contain three cysteine residues, two of which (C657 and C737) are homologous to the highly conserved residues Cys659 and Cys739 shown to bind zinc in the *C. albicans* Met6 protein[74]. NEM treatment of zinc-replete cells followed by *in vitro* treatment with PEG-maleimide demonstrated protection of 1, 2, or 3 cysteines, with most Met6 proteins having three protected cysteines (Figure 10A). Under zinc-deficient conditions, most proteins had

only one protected cysteine. This result is consistent with zinc binding protecting the two cysteine ligands while the third thiol is constitutively protected *in vivo*. Omitting the *in vitro* DTT reduction step had little effect on the pattern of bands detected indicating that the protective effect of zinc was not conferred by Met6 cysteine oxidation.

This conclusion was supported by the analysis of a mutation that disrupts one of the non-cysteine zinc ligands in the metal site. *S. cerevisiae* His 655 is homologous to H657 of the *C. albicans* and *S. pombe* Met6 proteins, which was identified as zinc binding by structural analysis[75,76]. An H655A mutation resulted in constitutive NEM sensitivity (Figure 10B). In contrast, a mutation in Asp 612 (D612A, homologous to Asp614 in *S. pombe*), which inactivates Met6 by disrupting homocysteine binding had no effect on cysteine sensitivity to NEM in replete conditions, indicating unaltered zinc binding[76]. In deficient conditions, the D612A mutant showed increased NEM protection when compared with the wild-type, suggesting that substrate binding affects zinc lability. Overall however, these observations indicate that the effect of the zinc ligand mutation on NEM sensitivity was not due simply to loss of enzyme activity but reflected a loss of zinc from the active site. Our results are consistent with Met6 accumulating in a zinc-metalated form in replete cells and is only partially metalated in deficient cells.

In this report, we cataloged the zinc proteome of yeast, and determined the abundance of many of its proteins, their response to zinc deficiency, and estimated the metalation state of those proteins in zinc-replete and deficient cells. This analysis has provided unprecedented insights into the cellular economy of zinc under replete and deficient conditions. The

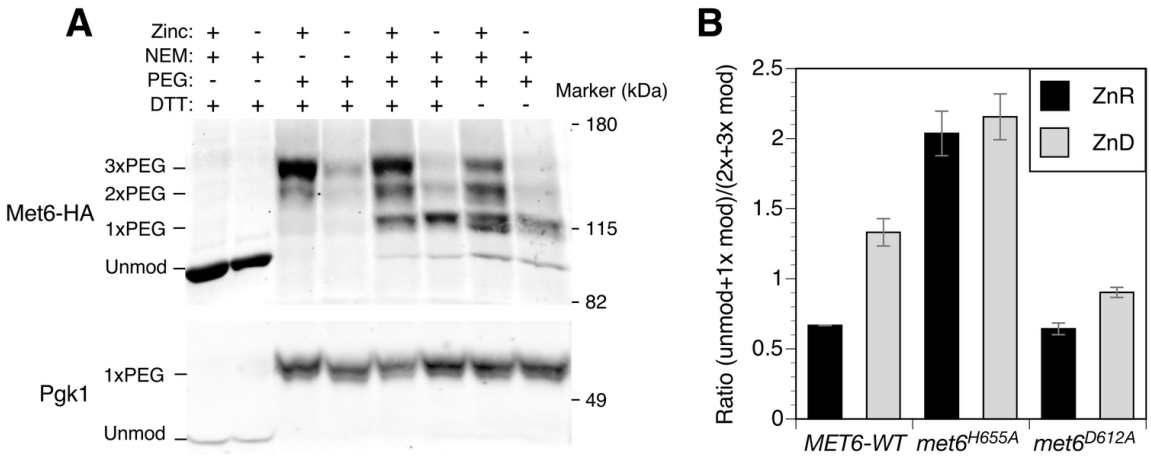


Figure 3.11: *In vivo* analysis of zinc binding by Met6.. (A) NEM/PEG-maleimide analysis of wild-type (BY4742) cells expressing Met6-HA. Zinc-replete (+) or deficient (−) cells were treated with and without NEM, proteins harvested, and then treated with and without DTT prior to PEG-maleimide treatment. The positions of molecular mass markers (kDa) and unmodified and PEG-modified forms of Met6-HA are shown. Pgk1 was detected as a control for PEG labeling efficiency. (B) Quantification of NEM/PEG-maleimide analysis of wild-type (BY4742) cells expressing Met6-HA, Met6^{H655A}-HA or Met6^{D612A}-HA as described for panel A. The ratios of unmodified + 1× PEG-modified to 2–3 × PEG-modified forms are shown and the error bars indicate ± 1 S.D. ($n = 3$)

majority of zinc proteins in yeast bind zinc as a structural cofactor and includes proteins with C_2H_2 zinc fingers, Zn_2C_6 zinc fingers, zinc ribbon domains, and other motifs. These proteins play key roles in gene regulation, transcription, translation, protein degradation, and many other functions. The zinc proteome also plays many catalytic roles and includes enzymes in all six general classes of enzymes. The catalog of the yeast zinc proteome we generated illustrates the diversity of function of this essential cofactor within a eukaryotic cell. We recognize, however, that this is likely not a perfect list of all zinc proteins in yeast. Zinc may play regulatory roles as a second messenger to control different aspects of cell physiology[10,77]. Given that such regulatory sites are probably of lower affinity and greater lability, they may not have been identified in our analysis of stable motifs and domains. Second, some enzymes (*i.e.* “cambialistic" proteins) are functional with different metal cofactors and the specific metals that metalate those enzymes likely depend on the cellular conditions of metal homeostasis and other factors[78]. Thus, to refer to such proteins as “zinc proteins" may not accurately reflect their cofactor requirements. Despite these caveats, however, the catalog of predicted zinc-binding proteins in yeast provides an essential tool for future studies of zinc function, trafficking, and homeostasis.

Using mass spectrometry we measured the abundance of almost half of the entire yeast proteome. Focusing on the zinc proteome, this analysis indicated the existence of 9×10^6 total zinc sites in replete cells. This proteomics-based estimate is remarkably similar to our experimentally determined measurement of the vacuolar and non-vacuolar pools of zinc under these same growth conditions[3]. Cells grown in LZM supplemented with

100 μ M $ZnCl_2$ contain 2.3×10^7 zinc atoms per cell. When the Zrc1 and Cot1 vacuolar zinc transporters were mutated so that zinc could no longer be stored in the vacuole, the accumulation of zinc in the cell dropped to $\sim 1 \times 10^7$ zinc atoms per cell. This result indicated that ~ 60% of the total zinc in a replete cell grown under these conditions is stored in the vacuole. Because labile zinc levels in cells are known to be very low, it is likely that the non-vacuolar zinc atoms are bound by proteins and this conclusion is very consistent with the number of zinc sites we found in these cells using proteomics. Our previous studies also established that the minimum zinc quota of a yeast cell was $\sim 1 \times 10^7$ zinc atoms per cell. This comparison suggests that the zinc proteome abundance of a replete cell is tuned for optimal growth and even slight deficits that decrease full metalation inhibits cell growth. One caveat to this analysis is that we were only able to measure the abundance of about half of the proteins that comprise the zinc proteome. The proteins missing from our analysis are likely to be of very low abundance and therefore may contribute little to our calculations of the cellular zinc requirement. A result very different from ours was obtained from a similar study of the Gram-negative bacterium *Cupriavidus metallidurans* where it was found that replete cells contain a large excess of zinc-binding sites relative to the minimum zinc quota[79]. A high number of zinc sites may add to zinc resistance of this bacterium by serving as a sink for excess metal. Organisms such as *S. cerevisiae* are not adapted to high zinc levels and we suspect it is more indicative of what occurs in most other organisms. A viable alternative hypothesis is that *C. metallidurans* has an excess of zinc sites to serve as a storage pool of the metal in the absence of intracellular organelles.

The total proteome and the zinc proteome of replete cells changes markedly during the transition to deficient conditions. Among the zinc proteins that increased in abundance were several transporter proteins induced by Zap1 to maintain zinc homeostasis. We also observed increased levels of several proteins involved in transcription and chromatin modification. These changes may represent compensatory regulatory responses to loss of these activities due to reduced metalation. The increased abundance of these zinc proteins was more than offset by the decreased abundance of many other zinc proteins such that the net effect was a decrease in the total zinc requirement of the cell by ~ 45% . We had anticipated that proteasome-mediated degradation of apoproteins may be a common effect during the transition to zinc deficiency and there are some potential cases of this in our results. For example, Map1 aminopeptidase levels drop dramatically in zinc-deficient cells with no detectable change in mRNA abundance. In addition, macroautophagy is induced during zinc deficiency and this process may decrease protein abundance more generally[68–70]. However, it is clear that changes in mRNA levels play at least some role in the down-regulation of numerous zinc proteins. While these regulatory effects may occur for many reasons, we hypothesize that a common purpose is to reduce the zinc requirement of the cell, *i.e.* zinc sparing.

One specific group of proteins that had a great impact on the cellular zinc requirement are the alcohol dehydrogenases. There are 13 different zinc-dependent alcohol dehydrogenases in yeast and we could measure the abundance of seven, Adh1, Adh3, Adh4, Adh5, Adh6, Bdh1, and Sfa1. All but Adh4 decrease in zinc-deficient cells for a net decrease in the

zinc requirement of $\sim 9 \times 10^5$ zinc atoms per cell. As mentioned above, Adh4 is induced and Adh1 and Adh3 are repressed by Zap1[25]. How are the other four Adh enzymes shut off in response to zinc deficiency? A clue for how Adh6 is regulated comes from our recent analysis of transcription start sites in the yeast genome. Loss of *ADH6* gene expression coincides with the induction of an antisense transcript that initiates downstream of the *ADH6* open reading frame[64]. While we have not mapped the 3' end of this antisense transcript, it may pass through the open reading frame and *ADH6* promoter and thereby interfere with *ADH6* transcription. This antisense transcript is likely to be Zap1-dependent because there is a consensus Zap1 binding site, 5'-ACCTTAAAGGT-3', located ~ 130 bp upstream of where the antisense RNA initiates transcription. Notably, a similar mechanism of antisense regulation was discovered controlling the *ADH1*gene of *S. pombe* in response to zinc[63]. While *ADH5* and *BDH1* also show decreased mRNA abundance, we have no clues as yet of how that occurs. While we have not directly examined *SFA1* mRNA levels, it was not detected as zinc regulated in previous transcriptome studies[20,57]. These observations suggest that Sfa1 protein levels decrease due to a post-transcriptional mechanism.

The broad effect of zinc deficiency on the abundance of proteins that comprise the translational machinery of the cell is also striking. Several zinc-dependent ribosomal subunits are decreased in expression. In addition, zinc-binding proteins of the ribosome biogenesis (RiBi) regulon, including ribosome biogenesis factors, translation initiation factors, and tRNA synthetases and modifying enzymes, also decrease[80]. Summing all of these proteins, the estimated number of zinc-binding sites drops in deficiency by $\sim 2.2 \times$

10^6 zinc atoms per cell. This collectively represents about half of the zinc sparing that we observed across the entire zinc proteome. This effect extends far beyond the zinc-dependent proteins and includes a large number of zinc-independent translation machinery proteins as well (Figure 3). It has been previously shown that expression of ribosomal proteins and the RiBi regulon is controlled in response to growth rate such that slower growing cells have reduced levels of these proteins than do faster growing cells[80,81]. Given that zinc deficiency results in slower growth, the underlying mechanisms linking growth rate to gene expression likely plays some role in their decreased abundance we observed in our studies.

An additional mechanism regulating the translational machinery in response to zinc deficiency is suggested by the studies from Chanfreau and colleagues who showed that RNAPI is specifically targeted for degradation during zinc deficiency[28]. Decreased RNAPI levels decrease 5.8S, 18S, and 25S rRNA synthesis, which would then trigger decreased ribosomal protein and RiBi regulon expression. It was previously suggested that targeted degradation of RNAPI was a zinc sparing response because of the zinc no longer required for this specific RNA polymerase complex. Based on our results, we estimate that the number of zinc atoms that are directly spared by degrading the five zinc-binding RNAPI subunits would be ~ 30,000 total. While not an insignificant amount, it represents only about 1% of the total zinc sparing that we observed. Mutant cells defective for RNAPI degradation are hypersensitive to EDTA, suggesting a greater disruption in zinc homeostasis than is expected from the level of zinc sparing provided by degrading this protein complex alone.

Therefore, we suggest that the effect of RNAPI degradation on zinc sparing may be much more extensive because of the resulting effects on the expression of ribosomal subunits and the RiBi regulon. In other words, zinc-responsive RNAPI degradation may be the master switch that controls the bulk of the zinc sparing response in zinc-deficient cells.

Despite the massive decrease in the number of zinc proteins that occurs during deficiency, we found that the number of zinc atoms that cells accumulated under these conditions was not sufficient to metalate even the reduced zinc requirement. In fact, we estimated that 70% of the zinc sites on proteins in a zinc-deficient cell were either not metalated (*i.e.* apoproteins) or mismetalated with a different cation. There is some question about how many potential zinc-binding ribosomal subunits in eukaryotes actually bind zinc *in vivo*[82].Nonetheless, if those proteins are removed from our calculations, the level of zinc atoms per cell are still insufficient for full metalation. Our previous studies of Tsa1 suggested that this protein's chaperone function was critical to tolerate this stress and that zinc apoproteins are the likely clients of Tsa1[23]. We tested the hypothesis of accumulated apoproteins for two abundant zinc proteins, Fba1 and Met6. Despite little change in protein levels, Fba1 enzyme activity was greatly reduced in zinc-deficient cells compared to replete cells. In addition, using a novel application of thiol-reactive agents to study *in vivo* metalation, we found that Fba1 in cells underwent a conformational change that was dependent on zinc availability and zinc ligands. These results suggested that Fba1 is not fully metalated in zinc-deficient cells. It was therefore surprising that reintroduction of zinc appeared to rapidly metalate the protein but did not restore enzyme activity to replete levels and new

protein synthesis was required to restore full activity. It is conceivable that the Fba1 protein is somehow damaged or modified (*e.g.* by phosphorylation) in zinc-deficient cells such that new Fba1 protein is required to restore activity. One possible mechanism of damage is glycation which has been previously observed for yeast Fba1[83]. Our studies of Met6 also indicated the accumulation of apoproteins during zinc deficiency. For Met6, we could probe reactivity of the cysteine ligands involved in direct zinc binding. These studies suggested that the many Met6 molecules in a cell are also unmetalated. It is conceivable that Fba1 and/or Met6 are mismetalated but if so, binding of that other cation fails to produce the same conformational change as zinc does for Fba1 and fails to protect the bound ligand from reaction with NEM in Met6. It was also surprising that many chaperones, co-chaperones, and chaperonin proteins (*e.g.* Ssa2, Ydj1, Scj1, Cct2) decreased in abundance during zinc deficiency (Figure 3, cluster 2) when our other results indicated an increase in apoprotein accumulation and unfolded protein stress. We note that many of the proteins that decreased in abundance during deficiency are “foldase" -type chaperones while Tsa1 is a “holdase" -type protein. These results make biological sense because apoproteins cannot be fully folded without resupply of their metal cofactor and foldase chaperones would not suffice.

Conclusion

In this report, we have cataloged the zinc proteome of yeast and determined the effects of zinc deficiency on the abundance of these proteins. This analysis highlighted the diverse functions of zinc proteins and mapped in detail the subcellular requirement of this essential

metal nutrient. Our studies demonstrated that the majority of zinc proteins detected decrease in abundance during deficiency and this apparent zinc-sparing response is due at least in part to transcriptional regulation of many of their respective genes. Our results also indicated that a surprisingly high number of zinc sites are not metalated with zinc during deficiency. Future studies will address the role of Tsa1 as a chaperone of these abundant zinc apoproteins. Future studies will also address the prevalence of zinc apoproteins more broadly and determine whether all proteins are similarly affected by zinc deficiency or whether some proteins can compete for zinc better than others. It will be exciting to learn whether prioritization of zinc distribution occurs in zinc trafficking among apoproteins.

Acknowledgements This work was supported by National Institutes of Health grant R01-GM56285 (DJE) and grants P41-GM10853 and R35GM118110 (JJC). The authors also acknowledge the support and the use of resources of Instruct-ERIC, a Landmark ESFRI project, and specifically the CERM/CIRMMP Italy Centre.

References

[1] C. Andreini, L. Banci, I. Bertini, and A. Rosato, "Zinc through the three domains of life," 2006.

[2] C. E. Outten and T. V. O'Halloran, "Femtomolar sensitivity of metalloregulatory proteins controlling zinc homeostasis," *Science*, vol. 292, pp. 2488–2492, jun 2001.

[3] C. W. MacDiarmid, "Zinc transporters that regulate vacuolar zinc storage in Saccharomyces cerevisiae," 2000.

[4] R. D. Palmiter and S. D. Findley, "Cloning and functional characterization of a mammalian zinc transporter that confers resistance to zinc.," 1995.

[5] D. A. Suhy, K. D. Simon, D. I. Linzer, and T. V. O'Halloran, "Metallothionein is part of a zinc-scavenging mechanism for cell survival under conditions of extreme zinc deprivation," 1999.

[6] P. J. P. J. Dittmer, J. G. Miranda, J. A. Gorski, and A. E. Palmer, "Genetically encoded sensors to elucidate spatial distribution of cellular zinc," jun 2009.

[7] C. A. Fierke and R. B. Thompson, "Fluorescence-based biosensing of zinc using carbonic anhydrase," 2001.

[8] J. L. Vinkenborg, M. S. Koay, and M. Merkx, "Fluorescent imaging of transition metal homeostasis using genetically encoded sensors," 2010.

[9] P. Chandrangsu, C. Rensing, and J. D. Helmann, "Metal homeostasis and resistance in bacteria," 2017.

[10] T. Hara, T. aki Takeda, T. Takagishi, K. Fukue, T. Kambe, and T. Fukada, "Physiological roles of zinc transporters: molecular and genetic importance in zinc homeostasis," 2017.

[11] A. Krężel and W. Maret, "The functions of metamorphic metallothioneins in zinc and copper metabolism," *International Journal of Molecular Sciences*, vol. 18, p. 1237, jun 2017.

[12] D. J. Eide, "Homeostatic and adaptive responses to zinc deficiency in," 2009.

[13] S. S. Merchant and J. D. Helmann, "Elemental Economy. Microbial Strategies for Optimizing Growth in the Face of Nutrient Limitation," 2012.

[14] E. M. Panina, A. A. Mironov, and M. S. Gelfand, "Comparative genomics of bacterial zinc regulons: Enhanced ion transport, pathogenesis, and rearrangement of ribosomal proteins," 2003.

[15] N. Ogasawara, S.-M. Park, F. Kawamura, K. Kobayashi, G. Akanuma, T. Kudo, R. Murayama, H. Nanamiya, K. Ochi, Y. Natori, and S. Kosono, "Zinc is a key factor in controlling alternation of two types of L31 protein in the Bacillus subtilis ribosome," 2004.

[16] S. E. Gabriel and J. D. Helmann, "Contributions of Zur-controlled ribosomal proteins to growth under zinc starvation conditions," 2009.

[17] S. Choi and A. J. Bird, "Zinc'ing sensibly: Controlling zinc homeostasis at the transcriptional level," 2014.

[18] M. C. Kersting and G. M. Carman, "Regulation of the Saccharomyces cerevisiae EKI1-encoded ethanolamine kinase by zinc depletion," 2006.

[19] A. Soto and G. M. Carman, "Regulation of the Saccharomyces cerevisiae CKI1-encoded choline kinase by zinc depletion," 2008.

[20] C.-Y. Y. Wu, A. J. Bird, L. M. Chung, M. A. Newton, D. R. Winge, D. J. Eide, M. A. Newton, C.-Y. Y. Wu, A. J. Bird, and L. M. Chung, "Differential control of Zap1-regulated genes in response to zinc deficiency in Saccharomyces cerevisiae," *BMC Genomics*, vol. 9, no. 1, p. 370, 2008.

[21] C. Y. Wu, A. J. Bird, D. R. Winge, and D. J. Eide, "Regulation of the yeast TSA1 peroxiredoxin by ZAP1 is an adaptive response to the oxidative stress of zinc deficiency," 2007.

[22] H. H. Jang, K. O. Lee, Y. H. Chi, B. G. Jung, S. K. Park, and J. H. Park ..., "Two Enzymes in One Two Yeast Peroxiredoxins Display Oxidative Stress-Dependent Switching from a ...," 2004.

[23] D. J. Eide, C. W. MacDiarmid, K. Kerdsomboon, K. Schelble, J. Taggart, M. Kubisiak, S. Panascharoen, K. Kerdsomboon, M. Kubisiak, S. Panascharoen, K. Schelble, and D. J. Eide, "Peroxiredoxin Chaperone Activity Is Critical for Protein Homeostasis in Zinc-deficient Yeast," oct 2013.

[24] J. Frydman, D. Kaganovich, and R. Kopito, "Misfolded proteins partition between two distinct quality control compartments," 2008.

[25] A. J. Bird, M. Gordon, D. J. Eide, and D. R. Winge, "Repression of ADH1 and ADH3 during zinc deficiency by Zap1-induced intergenic RNA transcripts," 2006.

[26] V. M. Williamson and C. E. Paquin, "Homology of Saccharomyces cerevisiae ADH4 to an iron-activated alcohol dehydrogenase from Zymomonas mobilis," *MGG Molecular & General Genetics*, vol. 209, pp. 374–381, sep 1987.

[27] U. Diisseldorf, "Overexpression, purification and properties of alcohol dehydrogenase IV from Saccharomyces cerevisiae Christel Drewke and Michael Ciriacy," 1988.

[28] Y. J. Lee, C. Y. Lee, A. Grzechnik, F. Gonzales-Zubiate, A. A. Vashisht, A. Lee, J. Wohlschlegel, and G. F. Chanfreau, "RNA polymerase I stability couples cellular growth to metal availability," 2013.

[29] W. Qiao, C. Ellis, J. Steffen, C. Y. Wu, and D. J. Eide, "Zinc status and vacuolar zinc transporters control alkaline phosphatase accumulation and activity in Saccharomyces cerevisiae," 2009.

[30] H. Zhao and D. Eide, "The yeast ZRT1 gene encodes the zinc transporter protein of a high-affinity uptake system induced by zinc limitation.," 2002.

[31] S. Braun, J. S. Weissman, N. J. Krogan, M. Schuldiner, D. M. Cameron, S. R. Collins, H. W. Newman, H. D. Madhani, D. K. Breslow, and J. Stewart-Ornstein, "A comprehensive strategy enabling high-resolution functional analysis of the yeast genome," 2008.

[32] R. D. Finn, A. Bateman, J. Clements, P. Coggill, R. Y. Eberhardt, S. R. Eddy, A. Heger, K. Hetherington, L. Holm, J. Mistry, E. L. L. Sonnhammer, J. Tate, and M. Punta, "Pfam: the protein families database," *Nucleic Acids Res*, vol. 42, pp. D222—-D230, jan 2014.

[33] S. R. Eddy, "Profile hidden Markov models.," *Bioinformatics (Oxford, England)*, vol. 14, no. 9, pp. 755–63, 1998.

[34] Y. Valasatava, A. Rosato, L. Banci, and C. Andreini, "MetalPredator: A web server to predict iron-sulfur cluster binding proteomes," 2016.

[35] A. Rosato, Y. Valasatava, and C. Andreini, "Minimal functional sites in metalloproteins and their usage in structural bioinformatics," *International Journal of Molecular Sciences*, vol. 17, may 2016.

[36] A. Bateman, M. J. Martin, C. O'Donovan, M. Magrane, E. Alpi, R. Antunes, B. Bely, M. Bingley, C. Bonilla, R. Britto, B. Bursteinas, H. Bye-AJee, A. Cowley, A. Da Silva, M. De Giorgi, T. Dogan, F. Fazzini, L. G. Castro, L. Figueira, P. Garmiri, G. Georghiou, D. Gonzalez, E. Hatton-Ellis, W. Li, W. Liu, R. Lopez, J. Luo, Y. Lussi, A. MacDougall, A. Nightingale, B. Palka, K. Pichler, D. Poggioli, S. Pundir, L. Pureza, G. Qi, S. Rosanoff, R. Saidi, T. Sawford, A. Shypitsyna, E. Speretta, E. Turner, N. Tyagi, V. Volynkin, T. Wardell, K. Warner, X. Watkins, R. Zaru, H. Zellner, I. Xenarios, L. Bougueleret, A. Bridge, S. Poux, N. Redaschi, L. Aimo, G. ArgoudPuy, A. Auchincloss, K. Axelsen, P. Bansal, D. Baratin, M. C. Blatter, B. Boeckmann, J. Bolleman, E. Boutet, L. Breuza, C. Casal-Casas, E. De Castro, E. Coudert, B. Cuche, M. Doche, D. Dornevil, S. Duvaud, A. Estreicher, L. Famiglietti, M. Feuermann, E. Gasteiger, S. Gehant, V. Gerritsen, A. Gos, N. Gruaz-Gumowski, U. Hinz, C. Hulo, F. Jungo, G. Keller, V. Lara, P. Lemercier, D. Lieberherr, T. Lombardot, X. Martin, P. Masson, A. Morgat, T. Neto, N. Nouspikel, S. Paesano, I. Pedruzzi, S. Pilbout, M. Pozzato, M. Pruess, C. Rivoire, B. Roechert, M. Schneider, C. Sigrist, K. Sonesson, S. Staehli, A. Stutz, S. Sundaram, M. Tognolli, L. Verbregue, A. L. Veuthey, C. H. Wu, C. N. Arighi, L. Arminski, C. Chen, Y. Chen, J. S. Garavelli, H. Huang, K. Laiho, P. McGarvey, D. A. Natale, K. Ross, C. R. Vinayaka, Q. Wang, Y. Wang, L. S. Yeh, and J. Zhang, "UniProt: The universal protein knowledgebase," *Nucleic Acids Research*, vol. 45, pp. D158–D169, mar 2017.

[37] C. Andreini, I. Bertini, and G. Cavallaro, "Minimal functional sites allow a classification of zinc sites in proteins," 2011.

[38] J. R. Wiśniewski, M. Y. Hein, J. Cox, and M. Mann, "A "Proteomic Ruler" for Protein Copy Number and Concentration Estimation without Spike-in Standards," 2014.

[39] S. Tyanova, Temu Tikira, Sinitcyn Pavel, Carlson Arthur, Hein Marco, Geiger Tamar, Mann Matthias, and Cox Jürgen, "The Perseus computational platform for comprehensive analysis of (prote)omics data," 2016.

[40] G. Dennis, B. T. Sherman, D. A. Hosack, J. Yang, W. Gao, H. C. Lane, R. A. Lempicki, M. J. Booth, W. Reik, S. Balasubramanian, L. Abramowitz, M. Bartolomei, F. Rambow, M. Bassi, T. Bruno, M. Fanciulli, C. Renner, A. Klein-Szanto, Y. Matsumoto, D. Kobi, I. Davidson, C. Alberti, L. Larue, and A. Bellacosa, "Genome-wide distribution of 5-formylcytosine in embryonic stem cells is associated with transcription and depends on thymine DNA glycosylase," *Genome Biology*, vol. 4, p. P3, apr 2003.

[41] C. W. Macdiarmid, J. Taggart, J. Jeong, K. Kerdsomboon, and D. J. Eide, "Activation of the yeast UBI4 polyubiquitin gene by Zap1 transcription factor via an intragenic promoter is critical for zinc-deficient growth," *Journal of Biological Chemistry*, vol. 291, pp. 18880–18896, sep 2016.

[42] C. W. MacDiarmid, M. A. Milanick, and D. J. Eide, "Induction of the ZRC1 metal tolerance gene in zinc-limited yeast confers resistance to zinc shock," *Journal of Biological Chemistry*, vol. 278, pp. 15065–15072, apr 2003.

[43] M. Cieśla, J. Mierzejewska, M. Adamczyk, A.-K. Ö. K. Ö. Farrants, and M. Boguta, "Fructose bisphosphate aldolase is involved in the control of RNA polymerase III-directed transcription," jun 2014.

[44] A. P. Freire, A. M. Martins, and C. Cordeiro, "An experiment illustrating metabolic regulation in situ using digitonin permeabilized yeast cells," *Biochemical Education*, vol. 26, pp. 161–163, apr 1998.

[45] P. Zipfel, I. D. Jacobsen, L. Schild, S. Brunke, M. Brock, P. Miramón, F. Citiulo, B. Hube, and D. Wilson, "Candida albicans Scavenges Host Zinc via Pra1 during Endothelial Invasion," 2012.

[46] A. Pagani, L. Villarreal, M. Capdevila, and S. Atrian, "The Saccharomyces cerevisiae Crs5 Metallothionein metal-binding abilities and its role in the response to zinc overload," 2007.

[47] Ò. Palacios, S. Atrian, and M. Capdevila, "Zn-and Cu-thioneins: A functional classification for metallothioneins?," *Journal of Biological Inorganic Chemistry*, vol. 16, pp. 991–1009, oct 2011.

[48] C. Andreini and I. Bertini, "A bioinformatics view of zinc enzymes," 2012.

[49] Huh, W. K., Falvo, J. V., Gerke, L. C., Carroll, A. S., Howson, R. W., Weissman, J. S., O'Shea, and E. K., "Global analysis of protein localization in budding yeast.," 2003.

[50] J. L. Y. Koh, C. Boone, H. Friesen, A. Moses, J. Moffat, Y. T. Chong, and B. J. Andrews, "CYCLoPs: A Comprehensive Database Constructed from Automated Analysis of Protein Abundance and Subcellular Localization Patterns in Saccharomyces cerevisiae," *G3: Genes|Genomes|Genetics*, vol. 5, pp. 1223–1232, jun 2015.

[51] J. J. Coon, D. J. Bailey, M. S. Westphall, A. S. Hebert, E. E. Coughlin, A. Ulbrich, A. L. Richards, A. S. Hebert, A. Ulbrich, D. J. Bailey, E. E. Coughlin, M. S. Westphall, and J. J. Coon, "One-hour proteome analysis in yeast," *Nature Protocols*, vol. 10, pp. 701–714, may 2015.

[52] J. A. Stefely, N. W. Kwiecien, E. C. Freiberger, A. L. Richards, A. Jochem, M. J. Rush, A. Ulbrich, K. P. Robinson, P. D. Hutchins, M. T. Veling, X. Guo, Z. A. Kemmerer, K. J. Connors, E. A. Trujillo, J. Sokol, H. Marx, M. S. Westphall, A. S. Hebert, D. J. Pagliarini, and J. J. Coon, "Mitochondrial protein functions elucidated by multi-omic mass spectrometry profiling," *Nature Biotechnology*, vol. 34, pp. 1191–1197, nov 2016.

[53] J. R. Wiśniewski, A. Zougman, N. Nagaraj, and M. Mann, "Universal sample preparation method for proteome analysis," *Nature Methods*, vol. 6, pp. 359–362, may 2009.

[54] S. Ghaemmaghami, W.-K. K. Huh, K. Bower, R. W. Howson, A. Belle, N. Dephoure, E. K. O'Shea, J. S. Weissman, S. Ghaemmaghami, K. Bower, W.-K. K. Huh, N. Dephoure, E. K. O'Shea, and R. W. Howson, "Global analysis of protein expression in yeast," *Nature*, vol. 425, pp. 737–741, oct 2003.

[55] R. Milo, "What is the total number of protein molecules per cell volume? A call to rethink some published values," 2013.

[56] J. J. Strain, D. R. Winge, L. T. Jensen, W. R. Howard, and V. C. Culotta, " Enhanced Effectiveness of Copper Ion Buffering by CUP1 Metallothionein Compared with CRS5 Metallothionein in Saccharomyces cerevisiae ," 2002.

[57] T. J. Lyons, A. P. Gasch, L. A. Gaither, D. Botstein, P. O. Brown, and D. J. Eide, "Genome-wide characterization of the Zap1p zinc-responsive regulon in yeast.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, pp. 7957–62, jul 2000.

[58] S. H. Han, G. S. Han, W. M. Iwanyshyn, and G. M. Carman, "Regulation of the PIS1-encoded phosphatidylinositol synthase in Saccharomyces cerevisiae by zinc," 2005.

[59] M. C. Walsh, J. T. Pronk, J.-M. Daran, R. De Nicola, E. A. F. De Hulster, T. A. Knijnenburg, M. J. T. Reinders, G. M. Walker, L. A. Hazelwood, and P. Daran-Lapujade, "Physiological and Transcriptional Responses of Saccharomyces cerevisiae to Zinc Limitation in Chemostat Cultures," 2007.

[60] C.-Y. Y. Wu, D. R. Winge, D. J. Eide, S. Roje, F. J. Sandoval, A. J. Bird, D. R. Winge, and D. J. Eide, "Repression of Sulfate Assimilation Is an Adaptive Response of Yeast to the Oxidative Stress of Zinc Deficiency," *Journal of Biological Chemistry*, vol. 284, pp. 27544–27556, oct 2009.

[61] N. Singh, K. K. Yadav, and R. Rajasekharan, "ZAP1-mediated modulation of triacylglycerol levels in yeast by transcriptional control of mitochondrial fatty acid biosynthesis," 2016.

[62] A. Soto-Cardalda, S. Fakas, F. Pascual, H. S. Choi, and G. M. Carman, "Phosphatidate phosphatase plays role in zinc-mediated regulation of phospholipid synthesis in yeast," 2012.

[63] K. M. Ehrensberger, C. Mason, M. E. Corkins, C. Anderson, N. Dutrow, B. R. Cairns, B. Dalley, B. Milash, and A. J. Bird, "Zinc-dependent regulation of the adh1 antisense transcript in fission yeast," 2013.

[64] Y.-H. H. Wu, J. Taggart, P. X. Song, C. MacDiarmid, and D. J. Eide, "An MSC2 promoter-lacZ fusion gene reveals zinc-responsive changes in sites of transcription initiation that occur across the yeast genome," *PLoS ONE*, vol. 11, p. e0163256, sep 2016.

[65] "Retrotransposon expression in ethanol-stressed {Saccharomyces} cerevisiae. - {PubMed} - {NCBI}."

[66] A.-L. Todeschini, A. Morillon, M. Springer, and P. Lesage, "Severe Adenine Starvation Activates Ty1 Transcription and Retrotransposition in Saccharomyces cerevisiae," 2005.

[67] V. A. Bradshaw and K. McEntee, "DNA damage activates transcription and transposition of yeast Ty retrotransposons," 1989.

[68] T. Kawamata, T. Horie, M. Matsunami, M. Sasaki, and Y. Ohsumi, "Zinc starvation induces autophagy in yeast," *J Biol Chem*, vol. 292, pp. 8520–8530, may 2017.

[69] M. D. Bucci, E. Weisenhorn, S. Haws, Z. Yao, G. Zimmerman, M. Gannon, J. Taggart, T. Lee, D. J. Klionsky, J. Russell, J. Coon, and D. J. Eide, "An autophagy-independent role for ATG41 in sulfur metabolism during zinc deficiency," *Genetics*, vol. 208, pp. 1115–1130, mar 2018.

[70] S. Yamashita, K. Hosaka, J. Nikawa, and T. Kodaki, "Regulation of phospholipid synthesis in yeast," 2001.

[71] J. L. Apuy, L. S. Busenlehner, D. H. Russell, and D. P. Giedroc, "Ratiometric Pulsed Alkylation Mass Spectrometry as a Probe of Thiolate Reactivity in Different Metalloderivatives of Staphylococcus aureus pI258 CadC," 2004.

[72] J. L. Apuy, D. H. Russell, D. P. Giedroc, X. Chen, and T. O. Baldwin, " Ratiometric Pulsed Alkylation/Mass Spectrometry of the Cysteine Pairs in Individual Zinc Fingers of MRE-Binding Transcription Factor-1 (MTF-1) as a Probe of Zinc Chelate Stability † ," 2002.

[73] V. L. Mendoza and R. W. Vachet, "Probing protein structure by amino acid-specific covalent labeling and mass spectrometry," 2009.

[74] D. Ubhi, G. Kago, A. F. Monzingo, and J. D. Robertus, "Structural analysis of a fungal methionine synthase with substrates and inhibitors," 2014.

[75] D. Ubhi, K. L. Kavanagh, A. F. Monzingo, and J. D. Robertus, "Structure of Candida albicans methionine synthase determined by employing surface residue mutagenesis," 2011.

[76] P. Prasannan, H. S. Suliman, and J. D. Robertus, "Kinetic analysis of site-directed mutants of methionine synthase from Candida albicans," 2009.

[77] S. Chen, R. Bleher, S. A. Garwin, T. V. O'Halloran, T. K. Woodruff, A. R. Bayer, B. Y. Kong, E. L. Que, F. E. Duncan, S. C. Gleber, V. P. Dravid, and S. Vogt, "Quantitative mapping of zinc fluxes in the mammalian egg reveals the origin of fertilization-induced zinc sparks," 2014.

[78] J. P. Barnett, A. Millard, A. Z. Ksibe, D. J. Scanlan, R. Schmid, and C. A. Blindauer, "Mining genomes of marine cyanobacteria for elements of zinc homeostasis," 2012.

[79] M. Herzberg, D. Dobritzsch, S. Helm, S. Baginsky, and D. H. Nies, "The zinc repository of Cupriavidus metallidurans," *Metallomics*, vol. 6, pp. 2157–2165, sep 2014.

[80] O. Rodríguez-Galán, V. Begley, F. Gómez-Herreros, S. Chávez, J. de la Cruz, and M. de la Cruz Muñoz-Centeno, "Feedback regulation of ribosome assembly," *Current Genetics*, vol. 64, pp. 393–404, apr 2017.

[81] D. Shore, H. Lempia, H. Lempiäinen, and D. Shore, "Growth control and ribosome biogenesis," *Current Opinion in Cell Biology*, vol. 21, pp. 855–863, dec 2009.

[82] N. Brandes, D. Reichmann, H. Tienson, L. I. Leichert, and U. Jakob, "Using quantitative redox proteomics to dissect the yeast redoxome," 2011.

[83] G. Graça, R. A. Gomes, H. V. Miranda, C. Cordeiro, M. S. Silva, A. V. Coelho, A. P. Freire, and A. E. Ferreira, " Yeast protein glycation in vivo by methylglyoxal ," 2006.

Chapter 4

DI-GLYCINE REMNANT ENRICHMENT REVEALS PUB1-MEDIATED OLEOSIN UBIQUITINATION AND LIPID MOBILIZATION IN THE LEGUME *MEDICAGO TRUNCATULA.*

This chapter is part of the manuscript in preparation:

Weisenhorn, E.M., Chakraborty, S., Maeda, J., Jayarman, D., Kleven, B., Sussman, M., Ané, J., Coon JJ. *Di-glycine remnant enrichment reveals PUB1-mediated oleosin ubiquitination and lipid mobilization in the legume Medicago truncatula.* **2019**

Abstract

Ubiquitination is a crucial post-translational protein modification that is involved in protein degradation and trafficking. Quantification of ubiquitination is technically challenging due to its low abundance and transient nature in the cell. Legumes are agriculturally important plants owing to their ability to form symbioses with rhizobial bacteria and mycorrhizal fungi. Here, we provide the first reported quantification of ubiquitination in the model legume *Medicago truncatula.* Plant U-box type E3 ubiquitin ligases (PUBs) are essential components of ubiquitination. The *pub1* mutant of *M. truncatula* has symbiotic defects, although the biological role of PUB1 remains mostly unexplored. We optimized a protocol to compare protein profiles between wild-type and *pub1* mutant seedlings. First, we optimized the enrichment for di-glycine peptides, which result from the tryptic digest of ubiquitinated proteins. We then applied this method to compare ubiquitination between *M. truncatula* wild-type and *pub1* seedlings. One hundred and sixty-nine ubiquitination sites on 112 unique proteins were quantified using tandem mass tags. In conjunction with ubiquitination sites, we quantified the proteome of each sample and identified 5,597 proteins. We detected ubiquitination in several oleosin proteins that bind lipid droplets and discovered that *PUB1* regulates their ubiquitination. We further observed that *PUB1* governs the accumulation of triglycerides, one of the chief constituents of lipid droplets, that associate with oleosins. Our findings shed light on the biological role of PUB1 in *M. truncatula* and allude to the possible involvement of this protein in lipid droplet mobilization.

Introduction

Ubiquitin (Ub) is a small protein (8.5 kDa) that can become bound to other proteins on their lysine residues to mark them for degradation by the 26S ubiquitin-proteasome system (UPS), alter cellular localization, or regulate protein-protein interactions[1]. Ubiquitination in the model plant *Arabidopsis thaliana* is well documented, where approximately 6% of the gene products are associated with ubiquitination and proteasome-related functions[2]. Protein degradation via the 26S proteasome system involves the sequential activities of three enzymes- E1, E2, and E3. E1 activates the Ub molecule and subsequently transfers it from the E1-Ub complex to E2. Finally, the E3 ligase forms an E2-E3-substrate complex, preparing the substrate for degradation, which, in plants is primarily mediated by its conserved E2-interacting, U-box domain[3]. Ubiquitination results in the C-terminus of Ub getting conjugated to the target protein[4]. From the C-terminus, the final three amino acids of Ub are glycine, glycine, and arginine. Hence, when Ub-conjugated proteins are subjected to tryptic digest, a di-glycine remnant attached to the modified peptide is left, while the remainder of the ubiquitin is removed. Recently, antibodies have become available that can selectively enrich for these di-glycine peptides[5].

Plant U-box type E3 ubiquitin ligases (PUBs) participate in a variety of biological processes such as cell proliferation, immunity, hormone signaling and stress response[6–9]. In the model legume *Medicago truncatula* as well as in Arabidopsis, 64 U-box proteins have been predicted, in contrast to eight in humans, thus, highlighting the potential importance

of PUBs in plants[3,10]. In addition to participating in ubiquitination, PUBs are involved in vesicular trafficking by interacting with membrane proteins and phospholipids[3].

Legumes are a group of plants whose roots form beneficial associations with symbiotic bacteria and fungi, developing the symbiotic root organ, nodule, and mycorrhiza, respectively. PUBs and other classes of E3 ubiquitin ligases have been widely studied in legumes, owing to their involvement in rhizobium-legume symbiosis[11] In *M. truncatula,* PUB1 and PUB2 play an essential role in rhizobium-legume symbiosis[12–14]; both are phosphorylated by the receptor kinase Does not Make Infection 2 (DMI2), also known as Nodulation Receptor Kinase (NORK)[13,14]. PUB1 is additionally phosphorylated by the plasma membrane-localized Lysin motif (LysM) receptor-like kinase 3 (LYK3)[12]. NORK and LYK3 are essential components of the initial stages of rhizobium-legume symbiosis.

PUB1 consists of a poorly-conserved U-box N-terminal Domain (UND), a highly conserved U-box domain, which is followed by a variant region of ARMADILLO (ARM) repeats[12]. The protein localizes to the plasma membrane and is composed of 694 amino acid residues[11,12]. In addition to rhizobium-legume symbiosis, PUB1 also controls the arbuscular mycorrhizal symbiosis. The substitution of an aspartate to an asparagine residue in the U-box domain in the *pub1* mutant led to an increase in rhizobial infection as well as fungal colonization[13]. Despite a regulatory role in these symbioses, no substrate of PUB1 has yet been identified.

PUBs regulate a myriad of cellular processes in plants[3]. Other than in the context of symbiosis, the biological role of PUB1 remains largely unexplored. To gain insight into the

function of PUB1, we compared protein profiles between wild-type *M. truncatula* seedlings and *pub1* mutants. We first validated the Ub enrichment reproducibility to ensure that biological comparisons can be reliably made and investigated the optimum quantity of *M. truncatula* protein input to maximize site identifications. Additionally, we leveraged the physical properties of the di-glycine peptides to tune our instrument parameters to maximize identifications. Our work revealed that ubiquitination of several oleosin proteins is selectively downregulated in the *pub1* mutants, resulting in an increased abundance of these proteins in the mutant. Triglycerides associated with oleosin proteins also increased in abundance, supporting this result. To the best of our knowledge, this is the first study on ubiquitination in *M. truncatula.*

Materials and Methods

Plant growth *Medicago truncatula* wild-type (Jemalong A17) and *pub1* mutant seeds were scarified with sulfuric acid for 15 minutes, surface sterilized with 30% Clorox for 10 minutes, imbibed at room temperature for 4-5 hours and stratified at 4 ° C for 1-4 days. Overnight germinated seedlings were plated on modified Fahräeus medium on top of a germination paper. A week later, seedlings were treated with 10^{-8} M purified Nod factors from *Sinorhizobium meliloti* GMI6390 or 0.005% ethanol as a control treatment in the dark for an hour. Whole seedlings were then ground in liquid nitrogen and stored at -80 ° C until further processing.

Seedling preparation and digestion Ground tissue samples were re-suspended in 5 mL 6 M guanidine, 100 mM Tris pH 8. The resulting slurry was probe-sonicated for approximately two minutes until homogenous. Protein was extracted using a 90% methanol precipitation, and the pellet was re-suspended in 3 mL lysis buffer (8 M Urea, 100 mM Tris, 20 mM TCEP, 80 mM Chloroacetamide). Protein digestion was performed overnight with LysC (1:200) before diluting to 1.5 M urea and digesting with trypsin for an additional 3 hours (1:200). Following digestion, the sample was centrifuged for five minutes at 10,000 G. Each sample was subsequently split into three aliquots which were desalted with 100 mg Strata C18 solid phase extraction cartridges and dried in a vacuum centrifuge. The resulting protein concentration was then measured using a Nanodrop One (Thermo Scientific).

Ubiquitin Enrichment and TMT Labeling To enrich for di-glycine peptides, the PTMScan Ubiquitin Remnant Motif Kit from Cell Signaling Technology was employed, and all samples and buffers were kept on ice during the method. From each sample, 10 mg of peptides were re-suspended in immunoaffinity purification (IAP) buffer, and pH was adjusted to 7 with 1 M Tris (not pH adjusted). Peptides were then centrifuged at 20,000 G for 5 minutes to remove insoluble material. Meanwhile, one vial of antibody beads was re-suspended in 1.5 mL IAP buffer. For each sample, 150 µl beads were aliquoted to a 1.5 mL tube. Peptides were then transferred to tubes containing beads and incubated for two hours at 4° C with end over end rotation. Samples were then centrifuged at 2,000 G for one minute following which beads were allowed to settle completely. The supernatant was removed and saved

as the un-ubiquitinated portion of the proteome. The beads were then washed with 1 mL cold IAP buffer, inverted five times, and then centrifuged at 2,000 G for one minute. After removing the supernatant, the beads were washed twice more with 1 mL cold phosphate buffered saline (PBS) as described previously. Following the third wash, 50 µl elution buffer (0.15% trifluoroacetic acid) was added, and the bottom of the tube was tapped several times. The beads were allowed to sit for five minutes at room temperature and then centrifuged at 2,000 G for one minute. The supernatant was collected, and elution was repeated twice more as described. Each sample was then desalted using a 10 mg Strata C18 solid phase extraction cartridges and dried in a vacuum centrifuge.

To label samples, the TMT 10plex isobaric label kit provided by Thermo Fisher was used. Two 8-plex labeling experiments were performed with two of the four replicates from each sample included in each trial. 0.8 mg aliquots of labeling reagent were brought to room temperature and spun down before re-suspending in 50 µl 100% acetonitrile. Di-glycine peptide samples were re-suspended in 100 µl 200 mM triethylammonium bicarbonate (TEAB) and combined with the label before vortexing for 6 hours at room temperature. A test mix was prepared by combining 5 µl of each sample with 50 µl water and dried down. The test mix was re-suspended in 0.2% formic acid (FA) and analyzed on an Orbitrap Fusion Lumos (Thermo Fisher). Remaining samples were then mixed according to ratios and desalted using 10 mg Strata C18 solid phase extraction cartridges.

Samples were fractionated using high-pH reverse-phase separation to increase proteomic depth. The solvent system consisted of mobile phase A (10 mM ammonium formate

pH 10) and mobile phase B (10 mM ammonium formate pH 10 80% methanol) which was run on an Ultimate 3000 UPLC system (Dionex Sunnyvale, CA) with a reverse-phase C18 column. Gradient elution was performed at 800 μ l min^{-1} with the gradient increased from 0 to 6% B over 5 min followed by an increase to 80% B until 24 minutes and a wash at 100% B for 3 minutes. Sixteen fractions were collected from each pooled TMT experiment and subsequently dried and re-suspended in 0.2% FA.

Mass spectrometry and high-performance liquid chromatography Samples were analyzed using an LC-MS instrument comprising an Orbitrap Fusion Lumos Tribrid mass spectrometer and Ultimate 3000 RSLCnano liquid chromatography system (Thermo Fisher Scientific). Mobile phase A consisted of 0.2% formic acid in water and mobile phase B consisted of 0.2% formic acid in 70% acetonitrile. A 75-min gradient ranging from 0% to 70% B was employed spanning a total runtime of 90 min. Analytes were injected onto a 1.7 μ m C18 column (75 μ m i.d.) packed in-house to a length of 35 cm and heated to 45° C. Survey scans of peptide precursors were collected every second from 300-1350 Th with an AGC target of 1x10^6 and a resolution of 60,000 in the Orbitrap. Precursors were isolated from a 1.6 Th window in the quadrupole, and HCD MS/MS scans at 35% collision energy were collected in the orbitrap with an AGC target of 5x10^4 from 100-1200 Th.

The resulting LC-MS proteomic data were processed using Maxquant software version 1.5.2.8 and searched against a database downloaded from *Medicago truncatula* Genome Database version 5.0 downloaded on 12/3/18. The digestion enzyme was set to trypsin with

up to two missed cleavages and oxidation of methionine, a nd protein N-terminal acetylation was set as variable modifications. Cysteine carbamidomethylation was established as a fixed modification. Reporter ion MS^2 quantification was used with the appropriate 8-plex TMT labels. The match between the runs feature was utilized to decrease missing data values within the data set. Precursor mass tolerance was 20 ppm, and product ions were searched at 4.5 ppm tolerances. Peptides were combined with protein groups based on the rules of parsimony. The data was filtered to a 1% FDR at the peptide-spectrum match and protein levels. Fold change of ubiquitination sites was normalized to protein fold change by subtracting the log 2 protein fold change from the ubiquitination site fold change.

Lipid Sample Preparation and Analysis Plant samples for lipid analysis were prepared as described previously with the exception that samples of roots, shoots, and whole seedlings were made separately for analysis. The ground tissue samples were re-suspended in 1.2 mL MTBE:MeOH (10:3 v/v) and combined with 100 µ L of 180-micron glass beads. The samples were homogenized for 10 minutes with a Mixer Mill MM 400 (Retsch) at 25 Hz and then vortexed for an additional 10 minutes. 225 µ L of water was added, and samples were briefly vortexed before centrifuging for 20 minutes at 13,000 G and 4° C. For lipid analysis 200 µ L of the organic phase was extracted and dried in a vacuum centrifuge before re-suspending in 100 µ L of MeOH:Toluene (9:1 with molecular sieves).

Lipid Data Analysis The resulting LC-MS raw files were converted to mgf files with MSConvertGUI (ProteoWizard, Dr. Parag Mallick, Stanford University)and analyzed with

Compound Discoverer (Thermo Scientific) to generate aligned and unaligned feature tables. The aligned workflow was assembled as follows: input files, select spectra, align retention times, detect unknown compounds, group unknown compounds, fill gaps, and mark background compounds. The unaligned workflow included input files, select spectra, and detect compounds. For both workflows the 'detect unknown compounds' parameters were set as follows: 10 ppm mass tolerance, 100% intensity tolerance, S/N threshold 3, [M+H]+1 and [M-H]-1 ions, and 100,00 minimum peak intensity. For 'fill gaps', parameters were set to 10 ppm mass tolerance, 0.2 minute RT tolerance, and 1.5 S/N threshold.

Using the lipid software LipiDex, feature tables were searched against acetate and plant lipid libraries based on the mobile phase utilized[15]. Finally, using LipiDex Peak Finder, MS^2 identifications are matched with peaks using files generated from Compound Discoverer. Features were required to be identified in a minimum of two files while keeping the defaults of a minimum of 75% of lipid spectral purity, an MS2 search dot product of at least 500 and reverse dot product of at least 700, as well as a multiplier of 2.0 for FWHM window, a maximum 15 ppm mass difference and 3.5 k$*$ σ retention filtering.

Results and Discussion

To make biological comparisons between our samples, we required a di-glycine enrichment method that was sufficiently reproducible. We compared the overlap of ubiquitination sites detected in replicate injections and the overlap between multiple enrichments. 394 di-glycine sites were repeatedly observed in three injection replicates with approximately 75

unique sites found in each injection (84% overlap) (Figure 2a). The coefficient of variation between injections was 10% . Ideally, one would have higher overlap enabling comparisons between samples. To mitigate the low overlap between injection replicates, we utilized TMT quantification which permits sample multiplexing and decreases injections.

When we compared reproducibility across multiple enrichments, we observed similar overlap with 285 ubiquitination sites in all enrichments and between 31-146 unique sites in each replicate. The coefficient of variation between enrichments was 21% (Figure 2b). It is expected that the variation increased since we were comparing different enrichments with separate injections. While the variation that is caused by separate enrichments cannot easily be decreased, TMT quantification can be employed to reduce the number of injections thereby decreasing variation between replicates.

We observed an inherent charge difference of modified peptides due to the presence of a second N-terminus[16]. Di-glycine peptides possess on average a 3+ charge state compared to the 2+ charge of unmodified peptides (Figure 3a). We leveraged this difference with an instrument method that selects peptides with charge states 3-4 instead of 2-4 for MS2 analysis. Figure 3b displays the resulting increase in identifications across all protein quantities tested (replicate enrichments were only performed for the 10 mg protein quantity). Focusing our instrument time on peptides with a higher charge state increases our likelihood of selecting modified peptides with each scan.

To maximize the number of identified ubiquitination sites we performed enrichments with input material ranging from 1-10 mg. Due to its low abundance and transient nature in

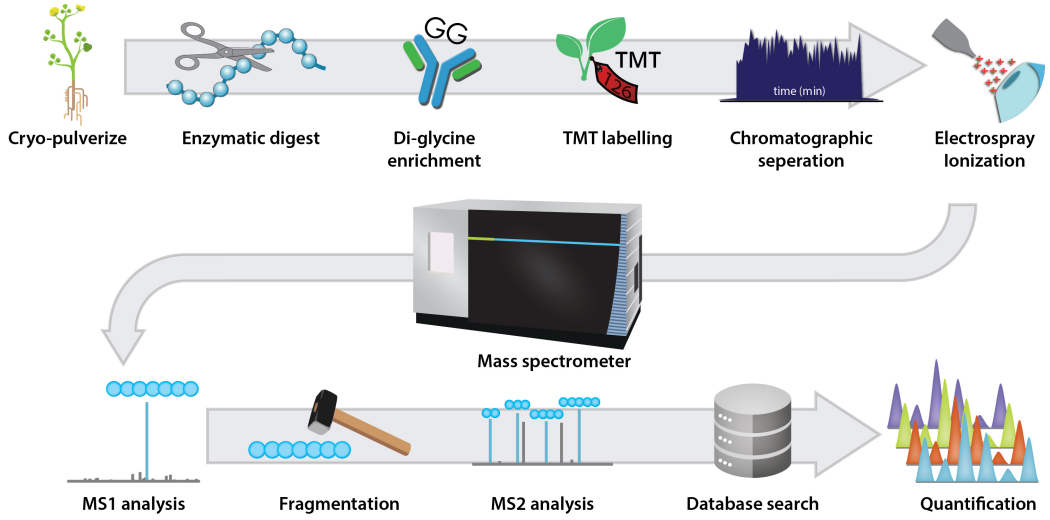


Figure 4.1: *Medicago truncatula* seedlings were digested, enriched for di-glycine peptides, and labeled with TMT tags before analyzing with a high-resolution Orbitrap Fusion Lumos.. Whole seedlings were ground under liquid nitrogen and digested with LysC and trypsin before analyzing with mass spectrometry and searching against a database downloaded from the*Medicago truncatula*genome database.

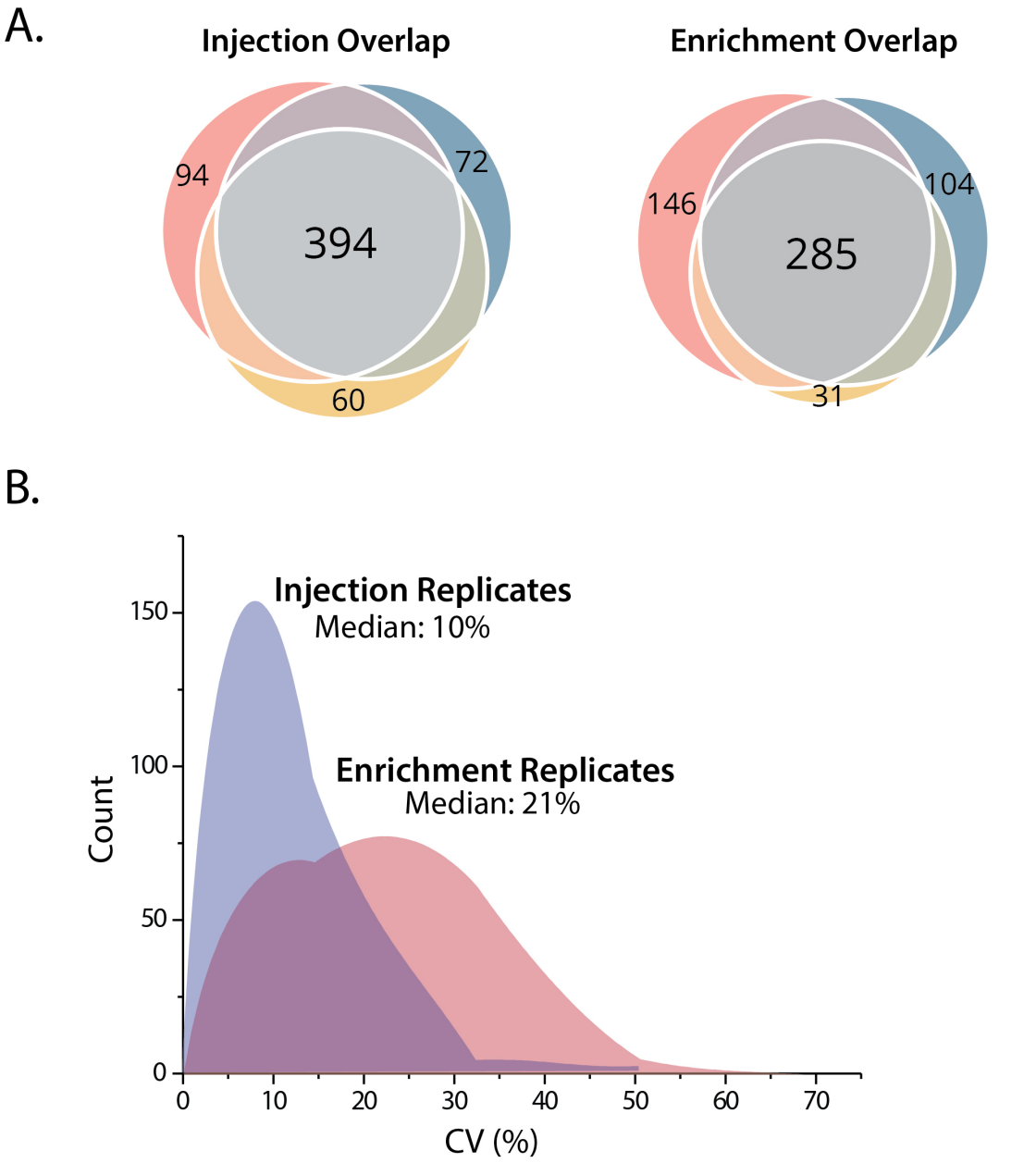


Figure 4.2: Good overlap is observed between both injection and enrichment replicates.. Three injection replicates were performed back to back from a single enrichment sample, and three enrichment replicates were each collected from 10 mg mouse brain protein sample. (**B**) The coefficient of variation was calculated between each set of three replicates as the standard deviation divided by the mean. The median standard deviation of injection replicates is 10% and the median standard deviation of enrichment replicates is 21% .

the cell, we expected large quantities of peptides to be necessary to detect the modification. Other attempts to enrich for ubiquitinated peptides have used from 10-40 mg of starting material[17–20]. We settled on an upper limit of 10 mg of material for our experiments due to the protein yield from *M. truncatula* seedlings and time required to grow additional plants. Figure 3b indicates a linear relationship between the input protein and the total number of sites identified. For this reason, all of our enrichments were performed with 10 mg of protein.

This method was then applied to *M. truncatula* wild type and *pub1* seedlings treated with and without rhizobial signal molecules called Nod factors. Four replicates of each condition were prepared for a total of sixteen unique samples. Using our developed enrichment with 10 mg of protein for each sample and selecting for triply and quadruply charged precursors, we identified 169 ubiquitination sites on 112 unique proteins (Supplementary Table 1). This experiment was technically challenging due to the magnitude of input material required; over 160 mg of protein was digested and prepared for enrichment. We avoided the use of proteasome inhibitors such as MG-132 as these treatments can activate the symbiosis pathway in the absence of any symbiont[21]. The use of inhibitors can also give a significant boost to the number of sites identified by preventing degradation by the proteasome and allowing the accumulation of ubiquitinated proteins[16].

In addition to identifying ubiquitination sites, we also separately performed analyses of the whole proteome of the same samples using label-free quantification. This resulted in the identification of 5,597 proteins (Supplementary Table 2). A principal component

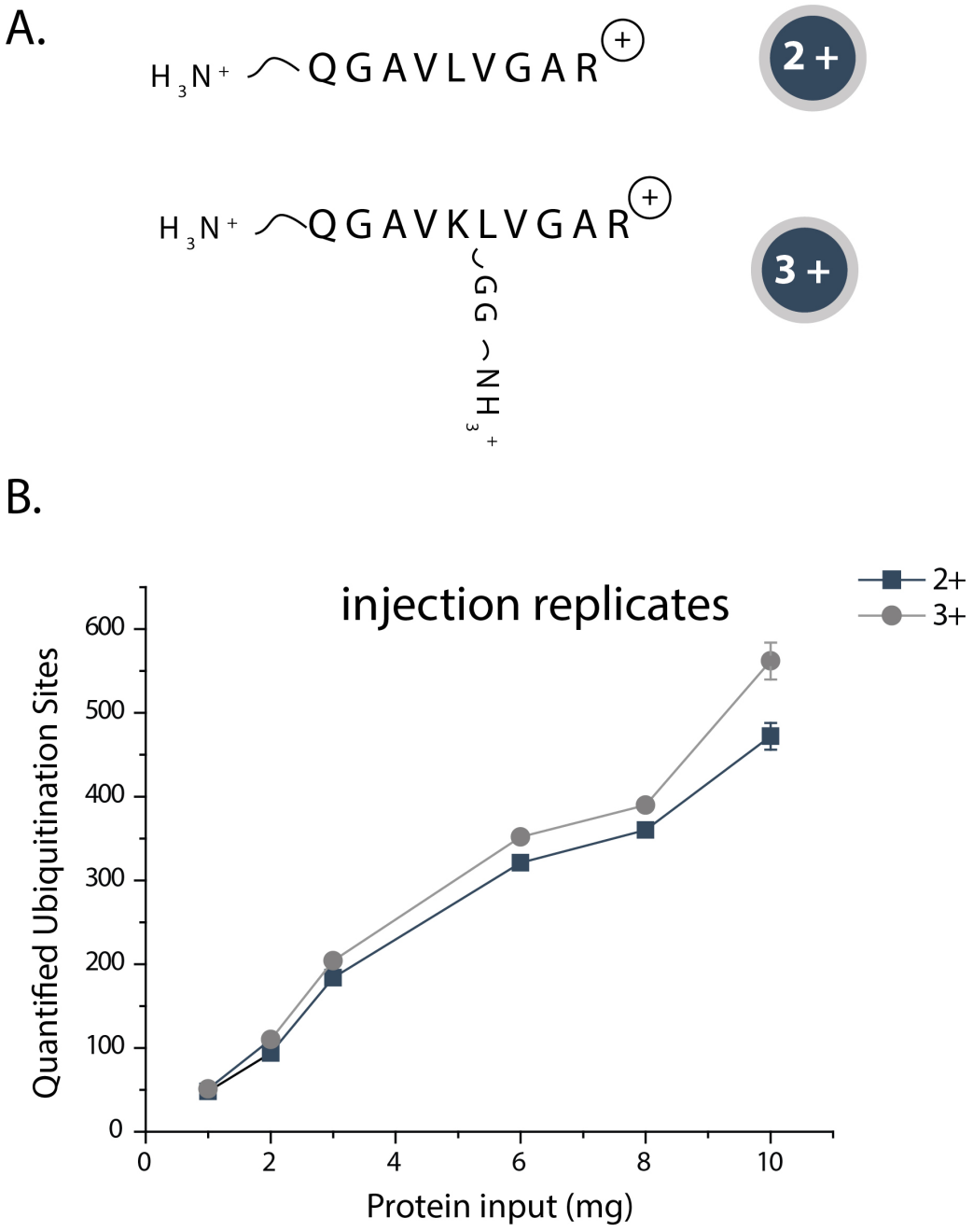


Figure 4.3: Di-glycine peptides have a higher charge state on average which can be exploited when designing instrument methods.. Due to the additional N-terminus on di-glycine peptides they possess an average charge state of 3+ instead of 2+. (**B**) Instrument methods that select charge states from 3-4 result in more di-glycine identifications than methods that select charge states 2-4. Di-glycine identifications linearly increase with the amount of protein input in the assay. 10 mg of protein is the maximum we could feasibly process for all samples. Injection replicates are only shown for the 10 mg protein input.

analysis (PCA) revealed that the plant genotype contributed to the majority of protein changes observed, (Fig. 4) while treatment with Nod factors had minimal effect[22].

We detected a significant reduction in the ubiquitination of multiple oleosin proteins in *pub1* compared to the wild-type seedlings, both in the presence and absence of Nod factors (Fig. 5 A, B). This reduction in ubiquitination corresponded with an increased protein abundance (Fig. 5 C, D).

Oleosins are found in abundance in lipid droplets inside a cell[23]. Lipid droplets are composed of a triglyceride matrix surrounded by a phospholipid monolayer penetrated by oleosins. Given the decreased ubiquitination of oleosins in *pub1*, we profiled wild-type and *pub1* seedlings, shoots, and roots for lipids. Our choice of tissue samples enabled us to determine where within the plant oleosins are accumulating. The lipidomics experiments revealed that the biggest difference in the lipid profiles between the gentotypes is in the roots, followed by shoots and then the seedlings. We detected several triglycerides in the roots whose abundance was significantly higher in the *pub1* mutant both in the presence and absence of Nod factor (Figure 6 A,B). We also found phosphatidylcholine and phosphatitidyl glycerol to be significantly different between the genotypes in the roots. While some triglycerides were up-regulated in all kinds of tissues profiled, the most significant number of triglycerides were found to be changing in the root. The increased abundance of multiple lipids in the *pub1* mutant strongly support the involvement of PUB1 in lipid droplet mobilization via ubiquitination of oleosins.

In the *M. truncatula* proteome atlas that we previously developed, we had detected

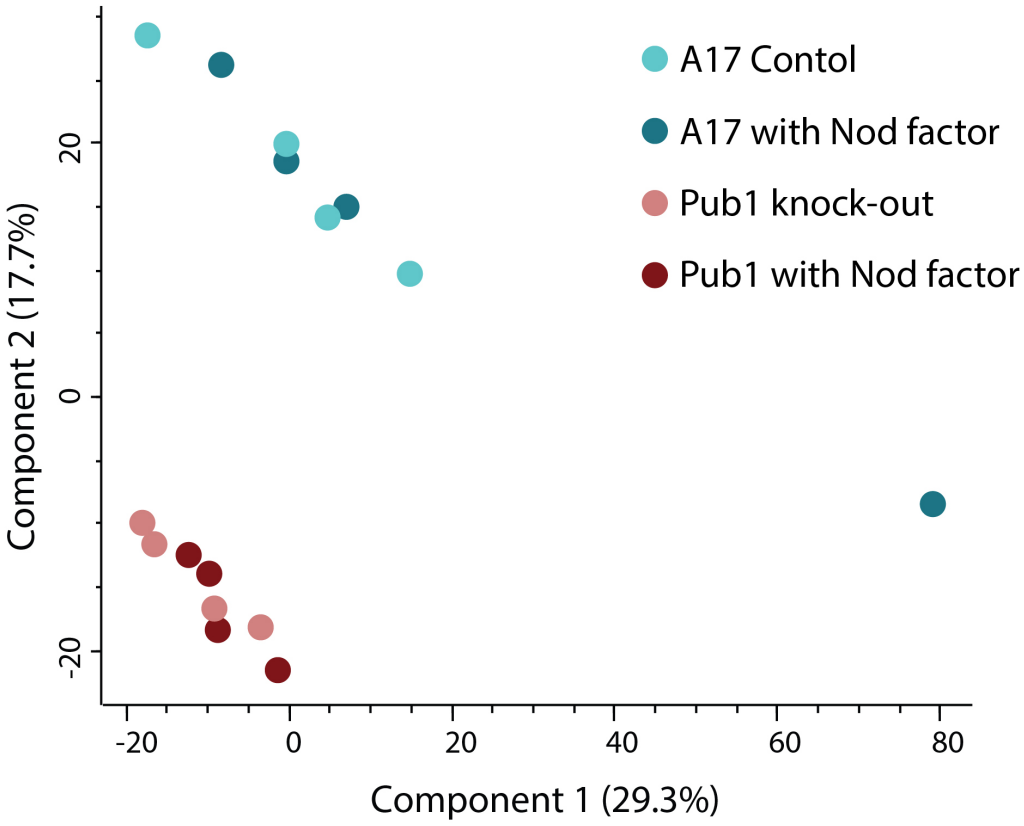


Figure 4.4: Principal component analysis of protein samples indicates that the majority of variation is caused by plant genotype.. A PCA plot was generated using Perseus and shows no discernable difference between the samples treated with and without Nod factors.

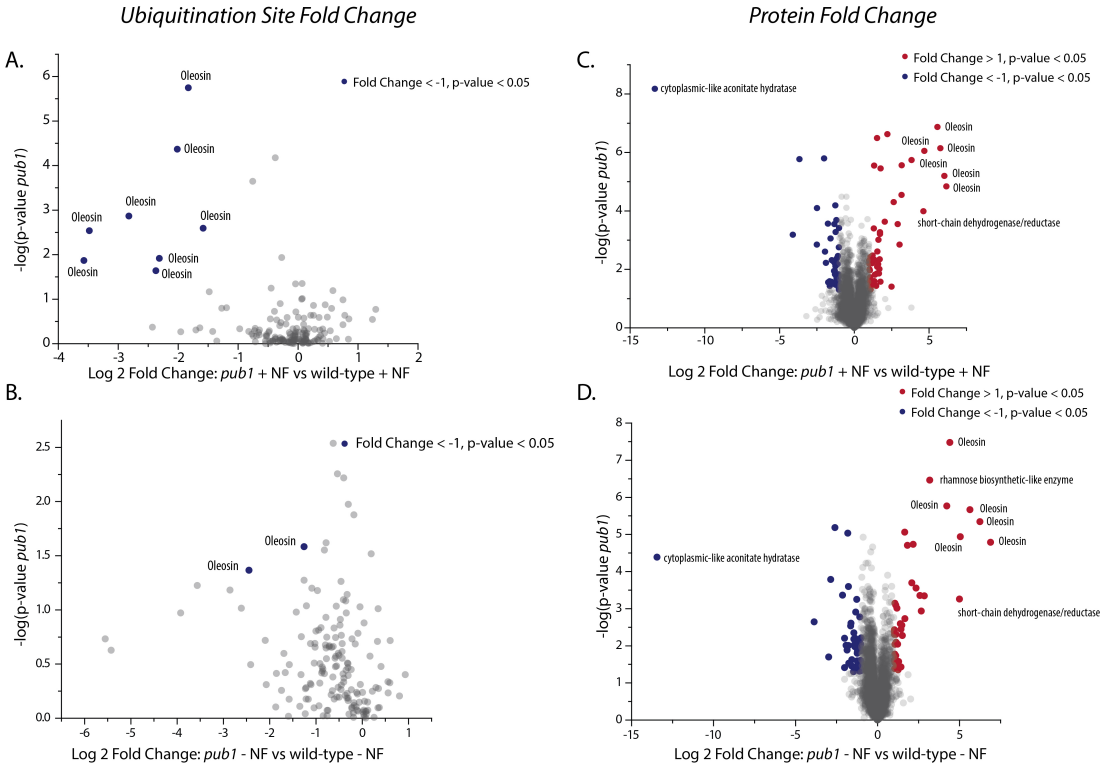


Figure 4.5: Volcano plots showing the significance and fold change of ubiquitination sites and protein abundance changing in *pub1* relative to the wild-type in the presence (A,C) or absence (B,D) of Nod factor in whole ground seedlings. Ubiquitination on oleosin proteins is significantly decreased in *pub1* both in the presence and absence of Nod factors, while protein abundance is increased.

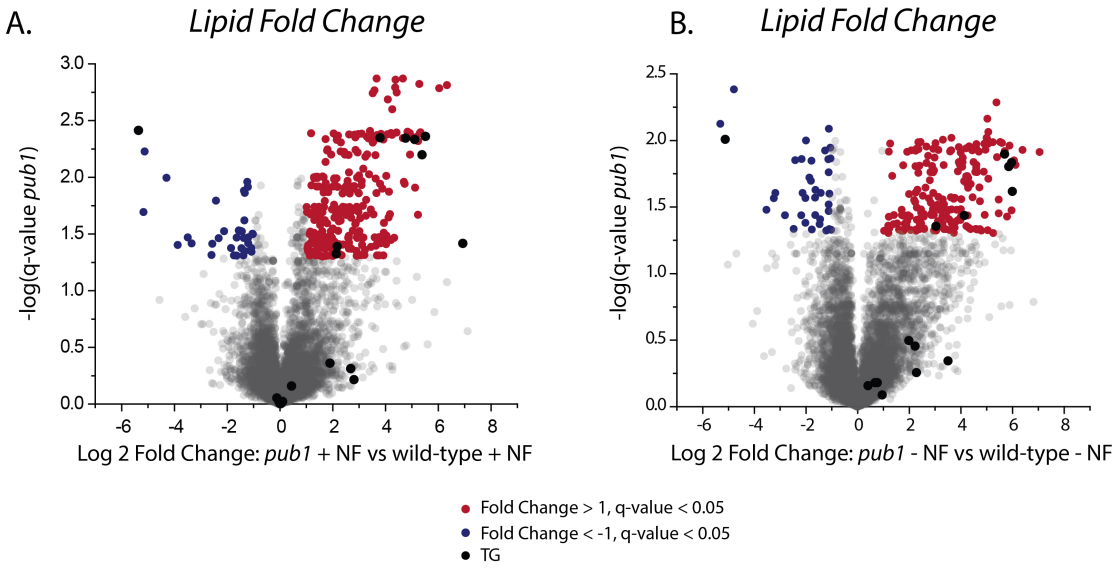


Figure 4.6: Volcano plot showing the significance and fold change of lipids changing in *pub1* relative to the wild-type in the presence (A) or absence (B) of Nod factor in plant roots. Lipid droplet-associated triglycerides, which are bound by oleosins are significantly up-regulated in *pub1* both in the presence and absence of Nod factor.

oleosins in the seed proteome as well as observed their phosphorylation[24]. Here, we found evidence of the ubiquitination of oleosins. Oleosins are present throughout the plant kingdom and are noted for their appearance in lipid droplet membranes, especially in seeds. They stabilize lipid droplets by forming long hydrophobic hairpins penetrating the triacylglycerol membrane of the droplets and also regulate lipid droplet synthesis[23]. During seed development in *M. truncatula,* the expression of specific oleosin genes corresponds to an increased lipid accumulation[25]. Post seed germination, lipid droplets are gradually mobilized to supply for the energy demands of the growing seedling, and the oleosins are degraded in this process[26]. In sesame, oleosins are ubiquitinated for degradation after seed germination[27]. Lipid droplets are degraded by lipases to generate fatty acids which, in turn, produce glucose via gluconeogenesis[23]. The fact that several oleosins are not degraded in *pub1* might indicate that these plants are not able to meet their energy requirements, and as a consequence, development is delayed. Indeed, a 10% reduction in total root length two weeks after inoculation with mycorrhizal fungi as well as in uninoculated plants have been reported in the *pub1* mutant[13]. This involvement of PUB1 in root growth is in accordance with our lipid profiling results, where we found several triglycerides to be significantly higher in the roots in the absence of PUB1, which in turn, might reflect inefficient lipid mobilization in the mutant. No such function of PUB proteins is yet known in plants; involvement of Ub-ligase in lipid droplet mobilization has been reported in the context of Troyer syndrome in humans[28].

Conclusion

Large-scale studies on ubiquitination have been conducted in *Arabidopsis thaliana* [29,30]. These studies used the proteasomal inhibitor MG132, which has been shown to mimic the effect of inoculation with rhizobia in the absence of this bacterium *M. truncatula* [21]. To circumvent this issue, here we developed an optimized method for the enrichment of ubiquitinated peptides and applied this method to the study of *M. truncatula* seedlings. To minimize the variation that we observed in injection replicates, we leveraged TMT quantification to label enriched di-glycine peptide samples. An optimum protein input of 10 mg was selected, and all of our samples were collected without the use of proteasome inhibitors which may artificially perturb results. While collecting data on the mass spectrometer, we selected peptides with a charge state 3-4 to increase the probability of choosing a di-glycine peptide.

These methods yielded 169 ubiquitination sites on 112 unique proteins from *M. truncatula* and to our knowledge is the first study of ubiquitination in this organism. We further observed decreased ubiquitination of several oleosin proteins and an increased abundance of multiple triglycerides in the *pub1*mutant compared to wild-type plants. The frequent association of PUBs with membrane proteins and lipids and the preponderance of oleosin in the membranes of lipid droplets within a cell allude to the potential role of PUB1 in lipid droplet mobilization. The precise mechanisms by which PUB1 regulates oleosin and triglyceride degradation and whether they play a role in rhizobia-legume symbiosis are important questions that remain to be addressed.

Acknowledgements We thank the University of Wisconsin-Madison and the National Science Foundation DBI0701846 (awarded to JJC) for providing financial support for this work.

References

[1] D. Komander and M. Rape, "The Ubiquitin Code," *Annual Review of Biochemistry*, vol. 81, pp. 203–229, jun 2012.

[2] B. Sharma, D. Joshi, P. K. Yadav, A. K. Gupta, and T. K. Bhatt, "Role of Ubiquitin-Mediated Degradation System in Plant Biology," *Frontiers in Plant Science*, vol. 7, jun 2016.

[3] M. Trujillo, "News from the PUB: Plant U-box type E3 ubiquitin ligases," *Journal of Experimental Botany*, vol. 69, pp. 371–384, jan 2018.

[4] A. Miricescu, K. Goslin, and E. Graciet, "Ubiquitylation in plants: Signaling hub for the integration of environmental signals," *Journal of Experimental Botany*, vol. 69, pp. 4511–4527, aug 2018.

[5] W. Kim, E. J. Bennett, E. L. Huttlin, A. Guo, J. Li, A. Possemato, M. E. Sowa, R. Rad, J. Rush, M. J. Comb, J. W. Harper, S. P. Gygi, A. Possemato, J. Rush, M. E. Sowa, E. L. Huttlin, W. Kim, J. Li, R. Rad, E. J. Bennett, A. Guo, J. W. Harper, S. P. Gygi, E. L. Huttlin, A. Guo, J. Li, A. Possemato, M. E. Sowa, R. Rad, J. Rush, M. J. Comb, J. W. Harper, and S. P. Gygi, "Systematic and quantitative assessment of the ubiquitin-modified proteome," *Molecular Cell*, vol. 44, pp. 325–340, oct 2011.

[6] M. Kai, N. Ueno, and N. Kinoshita, "Phosphorylation-dependent ubiquitination of paraxial protocadherin (PAPC) controls gastrulation cell movements," *PLoS ONE*, vol. 10, jan 2015.

[7] D. Marino, N. Peeters, and S. Rivas, "Ubiquitination during Plant Immune Signaling," *Plant Physiology*, vol. 160, pp. 15–27, sep 2012.

[8] D. Kelley, "E3 ubiquitin ligases: key regulators of hormone signaling in plants," *Molecular & Cellular Proteomics*, p. mcp.MR117.000476, jan 2018.

[9] M. Sharma, A. Pandey, G. K. Pandey, A. Ciechanover, A. Hershko, and I. Rose, "Role of Plant U-BOX (PUB) Protein in Stress and Development," *Plant Stress*, no. 1, pp. 1–9, 2013.

[10] J. Song, X. Mo, H. Yang, L. Yue, J. Song, and B. Mo, "The U-box family genes in Medicago truncatula: Key elements in response to salt, cold, and drought stresses," *PLoS ONE*, vol. 12, p. e0182402, aug 2017.

[11] C. Herve, B. Lefebvre, and J. Cullimore, "How many E3 ubiquitin ligase are involved in the regulation of nodulation?," *Plant Signaling and Behavior*, vol. 6, pp. 657–661, may 2011.

[12] M. Mbengue, S. Rivas, S. Camut, L. Deslandes, D. Klaus-Heisen, S. Froidure, T. Timmers, C. Hervé, J. Cullimore, B. Lefebvre, S. Moreau, and F. de Carvalho-Niebel, " The Medicago truncatula E3 Ubiquitin Ligase PUB1 Interacts with the LYK3 Symbiotic Receptor and Negatively Regulates Infection and Nodulation ," *The Plant Cell*, vol. 22, pp. 3474–3488, oct 2010.

[13] C. Camps, M. Mbengue, S. Camut, V. Gasciolli, C. Rembliere, R. Thompson, T. Timmers, C. Hervé, J. Cullimore, C. le Signor, T. Vernié, B. Lefebvre, and F. de Carvalho-Niebel, " PUB1 Interacts with the Receptor Kinase DMI2 and Negatively Regulates Rhizobial and Arbuscular Mycorrhizal Symbioses through Its Ubiquitination Activity in Medicago truncatula ," *Plant Physiology*, vol. 170, pp. 2312–2324, apr 2016.

[14] Z. Lu, J. Dong, P. Qin, T. Wang, F. Zhu, Y. Li, J. Deng, and J. Liu, "The MtDMI2-MtPUB2 Negative Feedback Loop Plays a Role in Nodulation Homeostasis," *Plant Physiology*, vol. 176, pp. 3003–3026, apr 2018.

[15] P. D. Hutchins, J. D. Russell, and J. J. Coon, "LipiDex: An Integrated Software Package for High-Confidence Lipid Identification," *Cell Systems*, vol. 6, pp. 621–625.e5, may 2018.

[16] C. M. Rose, M. Isasa, A. Ordureau, M. A. Prado, S. A. Beausoleil, M. P. Jedrychowski, D. J. Finley, J. W. Harper, and S. P. Gygi, "Highly Multiplexed Quantitative Mass Spectrometry Analysis of Ubiquitylomes," *Cell Systems*, vol. 3, pp. 395–403, oct 2016.

[17] V. G. Anania, V. C. Pham, X. Huang, A. Masselot, J. R. Lill, and D. S. Kirkpatrick, "Peptide Level Immunoaffinity Enrichment Enhances Ubiquitination Site Identification on Individual Proteins," *Molecular & Cellular Proteomics*, vol. 13, pp. 145–156, jan 2014.

[18] D. Bustos, C. E. Bakalarski, Y. Yang, J. Peng, and D. S. Kirkpatrick, "Characterizing Ubiquitination Sites by Peptide-based Immunoaffinity Enrichment," *Molecular & Cellular Proteomics*, vol. 11, pp. 1529–1540, dec 2012.

[19] N. D. Udeshi, P. Mertins, T. Svinkina, and S. A. Carr, "Large-scale identification of ubiquitination sites by mass spectrometry," *Nature Protocols*, vol. 8, pp. 1950–1960, oct 2013.

[20] S. A. Wagner, P. Beli, B. T. Weinert, M. L. Nielsen, J. Cox, M. Mann, and C. Choudhary, "A Proteome-wide, Quantitative Survey of In Vivo Ubiquitylation Sites Reveals Widespread Regulatory Roles," *Molecular & Cellular Proteomics*, vol. 10, p. M111.013284, oct 2011.

[21] H. Pan, C. Stonoha-Arther, and D. Wang, " Medicago Plants Control Nodulation by Regulating Proteolysis of the Receptor-Like Kinase DMI2 ," *Plant Physiology*, vol. 177, pp. 792–802, jun 2018.

[22] S. Tyanova, T. Temu, P. Sinitcyn, A. Carlson, M. Y. Hein, T. Geiger, M. Mann, and J. Cox, "The Perseus computational platform for comprehensive analysis of (prote)omics data," *Nature Methods*, vol. 13, pp. 731–740, jun 2016.

[23] A. H. Huang, "Plant Lipid Droplets and Their Associated Proteins: Potential for Rapid Advances," *Plant Physiology*, vol. 176, pp. 1894–1918, mar 2017.

[24] H. Marx, C. E. Minogue, D. Jayaraman, A. L. Richards, N. W. Kwiecien, A. F. Siahpirani, S. Rajasekar, J. Maeda, K. Garcia, A. R. Del Valle-Echevarria, J. D. Volkening, M. S. Westphall, S. Roy, M. R. Sussman, J. M. Ané, and J. J. Coon, "A proteomic atlas of the legume Medicago truncatula and its nitrogen-fixing endosymbiont Sinorhizobium meliloti," *Nature Biotechnology*, vol. 34, pp. 1198–1205, nov 2016.

[25] L. Wang, R. B. Peterson, and T. P. Brutnell, "Regulatory mechanisms underlying C4photosynthesis," *New Phytologist*, vol. 190, pp. 9–20, apr 2011.

[26] A. J. Sami, N. Ahmad, and A. R. Shakoori, "Computational analysis of an acidic lipase of Tribolium castaneum," *Pakistan Journal of Zoology*, vol. 46, pp. 805–812, feb 2014.

[27] E. S. L. Hsiao and J. T. C. Tzen, "Ubiquitination of oleosin-H and caleosin in sesame oil bodies after seed germination," *Plant Physiology and Biochemistry*, vol. 49, pp. 77–81, jan 2011.

[28] S. W. Eastman, M. Yassaee, and P. D. Bieniasz, "A role for ubiquitin ligases and Spartin/SPG20 in lipid droplet turnover," *Journal of Cell Biology*, vol. 184, pp. 881–894, mar 2009.

[29] V. Aguilar-Hernández, D. Y. Kim, V. Aguilar-Hernández, D. Y. Kim, R. J. Stankey, R. D. Vierstra, M. Scalf, and L. M. Smith, "Mass Spectrometric Analyses Reveal a Central Role for Ubiquitylation in Remodeling the Arabidopsis Proteome during Photomorphogenesis," *Molecular Plant*, vol. 10, pp. 846–865, jun 2017.

[30] D.-Y. Kim, M. Scalf, L. M. Smith, and R. D. Vierstra, " Advanced Proteomic Analyses Yield a Deep Catalog of Ubiquitylation Targets in Arabidopsis ," *The Plant Cell*, vol. 25, pp. 1523–1540, may 2013.

Chapter 5

PROTEIN TURNOVER IN THE MOUSE

This chapter is part of the manuscript in preparation:

Weisenhorn, E.M., Wilson, G., Kwiecien, N., Wilkerson, E., Rose, C., Keller, M., Attie, A., Coon JJ. *Protein Turnover in the Mouse* **2019**

Abstract

Protein turnover is a dynamic process determined by the rates of protein synthesis and degradation and rates can vary over several orders of magnitude. Due to the ease of incorporating metabolic labels, most large studies of turnover have focused on single-celled organisms. Mammalian studies, however, have been smaller in scope or limited to a handful of tissues. We calculated turnover rates in nine mouse tissues (totaling 8,149 unique proteins) and analyzed their dependence on physical and biological properties. Our study is limited in our ability to calculate protein half-lives by the number of time points available to perform curve fitting (four). This necessarily limits the range of half-lives we can measure and we report here half-lives ranging from 4-20 days although many rates outside this range were calculated and are reported as <4 or >20 days. Measuring these proteins across multiple tissues, we found that tissue localization often determines protein half-life. Our study shows that MS method can significantly impact measured half-lives, and improving signal-to-noise ratio with peptide fractionation results in the calculation of shorter and more accurate half-lives. This study catalogs a dimension of proteins regularly absent from proteomic analyses and provides a resource for studying diverse areas of biology, particularly those implicated in aging and disease.

Introduction

A hallmark of aging, the decline of protein turnover rate results in abnormal accumulation of protein inclusions and aggregates[1–6]. These and other disruptions to proteostasis contribute to a growing list of disorders, including neurodegenerative diseases, cancer, diabetes, drug withdrawal, and cardiovascular diseases[7–12]. Not surprisingly, then, protein half-life is one of the best indicators for drug target likeness[13,14]. If a drug is intended to bind a protein target and modulate activity then longer-lived proteins are most suitable for targeting. But because disruptions to turnover dynamics sometimes do not result in altered protein abundance, the altered protein half-life phenotype eludes detection in conventional quantitative proteomic experiments[15–19]. A large-scale characterization of turnover rates would therefore provide an opportunity to better understand the mechanisms of age-related diseases. Further, we expect such a catalog of turnover rates to serve as a resource for guiding target selection during drug development.

Using metabolic labeling and MS, we have assembled a global repository of protein half-lives from nine mouse tissues. Mouse tissues collected from C57BL/6J mice by Baughman et al. to monitor NeuCode label incorporation were re-purposed to calculate turnover rates[20]. We leveraged this resource to interrogate functional and biological roles and, more specifically, differential protein turnover between tissues of origin. Previous turnover studies have found that factors such as degradation motifs and cellular localization can regulate protein half-lives[21–23]. Most studies of turnover on a whole-proteome scale have relied on

cell culture and single-celled organisms, for which sample collection and labeling prove relatively straight forward[24–27]. Further, most studies conducted on a whole animal model are limited to three to four tissues[15,28–32]. The scope of our data allows us to further these studies and investigate the role of many biological and physical characteristics. Additional studies have been conducted in other model organisms, including *Danio rerio*, *Arabidopsis thaliana*, *Escherichia coli*, *Gallus gallus*, and *Myodes glareolus*[28,33–36]. We have chosen to collect these data from a whole mouse model: first, we are able to compare protein turnover rates across tissues by analyzing a wide range of tissues, and second, the ubiquitous use of the mouse for research promises wide utility for the scientific community.

Materials and Methods

Mice and labeling Male C57BL/6J mice (6 weeks; the Jackson Laboratory) were fed laboratory control diet for two weeks before being fed a customized lysine-free diet (Harlan, Madison, WI) combined with 1% natural light lysine (K000). Starting at ten weeks of age, mice were fed ad libitum lysine-free diet containing 1% K602 (n = 12) or 1% K080 (n = 12) (Cambridge Isotopes, Boston, MA).

A second mouse cohort were fed a custom leucine-free diet combined with 1% natural light leucine (L000). After ten weeks mice were fed ad libitum leucine-free diet containing 1% L601 (n=12). Food consumption and body weight were monitored throughout the experiment. Mice raised on isotope-labeled diets were sacrificed at four time points (3, 10, 20, 30 days). After sacrificing animals by cervical dislocation, tissues were dissected,

washed in phosphate-buffered saline (PBS), and frozen in liquid nitrogen. These mice are identical to those previously analyzed by Baughman and Rose et al. 2016[20].

Lysine mice preparation and digestion Tissue samples were pulverized using a Qiagen TissueLyzer II in 1-2ml of 9M urea in 20mM Hepes with ½X tablet of Complete protease inhibitor and 1 mM sodium orthovanadate, 2.5 mM sodium pyrophosphate, and 1 mM ß- glycerophosphate. Lysate was cleared by centrifugation at 20,000 g. Following protein concentration estimation by BCA assay, a small amount (<1 mg) of each lysate was taken for analysis. Protein disulfide bonds were reduced by addition of 5 mM dithiothreitol (DTT) and incubation for 45 min at 37° C. Free thiols were alkylated by the addition of 15 mM iodoacetamide and incubation in the dark at room temperature for 30 min. The alkylation reaction was quenched by addition of 5 mM DTT. Proteolytic digestion was performed by addition of Lys-C (Wako) at a 1:100 enzyme-to-protein ratio and incubation at 37° C for 2 hr. The urea concentration was then diluted to 4 M using 50 mM Tris, 3 mM CaCl2, and another bolus of Lys-C was added at a 1:100 enzyme-to-protein ratio. The sample was then incubated overnight at room temperature while rocking. The digestion was quenched by the addition of TFA to 1% and then desalted with tC18 Sep-Pak cartridges (Waters).

Leucine mice preparation and digestion Tissue samples were pulverized using a Qiagen TissueLyzer II in 1-2ml of 9M urea in 20mM Hepes with ½X tablet of Complete protease inhibitor and 1 mM sodium orthovanadate, 2.5 mM sodium pyrophosphate, and 1 mM

ß- glycerophosphate. Lysate was cleared by centrifugation at 20000 g. Following protein concentration estimation by BCA assay, a small amount (<1 mg) of each lysate was taken for analysis. Protein was extracted using a 90% methanol precipitation and the pellet was re-suspended in 150 µ l lysis buffer (8 M Urea, 100 mM Tris, 20 mM TCEP, 80 mM Chloroacetamide) and diluted with 50 mM Tris to a Urea concentration of 1.5 M. Protein digestion was performed overnight with trypsin (1:100) before de-salting with 10 mg Strata C18 solid phase extraction cartridges and dried in a vacuum centrifuge.

Samples were fractionated using high-pH reverse-phase separation to increase proteomic depth. The solvent system consisted of mobile phase A (20 mm ammonium bicarbonate) and mobile phase B (20 mm ammonium bicarbonate 80% acetonitrile) which was run on an Ultimate 3000 UPLC system (Dionex Sunnyvale, CA) with a reverse-phase C18 column. Gradient elution was performed at 400 µ l min− 1 with the gradient increased from 0 to 6% B over 5 min followed by an increase to 80% B until 24 min and a wash at 100% B for 3 min. Forty fractions were collected from each sample which were subsequently pooled, resulting in 20 total fractions per sample. In addition to the fractionated samples, an un-fractionated analysis of the leucine labeled brain and liver was collected for comparison.

Lysine mice mass spectrometry and high-performance liquid chromatography Online reverse-phase chromatography was performed using a nanoAcquity UPLC (Waters, Milford, MA) or Easy-nanoLC 1000 (Thermo Fisher Scientific, San Jose, CA). Peptides were eluted over an analytical column (75 µ m ID) heated to 60° C and packed with 30 cm of 1.7 µ

m diameter, 130 Å pore size, Bridged Ethylene Hybrid C18 particles (Waters). Mobile phase A was composed of water, 0.2% formic acid, and 5% DMSO. Mobile phase B was composed of acetonitrile and 0.2% formic acid. The gradient was optimized to ensure even elution of peptides over a 70 min period. Eluted peptide cations were converted to gas-phase ions by ESI and analyzed on an Orbitrap Elite mass spectrometer (Thermo Fisher Scientific). A survey scan was performed in the Orbitrap at 30,000 resolving power to identify precursors to sample for data-dependent, top-20 ion trap CAD MS/MS (rapid scan analysis). An additional quantitative 480,000 resolving power scan immediately followed the survey scan. Ion trap MS/MS scans were performed while the FT transient was collected by enabling “Preview Mode." Monoisotopic precursor selection was on and precursors with unknown charge or charge of +1 were excluded from MS/MS. MS1 and MS/MS target-ion accumulation values were set to $1x10^6$ and $5x10^3$, respectively. Dynamic exclusion was set to 45 s for -25 ppm and +15 ppm around the selected precursor.

Leucine mice mass spectrometry and high performance liquid chromatography Samples were analyzed using a LC-MS instrument comprising an Orbitrap Fusion Lumos Tribrid mass spectrometer and Ultimate 3000 RSLCnano liquid chromatography system (Thermo Fisher Scientific). Mobile phase A consisted of 0.2% formic acid in water and mobile phase B consisted of 0.2% formic acid in 70% acetonitrile. A 75-min gradient ranging from 0% to 60% B was employed spanning a total runtime of 90 min. Analytes were injected onto a 1.7 μ m C18 column (75 μ m i.d.) packed in-house to a length of 35 cm and heated to 45° C.

Survey scans of peptide precursors were collected every second from 300-1350 Th with an AGC target of $1x10^6$ and a resolution of 240,000 in the orbitrap. Precursors were isolated from a 1.6 Th window in the quadrupole and HCD MS/MS scans at 30% collision energy were collected in the orbitrap with an AGC target of $1x10^4$ from 100-1200 Th.

Data Analysis The resulting LC-MS proteomic data were processed using Maxquant software version 1.5.2.8 and searched against a *Mus musculus* database downloaded from Uniprot on 1/5/16. The digestion enzyme was set to LysC with up to two missed cleavages and oxidation of methionines and protein N-terminal acetylation were set as variable modifications. Leucine 601, and Leucine 601 with N^{15} loss were set as variable modifications for mice fed the heavy leucine diet. Mice fed either the K602 or K080 isotopologues were searched with the appropriate amino acid as a variable modification. Cysteine carbamidomethylation was set as a fixed modification. The match between runs feature was utilized to decrease missing data values within the data set. Precursor mass tolerance was 20 ppm and product ions were searched at 0.5 Da tolerances. Peptides were combined to protein groups based on the rules of parsimony. The data was filtered to a 1% FDR at the peptide-spectrum match and protein levels. This search was used for the calculation of RIA_{inf} as described below. The data were subsequently searched a second time in an identical fashion with the exception of heavy lysine or leucine set as a fixed-label modification to assess the extent of isotopologue incorporation into protein populations. These results across temporal replicates served as inputs for non-linear regression analysis and

the generation of protein turnover rates.

Calculation of precursor RIA and turnover rates A first-order curve was fitted to the di-leucine and di-lysine peptides to obtain the theoretical maximum RIA plateau (RIA_{inf}) using non-linear regression. The horizontal asymptote, or RIA_{inf}, was calculated for each tissue and was used as an estimate for RIA_p and the upper limit for non-linear curve fitting when calculating turnover rates. The calculated RIA_{inf} values for each tissue ranged from 0.8-1.2. In theory the RIA_{inf} should never exceed one and any result above this is due to experimental error. In the case where a RIA_{inf} value was calculated as above one, the upper limit was set as one when calculating protein turnover.

Once a value of RIA_p was calculated, protein half-lives were calculated using all lysine and leucine RIA values. RIA measurements at each time point were calculated from MaxQuant results as the ratio of heavy-labeled protein intensity to total protein intensity. The process of extracting heavy:total intensity ratios was automated with in-house software written in C# . Non-linear curve fitting was performed in all of the tissues using the log-logistic two parameter function (LL.2) in the drc R package[37]. This function accommodates the delayed exponential behavior that is observed from the delay in delivering amino acids from the digestion of the labeled food source[31]. A time point at zero days with an RIA value of zero was imposed on all proteins and the upper-limit was set at the RIA_{inf} calculated for each tissue. A minimum of three of four time points were required for a protein to be modeled and a maximum residual error was set at 0.15.

Results and Discussion

Protein Half-Lives in Mouse Tissue We calculated protein half-lives from a SILAM experiment by measuring the ratio of heavy amino acid incorporation[20,38], collecting triplicate samples after 3, 10, 20, and 30 days on the diet (Figure 1a). Using a shotgun proteomics method on an Orbitrap Fusion Lumos[39], we analyzed proteins from nine harvested tissues: plasma, heart, brain, lung, kidney, intestine, liver, muscle, and islets. In what we will refer to as the survey study, all tissues were analyzed using a 90-minute data-dependent acquisition method without pre-fractionation[40,41]. To obtain additional proteomic depth, peptides from brain and liver were further separated into 20 fractions using high-pH reverse-phase chromatography. In all, 7,424 unique protein half-lives were measured from the in-depth brain and liver study. The complete raw files and results are available for viewing in the PRIDE database under identifier PXD011838.

From these data we calculated relative isotope abundance (RIA) at each time point by calculating the ratio of heavy-labeled peptide to total peptide[42]. To correct for differences in the ratio of isotope abundance in the precursor pool (RIA_p) in each a tissue, di-leucine and di-lysine peptides were analyzed to determine a horizontal asymptote. This asymptote, or RIA∞ , is used as an approximation for the maximum abundance of heavy amino acid in each tissue[15,28]. Protein half-life is defined as the time for half of the protein in a pool to acquire the heavy labeled amino acid. Here they were calculated using the drc R package[37] with the upper limit set to the calculated RIA_∞ for each tissue and a stringent cut-off of

0.15 residual error. To accurately perform the fit, we required that each protein be detected in three of the four time points.

On average we measured 921 protein half-lives in each of the tissues (Figure 1b). Plasma—a challenging mixture consisting of several high-abundance proteins[43]—contained the fewest measured half-lives at 166. 130 protein half-lives were measured in eight of the nine tissues with plasma, displaying the least overlap with the others. These data present an excellent opportunity to compare protein half-lives across tissues and identify factors that regulate turnover rate (Supplementary Table 2).

In the in-depth study we significantly increased the number of calculated turnover rates. We selected brain and liver for this more in-depth analysis because they represent tissues with fast and slow half-lives. In brain tissue we detected 10,375 proteins, of which 6,020 half-lives could be calculated and 5,098 passed our residual error cut-off (less than 0.15, Figure 1d). In liver tissue we identified a total of 8,693 proteins. Of these, half-lives were calculated for 5,768 and 5,322 had a residual error less than 0.15. In total over 7,424 unique protein half-lives were measured between brain and liver while 2,946 were detected in both (Supplementary Table 3).

Experimental Reproducibility Replicate experiments were conducted in mice fed different isotopologues of heavy lysine. Mice were fed a diet of either heavy lysine with six C^{13} and two N^{15} atoms or eight H^2 atoms to verify half-life measurements. Figure 2 demonstrates that protein half-lives measured in all nine tissues have a Pearson correlation greater

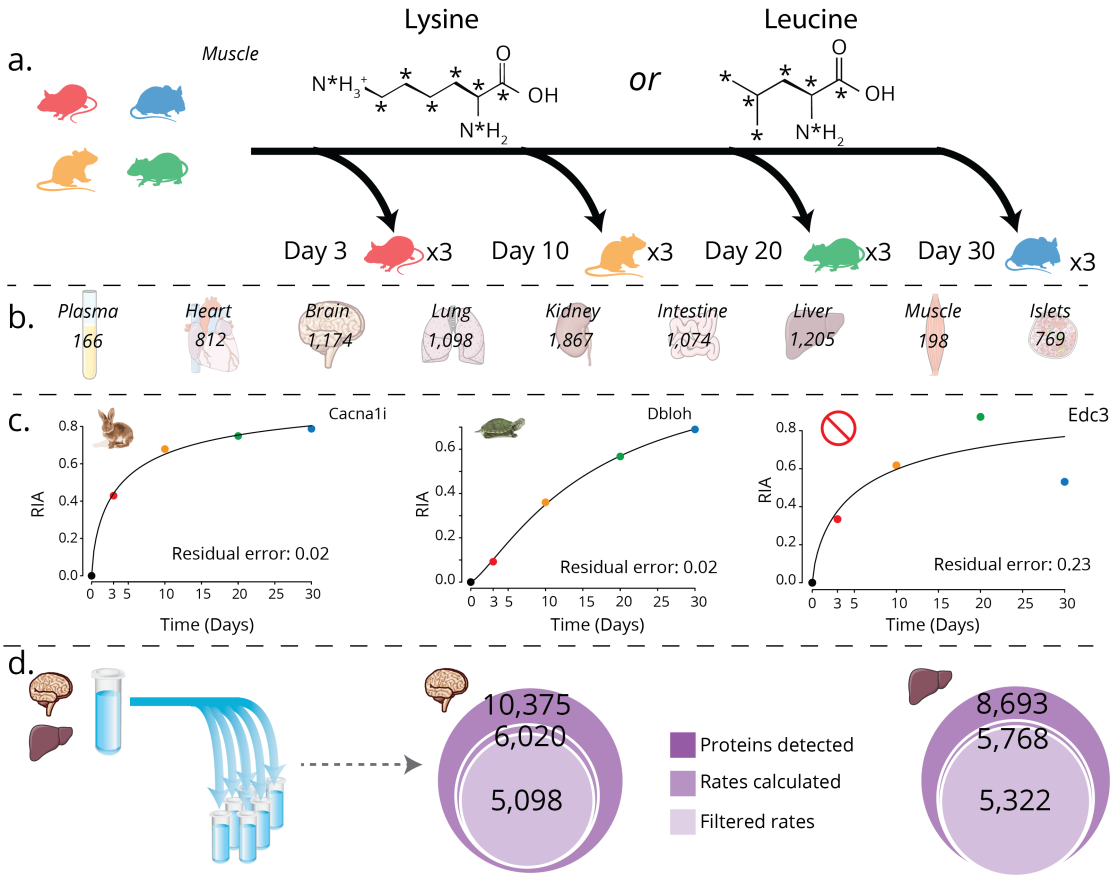


Figure 5.1: Protein turnover is calculated in nine tissues from mice using a diet including heavy lysine or leucine. (A) Three mice were fed a diet of chow with either lysine or leucine replaced with an amino acid containing heavy isotopes. Mice tissues were collected after either 3, 10, 20, or 30 days on the diet. (B) Nine tissues were collected from mice: plasma, heart, brain, lung, kidney, intestine, liver, muscle, and islets. (C) Turnover rates were calculated using the drc R package. A fit was included in the dataset if the residual error was less than 0.15. Proteins with relatively fast and slow turnover are shown for demonstration. (D) Brain and liver tissues were then fractionated into 20 fractions and analyzed using nLC-MS/MS, which measured half-lives in 5,098 and 5,322 proteins in brain and liver, respectively.

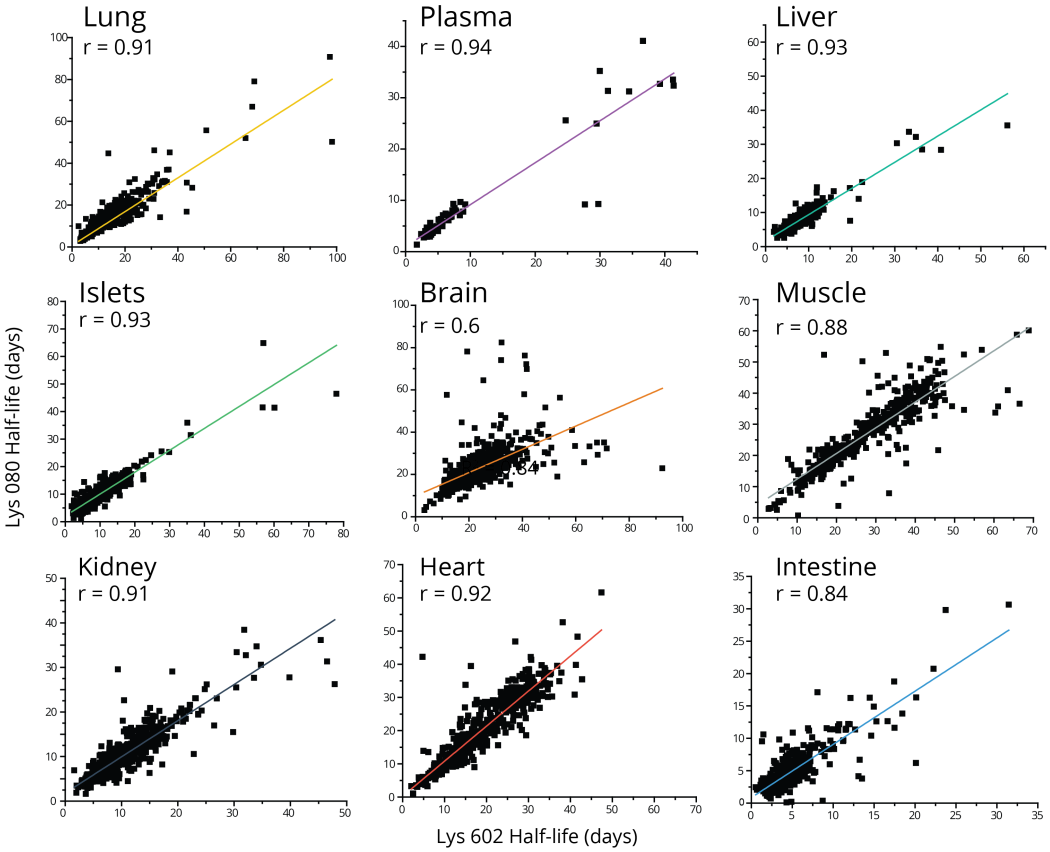


Figure 5.2: Protein half-lives calculated in mice that were fed lysine with six C^{13} and two N^{15} (602) show strong agreement with half-lives calculated from mice fed a diet of lysine containing eight H^2 (080). Pearson correlations were calculated between the two datasets and all nine tissues were found to be above 0.6.

than 0.6 between isotopologue experiments and our method delivers reproducible results.

To comment on reproducibility across experiments, we compared our results to previously published studies of protein turnover. Due in part to the large differences in cell division rate, we expect that our study of mammalian tissues will show minimal similarity with studies conducted in cell culture. Indeed, protein half-lives measured in mammalian tissues are more representative of human protein half-lives than those measured in yeast. In a recent comparable study performed by Hammond and colleagues, bank voles were fed a diet of heavy lysine for a SILAM experiment similar to ours[28]. Our data agree well with Hammond and colleagues' calculations of protein half-life in bank vole muscle, kidney, and heart—tissues we also analyzed (Figure 3); Pearson correlations for the protein half-lives range from 0.4-0.6. Additionally we observe good agreement of our half-life calculations in brain and liver with Price et al. and Fornasiero et al. which were both conducted in mouse[44,45]. Our calculations are most similar to Fornasiero et al. who used a similar experimental of mice fed a heavy amino acid, whereas Price et al. fed a diet of N^{15} labeled algae suggesting experimental method may contribute to technical differences.

Given the suitability of this comparison, we can also begin to consider the questions Hammond et al. pose regarding evolutionary conservation and the allometric scaling of half-life, as well as the scaling of biological processes with body size. Although further studies in other species are needed to confirm it, the correlation we observe may indicate that protein turnover is an evolutionarily conserved trait.

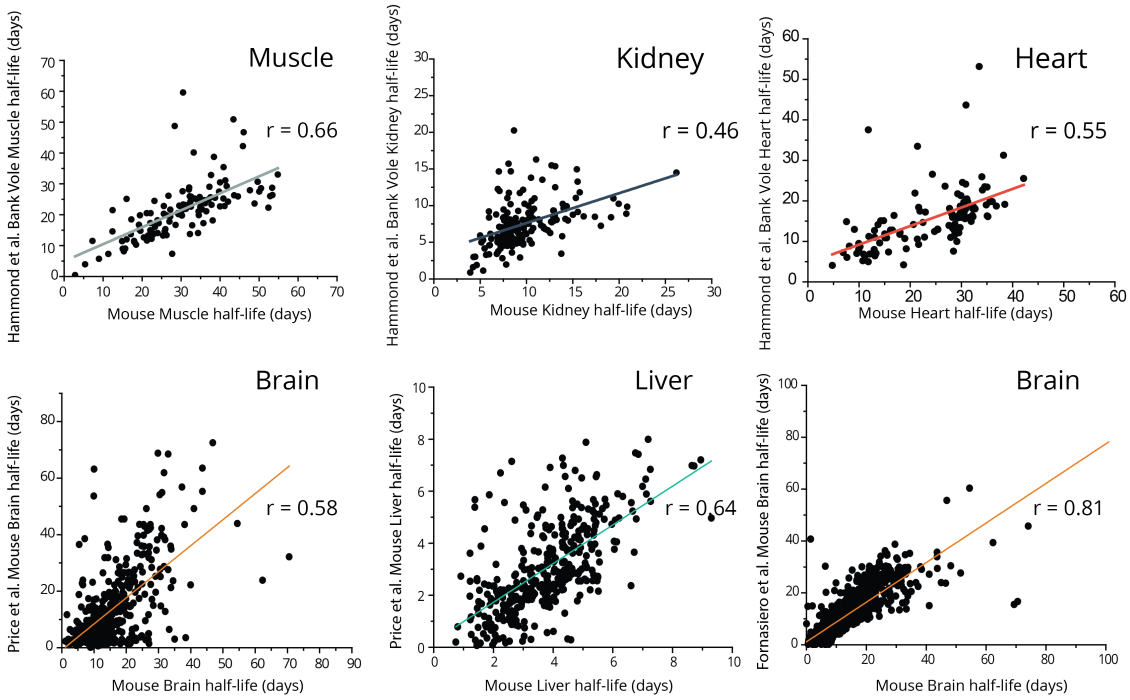


Figure 5.3: Protein half-lives measured in muscle, kidney, brain, liver, and heart showed good agreement with half-lives in mouse and bank vole measured by Hammond et al, Price et al., and Fornasiero et al. Pearson correlations are shown between the five tissues measured in three studies. Bank voles and mice analyzed by Hammond et al. and Fornasiero et al. were fed a diet of heavy lysine using a SILAC method similar to this study. The mice analyzed by Price et al. were fed a diet of N^{15} labeled spirulina.

Tissue-Dependent Turnover From our survey study, i.e., nine tissues, we observe a wide variation of median protein half-lives, ranging from four to >20 days. Our results confirm known trends: for example, brain and muscle protein half-lives are significantly longer than average protein half-lives across tissues, while liver half-lives are shorter[45,46]. In Figure 4a, we classify global turnover trends in additional tissues, noting that proteins from intestine, plasma, and islets have relatively short half-lives, while kidney and lung proteins have an intermediate turnover rate.

Leveraging our analysis to track changes in protein half-life across tissues, we measured 130 protein half-lives in eight out of nine tissues. (Plasma was excluded due to its poor overlap with other tissues). When tissues were sorted in order of increasing half-life, the median slope shows an increase of 2.3 days per tissue (Figure 4b). This increase indicates that protein turnover rate is highly dependent on the tissue where it is localized, and allows us to conclude that the turnover rate of most proteins depends more on the tissue of origin than the protein itself. That said, some proteins do display more consistent half-lives across tissues. These proteins may require differentially regulated half-lives across tissues for their function.

Determinants of Protein Half-life Using our in-depth brain and liver dataset, we set out to interrogate what properties of proteins might explain variations in protein half-life. A correlation between protein abundance and half-life has previously been reported in human cells; however, other studies conducted in yeast found no correlation between abundance

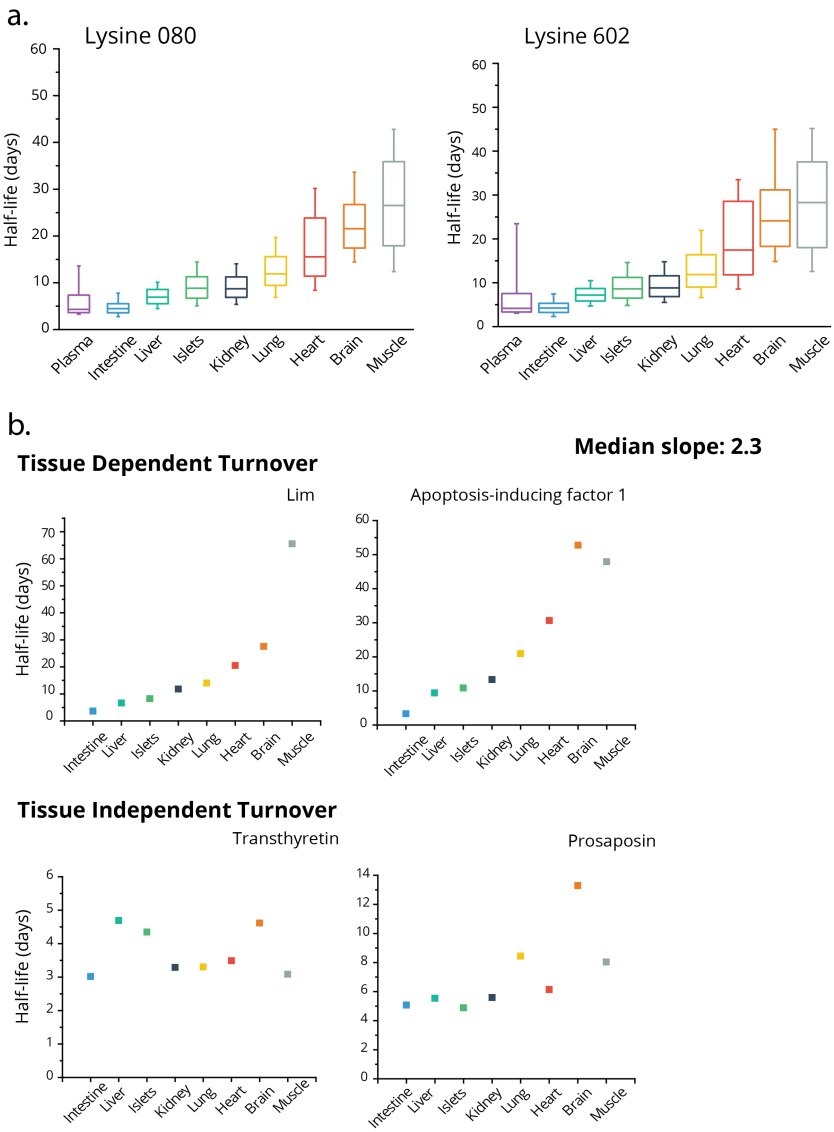


Figure 5.4: Protein turnover is highly dependent on tissue of origin. Box plot of distributions of protein turnover for each tissue. Line = median, Box = inner 50% , whisker range 10-90% , outliers are not shown.The median half-life for proteins in different tissues ranges from four days in intestines to 26 days in muscle. (A) While most proteins display tissue-dependent turnover as shown in the first three examples (B), others show a consistent half-life across many tissues (C).

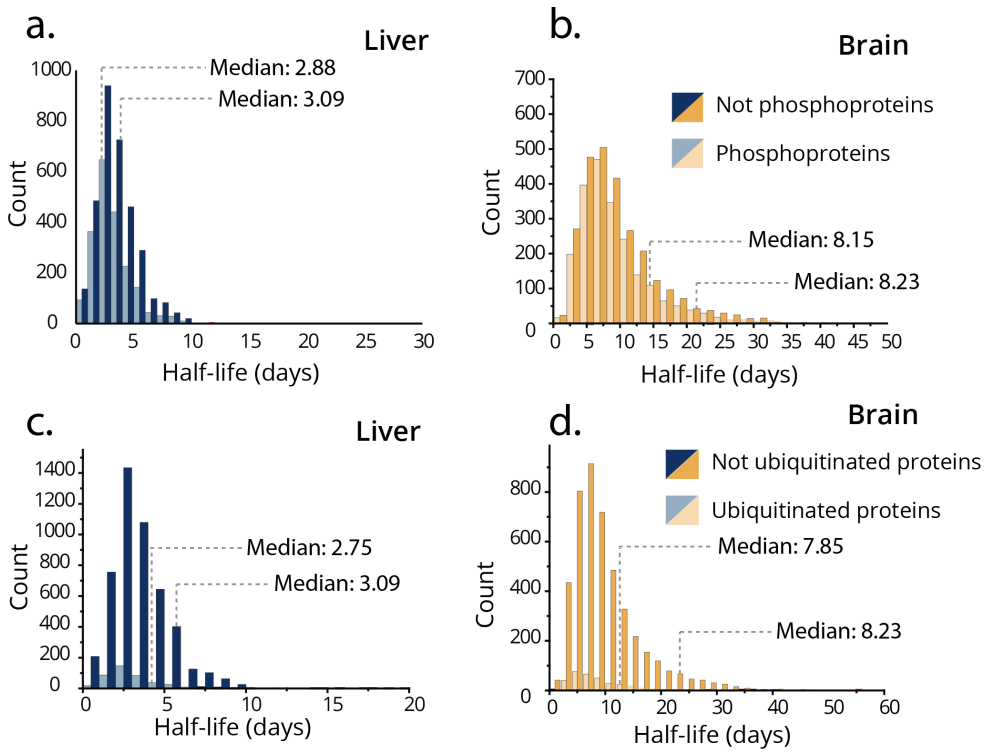
and turnover[21,26]. Figure 5b shows a significant correlation between protein abundance and half-life, with more abundant proteins having a longer half-life.

Altogether, our mouse protein turnover atlas includes half-lives ranging from 0.0018 to 380 days; however, as previously noted we are only able to confidently report half-lives ranging from 3-20 days. Further analysis of large scale turnover rates with a more complete set of time points would assist in validating this range of half-lives. We analyzed the functional and biological properties of proteins possessing different turnover rates to identify potential explanations for half-life variation. We investigated sub-cellular localization, post-translational modifications, sequence motifs at the N and C-termini, and functional properties of proteins. We found no significant correlation between aliphatic index, protein length, molecular weight, positively or negatively charged amino acids, and protein half-life using linear regression and ANOVA testing. Extinction coefficient, isoelectric point, instability index, and GRAVY score along with the following amino acids were all found to correlate with protein half-life: alanine, cysteine, glycine, methionine, proline, and arginine. Spearman correlation coefficients are reported alongside values calculated by Christiano and Martin-Perez et al (Supplementary Figure 1)[26,47].We also observed a trend in which proteins known to be ubiquitinated and phosphorylated as annotated in Uniprot have shorter half-lives in brain and liver; however, these differences were not statistically significant. (Supplementary Figure 2).

With the hypothesis that proteins in different organelles may possess varying half-lives, we also investigated the effect of sub-cellular localization on protein turnover rate. A few

	Spearman Correlation	Martin-Perez et al.	Christiano et al.
Alphatic Index	-0.05	**0.11**	**0.18**
Extinction Coefficient	**-0.14**		
GRAVY score	**0.12**	**0.15**	**0.17**
Instability Index	**-0.18**	**-0.09**	**-0.17**
Isoelectric point	**0.07**	**-0.07**	**-0.11**
Molecular Weight	-0.19	-0.04	-0.04
Protein length	-0.18	-0.04	-0.04
Carbon	-0.18	**0.20**	**0.10**
Hydrogen	-0.18	-0.05	**0.07**
Nitrogen	-0.18	**-0.15**	**-0.14**
Oxygen	-0.19	-0.03	**-0.07**
Sulfur	-0.15	-0.03	-0.04
Ala	**-0.14**	0.02	**0.13**
Arg	**-0.17**	**-0.13**	-0.05
Asn	-0.18	0.00	**-0.15**
Asp	-0.18	-0.01	0.05
Cys	**-0.14**	-0.01	0.01
Gln	-0.19	-0.04	-0.04
Glu	-0.19	**-0.08**	**0.07**
Gly	**-0.14**	0.04	**0.12**
His	-0.15	0.03	-0.03
Ile	-0.12	**0.07**	**0.09**
Leu	-0.17	**0.10**	**0.07**
Lys	-0.16	**-0.11**	0.00
Met	-0.13	-0.02	**-0.07**
Phe	-0.16	**0.12**	0.06
Pro	**-0.20**	**0.08**	-0.04
Ser	-0.20	-0.01	**-0.23**
Thr	-0.17	0.04	-0.06
Trp	**-0.13**	**0.09**	0.01
Tyr	-0.13	**0.11**	0.10
Val	**-0.14**	0.02	**0.20**

Supplementary Figure S5.1: Spearman correlation between physcial properties and protein half-life. Spearman correlations between protein half-life and physical properties are reported for this study and Christiano et al. and Martin-Perez et al. Correlations that are statistically significant ($p<0.05$) are shown in bold.

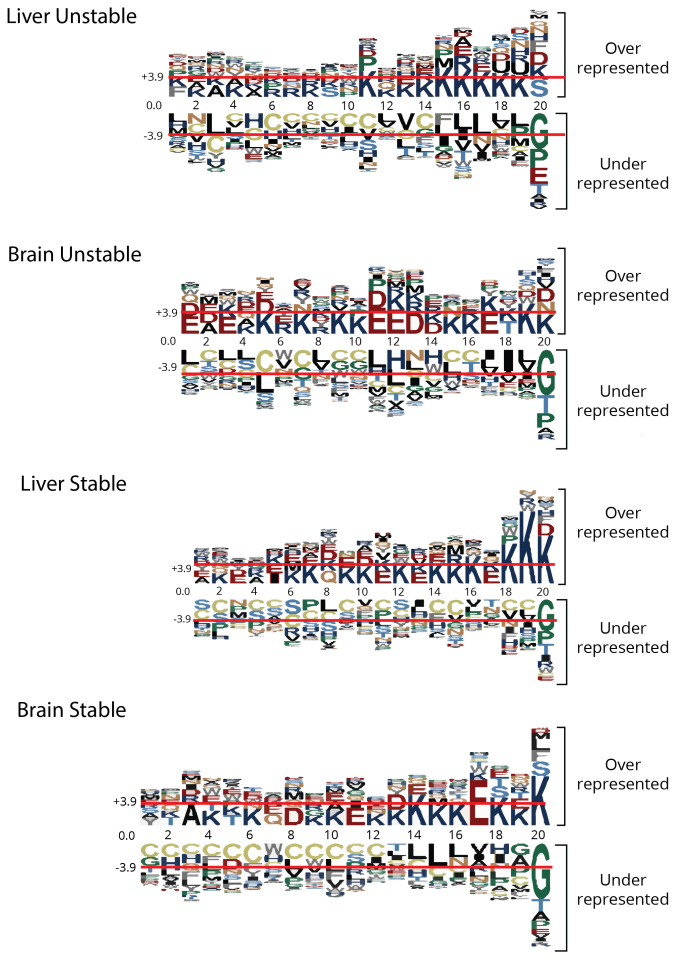


Supplementary Figure S5.2: No significant differences between half-lives of phosphorylated and ubiquitinated proteins with un-modified proteins were observed. Modified proteins were identified from Uniprot and plotted in a histogram. In brain and liver ubiquitinated and phosphorylated proteins had half-lives that trended shorter than un-modified proteins.

differences between organelles stood out: nuclear and peroxisomal proteins displayed some of the shortest half-lives in both brain and liver, while mitochondrial proteins trended toward longer half-lives (Figure 5c). Overall, however, we find that half-lives in proteins localized to different organelles are relatively consistent.

To examine the function of proteins with various turnover rates, we divided proteins of our in-depth dataset at the first and third quartile by half-life, labeling the slowest 25% of proteins “stable" and the fastest 25% “unstable." Using these groups, we performed motif analysis at the N- and C-termini of stable and unstable proteins to identify conserved sequences indicative of protein half-life using pLOGO[48]. Differences in the N-terminal sequence did not explain half-life variation. An enrichment for lysine at the C-terminus was observed in stable proteins in both liver and brain. Additionally, there is a slight enrichment for proline as the eleventh from last amino acid in unstable proteins, and proline is disfavored at this position in stable proteins (Supplementary Figure 3). Favored in disordered regions of proteins, proline’s position may allow these proteins to be more readily targeted for degradation[49].

Next, we performed GO enrichment using DAVID with the all proteins in our data set as background[50,51]. Statistically significant GO terms were detected in both the stable and unstable cohorts of brain and liver. Stable proteins in brain and liver are enriched for terms such as glycolysis/gluconeogenesis, TCA cycle, carbon metabolism, oxidative phosphorylation, and cytoskeleton. Among unstable proteins, we observed enrichment for pathways including kinases, DNA damage and repair, and helicases, as well as spliceosome,



Supplementary Figure S5.3: Motif analysis of the C-terminus of stable and unstable proteins in brain and liver generated using pLOGO. The proteins with half-lives in the longest 25% were labelled stable and those in the fastest 25% were labeled unstable.

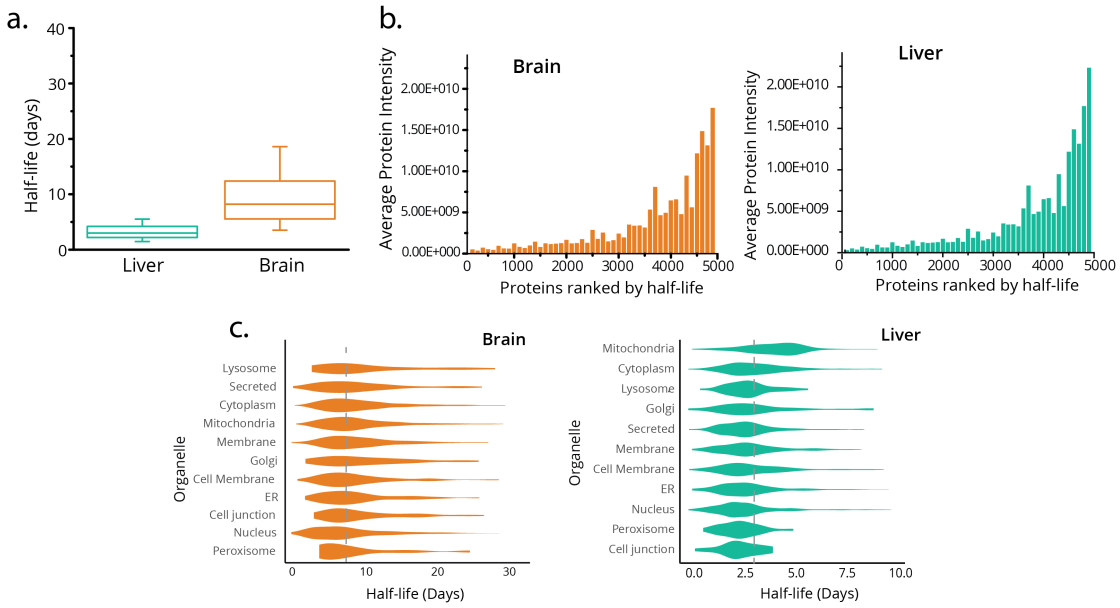


Figure 5.5: Liver and brain have differing rates of protein turnover and some variation can be explained by differences in protein abundance and sub-cellular localization. Box plot of distributions of protein turnover for fractionated liver and brain peptides. Line = median, Box = inner 50% , whisker range 10-90% , outliers are not shown. The median half-life of proteins in brain and liver is three and eight days respectively. (A) Protein abundance or intensity correlates with protein half-life. (B) Proteins localized to different organelles have small differences in half-life (C).

peroxisome, and ubiquitin-mediated proteolysis (Figure 6). None of the significant GO terms reported here for stable proteins were enriched in unstable proteins and *vice versa*.

Experimentally determined half-life values can vary with MS method Given we measured protein half-lives using two different MS methods for some brain and liver proteins (i.e., overlapping proteins measured in the survey study and the in-depth study), we compared median protein half-lives in leucine labeled mice collected from an in depth study and a survey analysis. Figure 7a displays this comparison for overlapping proteins from Leucine labelled livers analyzed with both the survey and in-depth method. To our surprise, our half-life estimate is quite different depending on the MS method. Specifically, measurements from the in-depth study estimated a given protein's half-life to be much shorter than the same measurement in the survey study (Figure 7b). Noting that the brain and liver study employed two-dimensional chromatography rather than the single-shot approach of the survey study, we reasoned that MS^1 dynamic range could be driving these differences. In fact, this is what we observe in many proteins; for example, heavy-labelled h2afy peptide is clearly visible in the in-depth study after three days while it cannot be seen in the survey study (Figure 7c). Pointedly, the total signal of this peptide is two orders of magnitude higher in the in-depth study than the survey study due to the improved dynamic range afforded by the two-dimensional chromatographic method. By improving our ability to detect low abundance peptides, we improve our ability to accurately measure the ratio of heavy peptide to light. Peptide ratios are used to calculate protein half-life

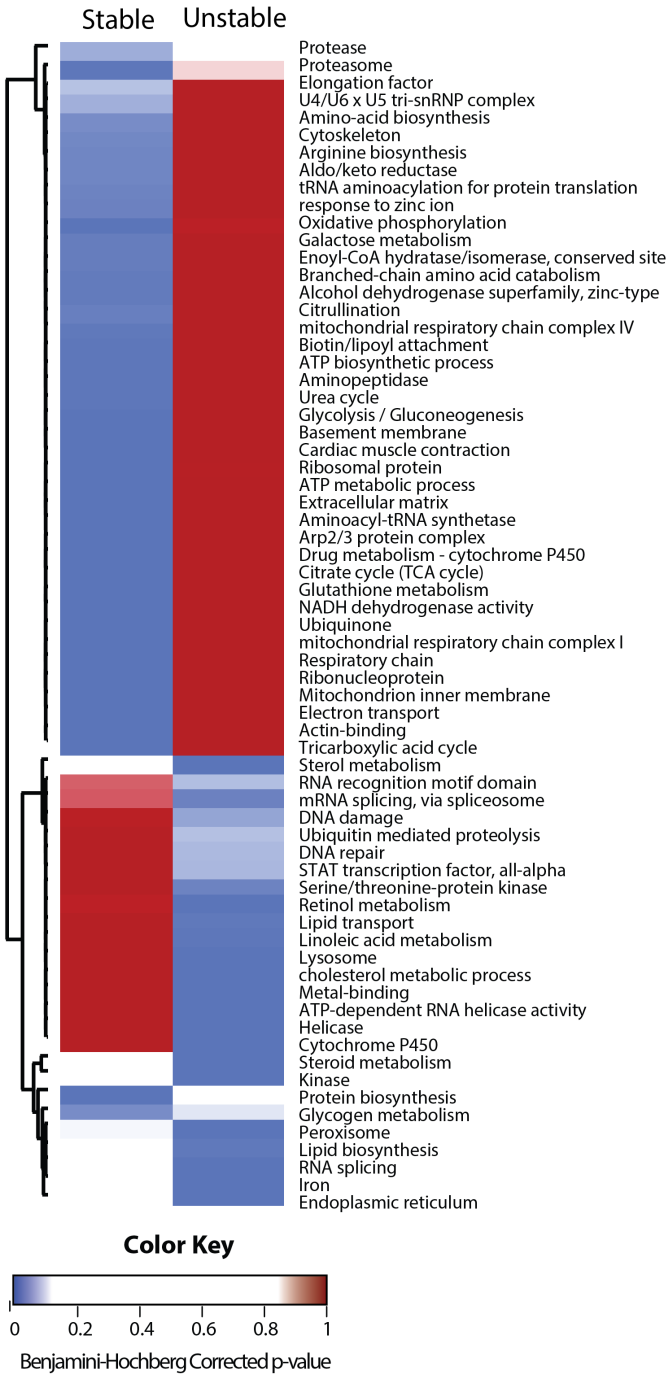


Figure 5.6: Stable and unstable proteins are enriched for distinct GO terms. Stable proteins are defined as those with a half-life above the third quartile; unstable proteins have a half-life below the first quartile. Stable and unstable proteins were enriched for GO terms using DAVID with a background set to the total proteins in our data set. GO terms that were only significant in either stable or unstable proteins are reported here. All reported p-values were corrected using Benjamini-Hochberg correction.

making dynamic range an essential determinant of half-life. (Figure 7d).

Conclusion

We conclude that this variation in dynamic may also contribute to the discrepancies observed in turnover experiments between labs. Any turnover rate calculated using MS methods will be limited by instrument sensitivity and is best interpreted as an approximation and relative rate. Beyond that, experiments having overall low incorporation rates and studies using limited chromatography to counter this will be most affected.

We measured protein half-lives in a whole mouse model spanning nine tissues and 8,149 unique proteins in total. Half-lives are reported from 4-20 days, however rates far outside this range were calculated. Mice were fed a diet containing heavy amino acids and turnover rates were calculated by measuring the ratio of heavy to light proteins over time. The data were fit using an R package to generate protein half-lives in a high throughput fashion. To generate a large atlas of protein half-lives in two diverse tissue types, we fractionated brain and liver peptides. After filtering by residual error, we measured half-lives of over 5,000 proteins in both brain and liver. Performing a replicate experiment in mice fed a diet containing a different isotopologue of heavy lysine allowed us to further validate our data. The strong correlation observed across all the tissues indicates that our method can reproducibly measure turnover rates.

Our data indicate that protein half-lives in different tissues vary widely. Specifically, tissues with a faster metabolic rate, such as liver and intestine, have faster turnover than

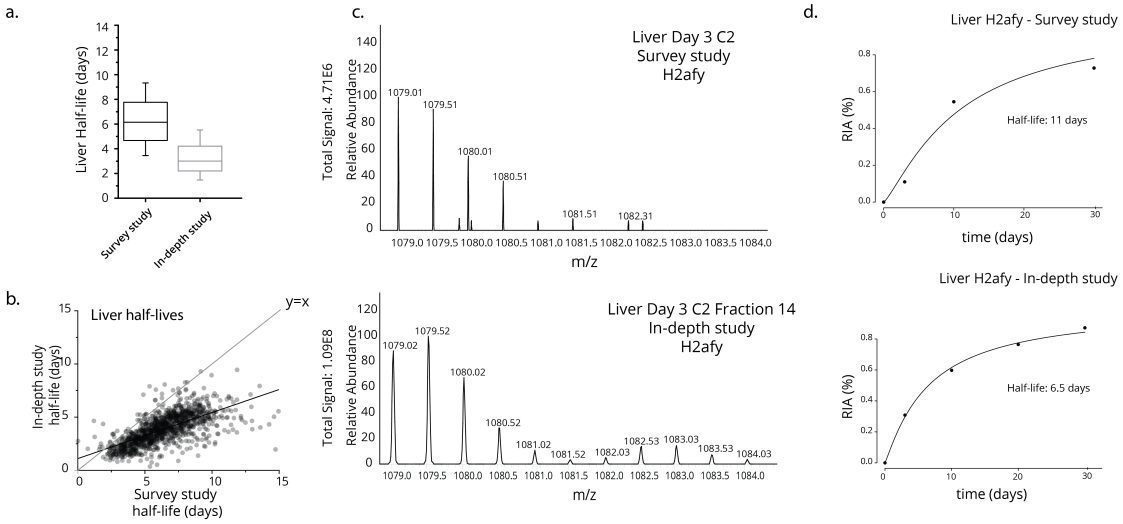


Figure 5.7: Protein fractionation results in the measurement of lower half-lives than single shot analysis. Significantly lower half-lives were measured using a single-shot method compared to fractionation. (A) Box plot of distributions of protein half-lives for liver from survey study and in-depth study. Line = median, Box = inner 50% , whisker range 10-90% , outliers are not shown. Proteins that were measured with both methods appeared to have a longer half-life when using fractionation. (B) For example, in histone protein H2afy we are able to detect heavy labeled peptide at day three in the fractionated in-depth sample, whereas no signal for heavy peptide can be detected in the survey study at day three. (C)The RIA plot for H2afy reflects this result. In the in-depth sample, heavy amino acid has been incorporated in 30% of peptides, but we do not observe any in the single-shot sample (D)

tissues such as muscle, heart, and brain, which have a lower metabolic rate. While turnover rates of the same protein across tissues depends primarily on tissue of origin, some proteins show conserved half-lives across tissues. These proteins may be differentially regulated across tissues resulting in consistent half-lives. We investigated the role of a number of properties which may contribute to turnover rate. Protein abundance, for example, robustly predicted protein half-life, with more abundant proteins having a slower rate of turnover. A fast turnover of abundant proteins would require a high-energy expenditure of the cell, which may explain this common trait among abundant proteins. Other factors such as hydrophobicity, isoelectric point, and amino acid abundance also showed correlation with turnover.

Our study shows that improving the signal-to-noise ratio with peptide fractionation results in the calculation of shorter and likely more accurate protein half-lives. Using two-dimensional chromatography, we detect heavy peptides as they first appear in the protein. However, in our survey experiment, low-abundance heavy peptides are obscured at early time pointes due to noise and spectral complexity. For this reason, we caution studies with limited separations and/or low isotope incorporation rates.

We found stable enriched proteins for GO terms, including oxidative-phosphorylation, mitochondria, citrate cycle, and ATP synthesis. These processes are of relatively constant need to the cell and less likely to be up or down-regulated rapidly in a stable system. Unstable proteins contain terms likely subject to more dynamic regulation, such as steroid metabolism, transcription regulation, mRNA splicing, tyrosine protein kinase, and blood

coagulation. We expect signaling proteins to exist in the cell at lower copy numbers compared to major metabolic pathways, supporting our previous findings regarding protein abundance.

Despite the crucial role protein turnover plays in indicating cellular state, it remains largely unexplored in the multi-omics era. The variation in turnover rates measured in different studies suggest that protein dynamics are likely highly sensitive to changes in cellular state. Our study finds that turnover rates depend on tissue of origin, which results in differences in the turnover rate of a single protein in different locations. Our goal is to provide a large atlas of turnover rates in a whole mouse model. This atlas can serve as a benchmark for research investigating other growth conditions or treatments: more studies are needed to investigate the effect of various stressors or disease states on protein turnover, as it represents an important indicator of metabolic state and a hallmark of aging. Indeed, turnover studies are complementary to other molecular readouts and may in fact prove more sensitive to biological perturbations than protein abundance.

Acknowledgements We thank the University of Wisconsin-Madison and the National Institute of General Medical Sciences R35GM118110 (awarded to JJC) for providing financial support for this work. The authors thank the lab of Alan Attie for their contributions and Alex Hebert and Evgenia Shishkova for thoughtful conversation.

References

[1] I. Dhondt, V. A. Petyuk, S. Bauer, H. M. Brewer, R. D. Smith, G. Depuydt, and B. P. Braeckman, "Changes of protein turnover in aging Caenorhabditis elegans," *Mol Cell*

Proteomics, p. mcp.000049.2017, jul 2017.

[2] H. Koga, S. Kaushik, and A. M. Cuervo, "Protein homeostasis and aging: The importance of exquisite quality control," *Ageing Research Reviews*, vol. 10, pp. 205–215, apr 2011.

[3] C. López-Otín, M. A. Blasco, L. Partridge, M. Serrano, and G. Kroemer, "The hallmarks of aging," *Cell*, vol. 153, pp. 1194–1217, jun 2013.

[4] S. I. Rattan, "Synthesis, modification and turnover of proteins during aging," *Advances in Experimental Medicine and Biology*, vol. 694, pp. 1–13, jan 2010.

[5] A. G. Ryazanov and B. S. Nefsky, "Protein turnover plays a key role in aging," *Mechanisms of Ageing and Development*, vol. 123, pp. 207–213, jan 2002.

[6] A. C. Thompson, M. D. Bruss, J. C. Price, C. F. Khambatta, W. E. Holmes, M. Colangelo, M. Dalidd, L. S. Roberts, C. M. Astle, D. E. Harrison, and M. K. Hellerstein, "Reduced in vivo hepatic proteome replacement rates but not cell proliferation rates predict maximum lifespan extension in mice," *Aging Cell*, vol. 15, pp. 118–127, feb 2016.

[7] W. E. Balch, R. I. Morimoto, A. Dillin, and J. W. Kelly, "Adapting proteostasis for disease intervention," *Science*, vol. 319, pp. 916–919, feb 2008.

[8] S. M. Day, "The ubiquitin proteasome system in human cardiomyopathies and heart failure," *AJP: Heart and Circulatory Physiology*, vol. 304, pp. H1283–H1293, may 2013.

[9] H. Kimmel, "Withdrawal from Repeated Cocaine Alters Dopamine Transporter Protein Turnover in the Rat Striatum," *Journal of Pharmacology and Experimental Therapeutics*, vol. 304, pp. 15–21, jan 2003.

[10] M. J. Kuhar, "On the use of protein turnover and half-lives," *Neuropsychopharmacology*, vol. 34, pp. 1172–1173, apr 2009.

[11] J. A. Norton, T. P. Stein, and M. F. Brennan, "Whole body protein synthesis and turnover in normal man and malnourished patients with and without known cancer," *Ann. Surg.*, vol. 194, pp. 123–128, aug 1981.

[12] J. R. Poortmans, A. Carpentier, L. O. Pereira-Lancha, and A. Lancha, "Protein turnover, amino acid requirements and recommendations for athletes and active populations," *Brazilian Journal of Medical and Biological Research*, vol. 45, pp. 875–890, jun 2012.

[13] S. C. Bull and A. J. Doig, "Properties of protein drug target classes," *PLoS ONE*, vol. 10, mar 2015.

[14] M. J. Kuhar and A. R. Joyce, "Slow onset of CNS drugs: Can changes in protein concentration account for the delay?," *Trends in Pharmacological Sciences*, vol. 22, pp. 450–456, sep 2001.

[15] A. J. Claydon, M. D. Thom, J. L. Hurst, and R. J. Beynon, "Protein turnover: Measurement of proteome dynamics by whole animal metabolic labelling with stable isotope labelled amino acids," *Proteomics*, vol. 12, pp. 1194–1206, apr 2012.

[16] I. V. Hinkson and J. E. Elias, "The dynamic state of protein turnover: It's about time," *Trends in Cell Biology*, vol. 21, pp. 293–303, may 2011.

[17] C. Hughes and J. Krijgsveld, "Developments in quantitative mass spectrometry for the analysis of proteome dynamics," *Trends in Biotechnology*, vol. 30, pp. 668–676, dec 2012.

[18] J. M. Pratt, J. Petty, I. Riba-Garcia, D. H. L. Robertson, S. J. Gaskell, S. G. Oliver, and R. J. Beynon, "Dynamics of Protein Turnover, a Missing Dimension in Proteomics," *Molecular & Cellular Proteomics*, vol. 1, pp. 579–591, aug 2002.

[19] B. H. Toyama and M. W. Hetzer, "Protein homeostasis: Live long, won't prosper," *Nature Reviews Molecular Cell Biology*, vol. 14, pp. 55–61, jan 2013.

[20] J. M. Baughman, C. M. Rose, G. Kolumam, J. D. Webster, E. M. Wilkerson, A. E. Merrill, T. W. Rhoads, R. Noubade, P. Katavolos, J. Lesch, D. S. Stapleton, M. E. Rabaglia, K. L. Schueler, R. Asuncion, M. Domeyer, J. Zavala-Solorio, M. Reich, J. DeVoss, M. P. Keller, A. D. Attie, A. S. Hebert, M. S. Westphall, J. J. Coon, D. S. Kirkpatrick, and A. Dey, "NeuCode Proteomics Reveals Bap1 Regulation of Metabolism," *Cell Reports*, vol. 16, no. 2, pp. 583–595, 2016.

[21] M. Scott, Y. Ahmad, D. Lamont, A. I. Lamond, F. Charrière, G. Barton, F.-M. Boisvert, and M. Gierliński, "A Quantitative Spatial Proteomics Analysis of Proteome Turnover in Human Cells," *Molecular & Cellular Proteomics*, vol. 11, p. M111.011429, mar 2011.

[22] S. B. Cambridge, F. Gnad, C. Nguyen, J. L. Bermejo, M. Krüger, and M. Mann, "Systems-wide proteomic analysis in mammalian cells reveals conserved, functional protein turnover," *Journal of Proteome Research*, vol. 10, pp. 5275–5284, dec 2011.

[23] S. Rogers, R. Wells, and M. Rechsteiner, "Amino acid sequences common to rapidly degraded proteins: The PEST hypothesis," *Science*, vol. 234, pp. 364–368, oct 1986.

[24] A. O. Helbig, P. Daran-Lapujade, A. J. Van Maris, E. A. De Hulster, D. De Ridder, J. T. Pronk, A. J. Heck, and M. Slijper, "The diversity of protein turnover and abundance under nitrogen-limited steady-state conditions in Saccharomyces cerevisiae," *Molecular BioSystems*, vol. 7, pp. 3316–3326, jan 2011.

[25] M. Martin-Perez and J. Villén, "Feasibility of Protein Turnover Studies in Prototroph Saccharomyces cerevisiae Strains," *Analytical Chemistry*, vol. 87, pp. 4008–4014, apr 2015.

[26] M. Martin-Perez and J. Villén, "Determinants and Regulation of Protein Turnover in Yeast," *Cell Systems*, vol. 5, pp. 283–294.e5, sep 2017.

[27] B. Schwanhüusser, D. Busse, N. Li, G. Dittmar, J. Schuchhardt, J. Wolf, W. Chen, M. Selbach, B. Schwanhäusser, D. Busse, N. Li, G. Dittmar, J. Schuchhardt, J. Wolf, W. Chen, and M. Selbach, "Global quantification of mammalian gene expression control," *Nature*, vol. 473, pp. 337–342, may 2011.

[28] P. Stockley, R. J. Beynon, A. J. Claydon, D. E. Hammond, J. L. Hurst, D. M. Simpson, and D. Edward, "Proteome Dynamics: Tissue Variation in the Kinetics of Proteostasis in Intact Animals," *Molecular & Cellular Proteomics*, vol. 15, pp. 1204–1219, apr 2016.

[29] E. J. Hsieh, N. J. Shulman, D.-F. Dai, E. S. Vincow, P. P. Karunadharma, L. Pallanck, P. S. Rabinovitch, and M. J. MacCoss, "Topograph, a Software Platform for Precursor Enrichment Corrected Global Protein Turnover Measurements," *Molecular & Cellular Proteomics*, vol. 11, pp. 1468–1474, nov 2012.

[30] E. Lau, Q. Cao, D. C. Ng, B. J. Bleakley, T. U. Dincer, B. M. Bot, D. Wang, D. A. Liem, M. P. Lam, J. Ge, and P. Ping, "A large dataset of protein dynamics in the mammalian heart proteome," *Scientific Data*, vol. 3, mar 2016.

[31] J. C. Price, S. Guan, A. Burlingame, S. B. Prusiner, and S. Ghaemmaghami, "Analysis of proteome dynamics in the mouse brain," *Proceedings of the National Academy of Sciences*, vol. 107, pp. 14508–14513, aug 2010.

[32] M. Rahman, S. F. Previs, T. Kasumov, and R. G. Sadygov, "Gaussian Process Modeling of Protein Turnover," *Journal of Proteome Research*, vol. 15, pp. 2115–2122, jul 2016.

[33] B. J. Cargile, J. L. Bundy, A. M. Grunden, and J. L. Stephenson, "Synthesis/Degradation Ratio Mass Spectrometry for Measuring Relative Dynamic Protein Turnover," *Analytical Chemistry*, vol. 76, pp. 86–97, jan 2004.

[34] M. K. Doherty, L. McClean, I. Edwards, H. McCormack, L. McTeir, C. Whitehead, S. J. Gaskell, and R. J. Beynon, "Protein turnover in chicken skeletal muscle: Understanding protein dynamics on a proteome-wide scale," *British Poultry Science*, vol. 45, pp. S27—-S28, apr 2004.

[35] K. T. Fan, A. K. Rendahl, W. P. Chen, D. M. Freund, W. M. Gray, J. D. Cohen, and A. D. Hegeman, "Proteome Scale-Protein Turnover Analysis Using High Resolution Mass Spectrometric Data from Stable-Isotope Labeled Plants," *Journal of Proteome Research*, vol. 15, pp. 851–867, mar 2016.

[36] B. Geary, K. Magee, P. Cash, I. S. Young, P. D. Whitfield, and M. K. Doherty, "Determining synthesis rates of individual proteins in zebrafish (Danio rerio) with low levels of a stable isotope labelled amino acid," *Proteomics*, vol. 16, pp. 1398–1406, may 2016.

[37] C. Ritz, F. Baty, J. C. Streibig, and D. Gerhard, "Dose-response analysis using R," *PLoS ONE*, vol. 10, p. e0146021, dec 2015.

[38] K. A. Overmyer, S. Tyanova, A. S. Hebert, M. S. Westphall, J. Cox, and J. J. Coon, "Multiplexed proteome analysis with neutron-encoded stable isotope labeling in cells and mice," *Nature Protocols*, vol. 13, pp. 293–306, feb 2018.

[39] A. E. Merrill and J. J. Coon, "Quantifying proteomes and their post-translational modifications by stable isotope label-based mass spectrometry," *Current Opinion in Chemical Biology*, vol. 17, pp. 779–786, oct 2013.

[40] J. J. Coon, A. S. Hebert, D. J. Bailey, A. Ulbrich, M. S. Westphall, E. E. Coughlin, and A. L. Richards, "The One Hour Yeast Proteome," *Molecular & Cellular Proteomics*, vol. 13, pp. 339–347, jan 2013.

[41] A. L. Richards, A. S. Hebert, A. Ulbrich, D. J. Bailey, E. E. Coughlin, M. S. Westphall, and J. J. Coon, "One-hour proteome analysis in yeast," *Nat. Protocols*, vol. 10, pp. 701–714, may 2015.

[42] R. J. Beynon, "The dynamics of the proteome: Strategies for measuring protein turnover on a proteome-wide scale," *Briefings in Functional Genomics and Proteomics*, vol. 3, pp. 382–390, feb 2005.

[43] L. Dayon and M. Kussmann, "Proteomics of human plasma: A criticalcomparison of analytical workflows in terms ofeffort, throughput and outcome," *EuPA Open Proteomics*, vol. 1, pp. 8–16, jan 2013.

[44] E. F. Fornasiero, S. Mandad, H. Wildhagen, M. Alevra, B. Rammner, S. Keihani, F. Opazo, I. Urban, T. Ischebeck, M. S. Sakib, M. K. Fard, K. Kirli, T. P. Centeno, R. O. Vidal, R. U. Rahman, E. Benito, A. Fischer, S. Dennerlein, P. Rehling, I. Feussner, S. Bonn, M. Simons, H. Urlaub, and S. O. Rizzoli, "Precisely measured protein lifetimes in the mouse brain reveal differences across tissues and subcellular fractions," *Nature Communications*, vol. 9, p. 4230, oct 2018.

[45] J. C. Price, S. Guan, A. Burlingame, S. B. Prusiner, and S. Ghaemmaghami, "Analysis of proteome dynamics in the mouse brain," *Proceedings of the National Academy of Sciences*, vol. 107, pp. 14508–14513, aug 2010.

[46] A. J. Claydon, M. D. Thom, J. L. Hurst, and R. J. Beynon, "Protein turnover: Measurement of proteome dynamics by whole animal metabolic labelling with stable isotope labelled amino acids," *PROTEOMICS*, vol. 12, pp. 1194–1206, apr 2012.

[47] R. Christiano, N. Nagaraj, F. Fröhlich, and T. C. Walther, "Global Proteome Turnover Analyses of the Yeasts S.cerevisiae and S.pombe," *Cell Reports*, vol. 9, pp. 1959–1966, dec 2014.

[48] J. P. O'Shea, M. F. Chou, S. A. Quader, J. K. Ryan, G. M. Church, and D. Schwartz, "PLogo: A probabilistic approach to visualizing sequence motifs," *Nature Methods*, vol. 10, pp. 1211–1212, dec 2013.

[49] V. N. Uversky, "The alphabet of intrinsic disorder," *Intrinsically Disordered Proteins*, vol. 1, p. e24684, apr 2013.

[50] D. W. Huang, B. T. Sherman, and R. A. Lempicki, "Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists," *Nucleic Acids Research*, vol. 37, pp. 1–13, jan 2009.

[51] D. W. Huang, B. T. Sherman, and R. A. Lempicki, "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources," *Nature Protocols*, vol. 4, pp. 44–57, jan 2009.

Chapter 6

CONCLUSIONS AND FUTURE DIRECTIONS

Project Summary The recent surge of technological advances in mass spectrometry has enabled new biological studies that were recently unfeasible. These advances have allowed once challenging experiments to become routine, and whole proteomes of simple organisms can now be sequenced in several hours. In this thesis, mass spectrometry is applied to diverse biological problems. Each requires unique optimization in terms of sample preparation, instrumentation, computation, and analysis.

First, red blood cells were analyzed from a cohort of monozygotic and dizygotic twins using proteomics and metabolomics. Maintaining blood efficacy during storage is critical to the many life-saving blood transfusion procedures that are performed each year. A high-throughput label free method was employed to quantify proteins in 18 twin pairs. One challenge is the large dynamic range of red blood cells with over 90% of the protein composed of hemoglobin. To circumvent this, cells were lysed and enriched for the membrane component where most low abundance proteins of interest reside. The results were used to calculate protein concentration heritabilities. Blood quality degradation during storage is highly variable, and identifying heritable markers raises the possibility of cataloging blood donors for longer or shorter storage based on their phenotype.

Next, yeast samples from the model organism *Saccharomyces cerevisiae* were measured over a time course in a zinc deficient environment. This experiment also utilized a label free quantitation method along with software to calculate copy number of proteins identified.

One goal of this study was to identify which (if any) proteins exist as apo-proteins in a zinc deficient environment. Since zinc atoms are often bound by cysteine residues in zinc fingers, differential cysteine labeling was used to identify changes in metalation status. Whole cells were first treated with the cell permeable reagent NEM which reacts with solvent accessible cysteines. Cysteines that are binding zinc are unable to react with NEM while apo-proteins will become labeled. The cells are then washed and lysed prior to reduction and alkylation with CAA. Differential labeling between zinc replete and zinc deficient cells can indicate which proteins are binding zinc. This experiment led to the identification of Fba1 and Met6 as apo-proteins during zinc deficiency which was validated as described in chapter 3.

Our focus was next turned to a plant based model system comprising the legume *Medicago truncatula* and the bacterium *Sinorhizobium meliloti*. While the ubiquitin ligase, PUB1, has been identified as essential in initiating the symbiotic relationship formed between these organisms, no Pub1 substrates have yet been identified. To study this low abundance modification, enrichment methods and large quantities of protein were required. This method was combined with TMT quantification to compare ubiquitination level. As a result, a group of lipid binding proteins were identified as putative PUB1 substrates and this result was further supported by lipidomic analyses.

Finally, protein turnover rate measurements were tackled in the context of a range of mouse tissues. Initially, a data set was collected by Baughman et al. (2016, Cell Reports) to assess whether the Neucode quantitative labels highlighted in the manuscript incorporate at similar rates. The rates were found to be indistinguishable between isotopologues in

nine mouse tissues that were collected from 3-30 days on the Neucode diet. It was later realized that such a dataset would potentially allow for the calculation of protein turnover rates. Further data was then acquired from the same stored mouse tissues. While re-purposing animals for multiple experiments is always a noble cause, had this experiment been designed to measure protein turnover the design would have been quite different. It is recommended that fewer mice be collected at each time point, and additional time points be collected starting at 0.5 days and ending at 40-50 days for a total of seven to nine time points. Additionally, while protein-half lives are an intuitive and commonly reported result, it is more correct to perform analyses with the resulting rate constants which prevents compression of data.

The results described in chapter 5 were collected from the available mouse tissues and show good agreement with previously published studies in both the calculated half-lives, and the conclusions that are derived from them. We also observed that experimental method can have a large effect on the rates measured. Turnover calculations directly rely on the ratio between a 'light' and heavy' labeled peptide, and thus the instrument sensitivity and signal to noise are imperative in detecting low abundance peptides. Peptide fractionation results in decreased spectral complexity and improved signal to noise. As a result, half-lives measured using fractionation are shorter and we believe more accurate.

For these reasons this dataset could be a valuable resource and is believed to be the largest compendium of protein turnover rates collected to date. However, its weaknesses in design limit the range of rates that can be confidently calculated and casts doubt on the

conclusions that are drawn from them.

Experimental Challenges The collection of such large quantities of data necessitates computational approaches to data analysis. Beyond the identification and quantitation of proteins, proteomics enables other analyses discussed here such as calculation of protein concentration heritabilities, and turnover rates. Similar calculations have previously been done on proteins of interest; however, proteomics facilitates expansion to a global scale. These large scale calculations, combined with other phenotype data, help create the most complete picture of an organism's biology.

Addressing questions posed by biologists will always require adaptation and optimization of sample preparation strategies to fit the organism or sample being analyzed. Different cell types may be difficult to lyse, offer a limited sample quantity, or contain a large dynamic range of proteins. Examples of many of these conditions were seen in the chapters contained here.

Additionally, depending on the number of samples to be analyzed, the method may need to be further optimized for use with high throughput analysis. Methods that may be feasible when processing a dozen samples are not attainable when analyzing hundreds. Large scale studies that require more than a week of instrument time behoove further consideration in maintaining consistent and quality instrument performance. Drifts in retention time or mass accuracy will cause difficulties in comparing samples in an experiment that ranges over many months. Instrument performance must be carefully monitored in these experiments

with calibrations and quality control methods to ensure reproducible data are collected. Despite these efforts, post-acquisition data normalization is typically necessary to compare data collected over an extended time period. Data are often normalized to a pooled wild type, or time zero control that is analyzed across the experiment and prepped alongside each batch of samples to control for technical variability. Drift between batches can be quite large and entirely confound meaningful biological changes if not corrected appropriately.

Future Directions As proteomic analyses become more routine, large scale analyses of hundreds of samples that currently requires a heroic effort will likely become increasingly automated, especially in the area of sample preparation. Currently this transition is underway with single celled organisms and samples that can be prepared in cell culture. While these techniques cannot manage more complicated experimental design or tissues, many large experiments could be streamlined with their implementation, increasing reproducibility and conserving experimenter's time. Mammalian and other heterogeneous tissues pose additional challenges for lyseing on a large scale which will need to be addressed. Additionally, many of the other techniques discussed here such as enrichment for PTMs are currently not easily scaled to large numbers of samples.

Since the dawn of proteomics there has been a continual push to maximize protein identifications. For single celled organisms the sequencing of an essentially complete proteome was achieved several years ago, and the same feat is close to being achieved in mammalian cell lines. However, solely because the sequencing of a complete proteome is possible,

does not mean that every experiment requires this depth of analysis. Many experimental conditions result in dramatic changes in phenotype which can easily be cataloged with fewer protein identifications. Accepting a lower threshold of protein identifications can allow for significantly faster experimental analysis and would facilitate these large scale analyses that are increasing in popularity.

COLOPHON

This document was typesetted with LaTeX 2$_\varepsilon$ using the MiKTeX project. It is based on the University of Wisconsin dissertation template created by William C. Benton (available at `https://github.com/willb/wi-thesis-template`).