Improvements to the Identification and Quantification of Peptides and Proteins

By

Christopher M. Rose

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

Doctor of Philosophy

(Chemistry)

At the

UNIVERSITY OF WISCONSIN-MADISON

2014

Date of final oral examination: 8/25/2014

This dissertation is approved by the following members of the Final Oral Committee:
Joshua J. Coon, Professor, Chemistry and Biomolecular Chemistry
Lloyd M. Smith, Professor, Chemistry
Michael R. Sussman, Professor, Biochemistry
Lingjun Li, Professor, Pharmaceutical Sciences and Chemistry
David J. Pagliarini, Assistant Professor, Biochemistry

© Copyright by Christopher M. Rose 2014

All Rights Reserved

IMPROVEMENTS TO THE IDENTIFICATION AND QUANTIFICATION OF PEPTIDES AND PROTEINS

Christopher M. Rose

Under the supervision of Professor Joshua J. Coon

At the University of Wisconsin-Madison

The following chapters contain a plethora of proteomic techniques aimed at the improvement of methods for the identification and quantification of peptide and protein species. The genesis of mass spectrometry for the analysis of biological molecules and the advances that have resulted in the large-scale identification and quantification of proteins are detailed in Chapter 1. Chapter 2 discusses the application of quantitative proteomic techniques for the large-scale analysis of a biological system, *Medicago truncatula*. The third chapter discusses the implementation of an online algorithm to identify peptides in real time (*InSeq*), enabling alterations to the instrument method that increase the quantitative accuracy of measurements or aid in post translational modification localization. Chapters 4 to 6 outline advances to the implementation of electron transfer dissociation aimed at increasing the ability to identify protein (Chapter 4) or peptide (Chapters 5 and 6) species. The remainder of this thesis details the application of neutron encoded (NeuCode) mass labels to peptide identification without tandem MS (Chapter 7), top-down quantification of intact proteoforms (Chapter 8), labeling of mammals that enables multiplexed quantitative analysis of mammalian tissue after only 10 days of labeling time (Chapter 9), and targeted quantification of more than 1,000 peptides (Chapter 10). Conclusions and future directions relating to the content of this thesis are discussed in Chapter 11.

Acknowledgements

I would like to thank Professor Coon for his guidance and mentorship over the past four years. Upon my first visit to Wisconsin his passion for his research and his drive to accomplish research goals was evident and throughout my graduate career he has done nothing short of live up to those first impressions. In addition to providing a productive work atmosphere, he has provided the Coon group with an abundance of state-of-the-art facilities and instrumentation. Beyond lab space and instruments, Professor Coon has provided ample opportunities for me personally, specifically speaking at the American Society for Mass Spectrometry (ASMS) conference and the plant phosphorylation conference in Keystone, CO, as well as visiting Thermo Fisher Scientific in San Jose, CA on two separate occasions. These opportunities, in addition to the numerous dinners and events he has arranged, have helped build a network of contacts that will be invaluable as I continue throughout my career. In these last four years, perhaps the greatest lesson I have learned from Professor Coon is the importance of thinking big – identifying needs of the entire proteomic community and trying to meet those with innovative techniques. As researchers it is often too easy to get hyper-focused on an experiment and lose sight of the goals and impact it will make in the field. As I continue on I will always remember that the beginning of my career is rooted in the Coon group.

Graduate school has been a humbling experience, full of many high and low points. I have been fortunate enough to experience more highs than lows, but my productivity in graduate school would not have been possible with the help of the current and former Coon group members. Specifically, I would like to thank those students older than I that were particularly vital in my development. When I first arrived in the Coon group I was assigned to work with Doug Phanstiel and his guidance and passion for science were truly inspiring. His patience and willingness to answer my many questions truly helped bring me up to speed quickly. Throughout my first two years, Craig Wenger was crucial in helping me understand the many complex proteomics concepts, such as protein grouping, and his wit and wisdom were always a welcome break from the sometime monotonous research work. When I made my sortie into instrumentation work, Aaron Ledvina was vital in teaching me about instrumentation and how to manipulate the instrument

code, without his guidance many of the projects detailed here would not have been possible. Justin Brumbaugh, who helped me appreciate the importance of asking the right biological question. Jason Russell was the consummate mentor and his patience and knowledge made for a combination that was truly refreshing if I ever found myself facing problems in the lab. Throughout these past four years no member of the group has provided me more help than Alex Hebert. Within days of starting an early project I found myself staring at an HPLC that I had let run dry, with my sample still on the column no less; sufficiently freaked out I found the closest person to ask for help and since that day Alex has been a constant that is always there when I have a question. In addition to his troubleshooting capabilities he has provided an outlet for discussion whenever I find a paper I am excited or disappointed about. His passion for science and his ability to think of outside of the box ideas is one aspect of the Coon group I will miss the most. Derek Bailey has been a close friend and lab-mate throughout my graduate career and his guidance and help teaching me about computer programing was vital to the completion of many of the projects contained in this thesis. Specifically, his enthusiasm for making programs more efficient enabled the timely analysis of hundreds of raw files making projects like *Medicago* phosphoproteomics (Chapter 2) and NeuCode Mouse (Chapter 9) possible. I would also like to thank those members of my Coon group cohort Anna Merrill and Catie Minogue for sharing this journey through graduate school. Lastly, I would like to thank Mike Westphall who has been a constant source of help, advice, knowledge, and entertainment. Mike helped guide me on my very first task in graduate school and even though I had no idea what I was doing his patience and willingness to explain instrumentation was vital to my development. Throughout these past four years he has remained a resource that I constantly go to for advice, explanations, and laughs. Perhaps one of the aspects of the Coon group I will miss most is the time spent troubleshooting instruments with Mike, as those situation provided invaluable learning experiences that can only be experienced.

I have had the opportunity to work and publish with amazing collaborators throughout my graduate career including Michael Sussman, Lloyd Smith, Jean Michele Ané, Dave Pagliarini, Alan Attie, John Syka, Jae Schwartz, Don Kirkpatrick, James Thomson, Raymond Deshaies, and Ben Garcia among others. I

would like to thank all of them as well as the members of their labs who I worked with directly on those projects. I would also like to thank Joshua Coon, Lloyd Smith, Michael Sussman, Lingjun Li, and Dave Pagliarini for their service on my committee. I feel fortunate to have had meaningful interactions with all of the members of my committee, specifically Lingjun Li whose kindness and willingness to reach out to an undergraduate from California was the reason I first decided to consider graduate school at Wisconsin. I would also like to thank the NSF for a graduate research fellowship and the NIH for the Chemistry Biology Training program fellowship.

I would also like to thank those vital to my undergraduate education and research experience. My undergraduate research advisor at Santa Clara University, Dr. Steven Suljak, was vital to my initial growth as a research scientist. His advising style enabled me to perform research in a mostly independent fashion. His guidance helped point me in the right direction, but he would allow me to make the mistakes that are necessary when first being introduced into research. Outside of the lab Dr. Suljak was instrumental to my career development by enabling me to travel to three professional conferences (PittCon and Neuroscience) and prompting me to apply to for an NSF graduate research fellowship that enabled me to start research full time upon entering graduate school. Even after leaving Santa Clara, Dr. Suljak has continued to be a close friend and given me the opportunity to travel back to my old school to give a lecture about my research. In addition to Dr. Suljak, I would like to thank Dr. John Birmingham who helped with the design and publication of my first research article, and Dr. Tracy Ruscetti whose molecular biology lab was the point at which I put away notions of becoming a medical doctor and decided I wanted pursue graduate research.

While all of the people above have helped make me a better researcher and man, they were only given the opportunity to shape me because of the constant support of my family. Throughout my life my parents have continued to make sacrifices that have given me opportunities some kids can only dream of and for that I will be forever grateful. They have enabled me to follow my dreams which seem to consistently take me further from them, and despite this I know I always have their loving support. I would like to also thank my sister and my grandparents for their help in making me who I am today. Lastly, I

would like to thank my fiancé Ashley Cummock, whose love and support has been a constant force driving me through graduate school. When we started this journey four years ago she made a leap of faith moving to Wisconsin with me. She has enabled me to consume myself with my research by being there to help take care of duties we should probably share more equally. She is also my best friend and the one who has helped me to look away from work, even if only for a while, before I get burned out. Her constant sense of joy helped ensure that I made it through these four years as a happy, healthy, and productive individual. As I continue on my career I am comforted by the fact that we will be on the journey together.

Table of Contents

Table of Contents (continued)

Table of Contents (continued)

List of Figures

List of Figures (continued)

List of Figures (continued)

List of Figures (continued)

List of Figures (continued)

List of Abbreviations and Acronyms

ACN	Acetonitrile
AGC	Automatic gain control
AQUA	Absolute quantification
BCA	Bicinchoninic acid (assay)
CAD	Collisionally activated dissociation
CD	Control diet
CI	Chemical ionization
CID	Collision-induced dissociation
COMPASS	Coon OMSSA proteomic analysis software suite
Da	Dalton, unit of mass measurement (1 Da = 1 g/mol)
DMSO	Dimethyl sulfoxide
DNA	Dexoyribonucleic acid
DTT	Dithiothreitol
EDTA	Ethylenediaminetetracetic acid
ESI	Electrospray ionization
ETD	Electron transfer dissociation
E-value	Expectation value
FDR	False discovery rate
FT	Fourier transform
FTICR	Fourier transform ion cyclotron resonance
FWOM	Full width at 1% maximum
FWTM	Full width at 10% mximum

List of Abbreviations and Acronyms (continued)

GO	Gene ontology
HCD	Higher-energy collisional dissociation
HPLC	High-performance liquid chromatography
i.d.	Inner diameter
InSeq	Instant sequence assignment
IT	Ion trap
iTRAQ	Isobaric tag for relative and absolute quantification
JAM	Joint accumulation multiplexing
LC	Liquid chromatography
Lys-C	Endoproteinase lysine C
m/z	mass-to-charge ratio
MALDI	Matrix-assisted laser desorption ionization
mDa	milli-Dalton
MRM	multiple reaction monitoring
mRNA	messenger RNA
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
MS1	Precursor mass analysis
MS2	Tandem mass spectrometry
mTh	milli-Thomson
NeuCode	Neutron encoding
nHPLC	Nanoflow high performance liquid chromatography

List of Abbreviations and Acronyms (continued)

OMSSA	Open mass spectrometry search algorithm
OT	Orbitrap
ppm	parts per million
PRM	Parallel reaction monitoring
PSM	Peptide-spectrum match
PTM	Post-translational modification
PTR	Proton transfer reaction
Q	Quadrupole
qIT	Quadrupole linear ion trap, synonymous with QLT and LTQ
R	Resolving power
RNA	Ribonucleic acid
RP	Reversed phase
RT	Retention time
RTF	Real time filtering
S/N	Signal-to-noise ratio
SCX	Strong cation exchange (chromatography)
SILAC	Stable isotope labeling by amino acids in cell culture
TCA	Tricarboxylic acid cycle
TFA	Trifluoroacetic acid
Th	Thomson (unit of m/z ratio)
TIC	Total ion chromatogram
TMT	Tandem mass tag

List of Abbreviations and Acronyms (continued)

WT	Wild type
XIC	Extracted ion chromatogram
z	charge

Chapter 1

Introduction

Portions of this chapter have been published:

Brumbaugh J., **Rose C. M.**, Phanstiel D. H., Thomson J. A., Coon J. J. "Proteomics and Pluripotency" *Critical Reviews in Biochemistry and Molecular Biology*, **2011**, 1-14.

Background

Within a single life time, genomic sciences research has progressed from identifying the molecule that contains the instructions for life (DNA),[1] to reading and documenting those instructions, to writing our own set of instructions to create a semi-synthetic organism.[2,3] Along the way it became apparent that while a genome contains the instruction for life, a living organism comprises many diverse biological molecules including nucleic acids (e.g., DNA and RNA), amino-acid polymers (e.g., proteins), and small molecules (e.g., hormones) among others. Further muddying the picture, many of these biological molecules exist in many forms with altered functions; for example proteins are often modified with phosphate groups that might activate or deactivate their function.[4] Even DNA, often thought of as static, exists in a dynamic state that is capable of being turned "on or off" through methylation events and histone modification.[5-7] The overwhelming complexity of biological systems necessitates orthogonal approaches that, when combined, yield biological insight and further our understanding of life itself. These approaches, typically referred to as genomics, transcriptomics, and proteomics, are all related through the central dogma of biology.[8]

Central Dogma of Biology

The flow of genetic information is classically diagrammed as starting with DNA which encodes for genes that are transcribed into messenger RNA (mRNA) before finally being translated into amino acid polymers, or proteins (**Figure 1**).[8] The large scale study of DNA, genomics, was realized in 2000 when the first draft of the Human genome was published.[9,10] Since this time the genomes of thousands of organisms have been documented as technology has advanced such that sequencing an entire genome costs ~ $5,000.[11] These technological improvements have been adopted to enable the large-scale study of mRNA transcripts, called transcriptomics, which serve as the intermediate molecule connecting genes coded in DNA and their product proteins (**Figure 1**).[12] The mRNA intermediate is biologically important, enabling one gene to give rise to multiple protein products through alternative splicing, while also being a relatively short

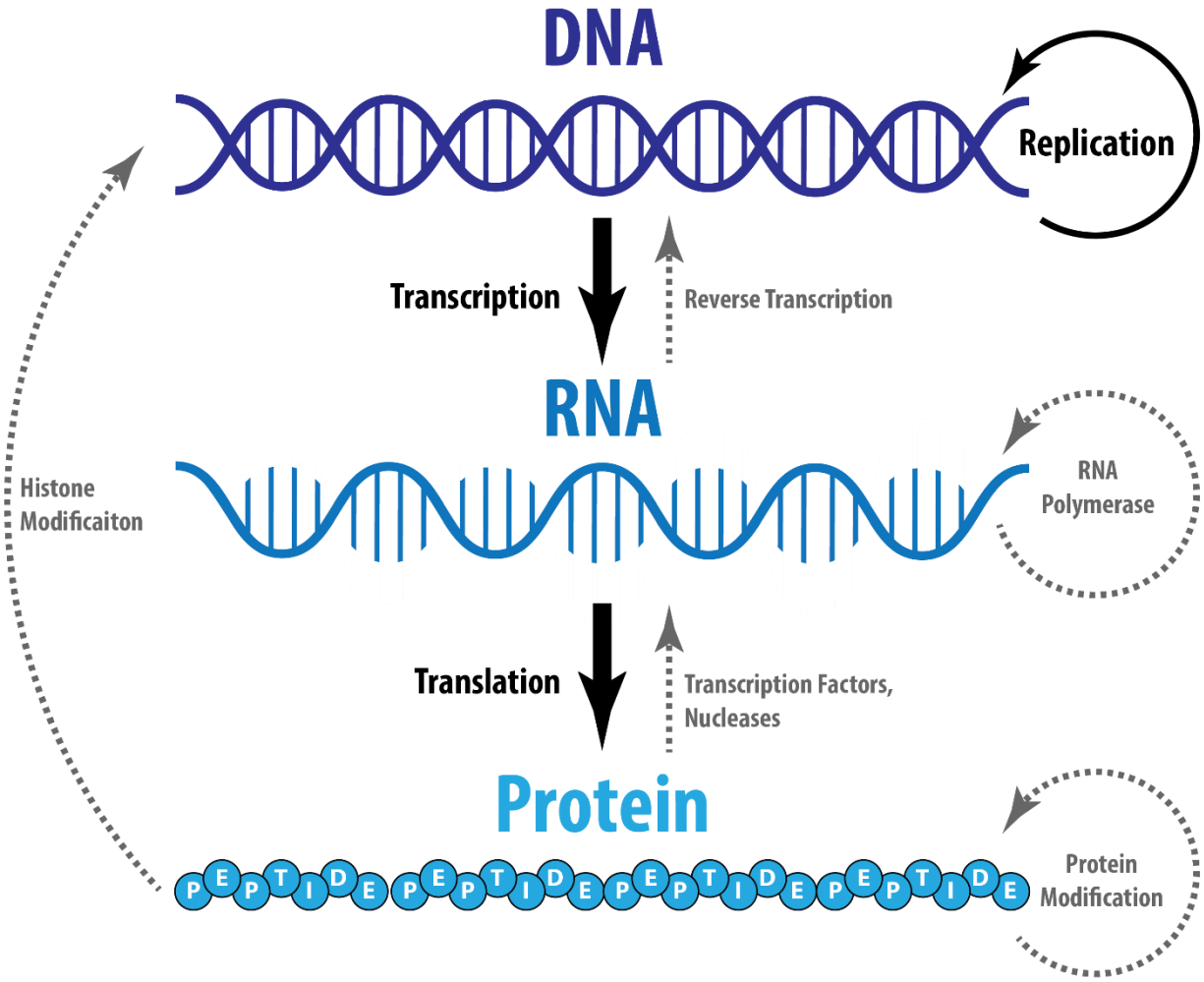


Figure 1. Central dogma of biology

The central dogma of biology describes the flow of genetic information from DNA, to RNA, to protein. Lines and labels in black represent the classical view of the central dogma, while those in grey represent processes that facilitate the flow of information in non-canonical directions.

lived molecule that ensures proteins are only made when needed.[12] Proteins are the product of translating mRNA sequences to amino-acid polymers (**Figure 1**) and act alone or with other proteins within complexes to perform functions within the cell such as promoting gene transcription or catalyzing reactions that add or remove post-translational modifications (PTMs).[4,13] Since the first outline of the central dogma of biology, many new types of interactions have arisen that suggest the true interplay of DNA, RNA, and proteins is more complex with information flowing in both directions. Despite this, measurement of mRNA levels can serve as a proxy for protein expression, especially in steady state systems; however, mRNA expression may not accurately reflect protein abundance if the system undergoing change and does not reveal information regarding protein modification.[14] Within this thesis I will describe improvements to mass spectrometry methods to improve the identification and quantification of proteins, the main actors within the cell.

Mass spectrometry and proteomics

Mass spectrometry (MS) is well suited to proteomic studies because it is highly sensitive, it can be quantitative, and, unlike many antibody-based methods, it does not require *a priori* knowledge of protein targets. The readout from a typical MS experiment is a mass-to-charge ratio, from which a mass is inferred. The primary sequence for a given peptide is determined using tandem mass spectrometry and, because PTMs cause a characteristic mass shift, protein modifications are readily identified and often localized to a single amino acid. Note that these tools are now applicable on the large-scale, owing largely to recent improvements in instrumentation, sample preparation, and software development for downstream analysis.

Large scale analysis of proteins has experienced fantastic growth since its inception, expanding from the identification of roughly 1,500 proteins[15] to more than 15,000 proteins as reported within the initial draft map of the Human proteome.[16,17] As with nucleotide analysis, key advances in technology have driven this meteoric growth.[18] Mass spectrometry analysis necessitates an analyte be transformed into a gaseous

ion, so that it can be manipulated by the electrical fields within the instrument. In the 1980's electrospray (ESI)[19] and matrix-assisted laser desorption ionization (MALDI)[20] enabled researchers to ionize and routinely analyze biological molecules. The direct analysis of proteins, referred to top-down proteomics,[21] is possible, but challenges exist with the preparation, separation, fragmentation, and analysis of the data. Due to these limitations early researchers focused on the large scale analysis of peptides created from the enzymatic digestion of proteins.[22] ESI was a monumental achievement as it provided a straightforward interface between liquid chromatography (LC) columns and mass spectrometers, enabling the direct ionization of peptides after they are separated. Yates and coworkers provided the first LC-MS/MS analysis of a proteome when they used two dimensional strong cation exchange (SCX) and reversed-phase (RP) chromatography coupled to a three dimensional ion trap mass spectrometer.[15] Prior to this experiment Yates and co-workers devised a strategy to utilize computer based peptide search algorithm (SEQUEST),[23] which they used in this large scale analysis to identify 1,500 proteins.

One of the most critical movements in modern proteomics has been the development and rapid advancement of hybrid mass spectrometers. A key, defining feature of these mass spectrometers is the inclusion of two or more mass analyzers, which are the physical components that separate analytes based on mass and charge. Whereas some mass analyzers excel at isolation and fragmentation of peptide precursors (i.e., quadrupole mass filters and ion traps), others offer superior mass accuracy and resolution (i.e., time-of-flight, Fourier transform ion cyclotron resonance, and Orbitrap mass analyzers).[24] Hybrid mass spectrometers combine different types of mass analyzers to harness the unique benefits of each, which has drastically enhanced the ability to detect and quantify proteins and post-translational modifications in complex biological mixtures.[24-27] As opposed to the instrument Yates first used, which operated around 1 scan/sec (Hz), the newest generation mass spectrometer operates close to 20 Hz.[18] The improvements in duty cycle have enabled the recent identification of 4,000 yeast proteins in ~ 1 hour of analysis time[18] and more than 15,000 proteins from multiple human tissue samples. And while the mass spectrometers and LCs have improved, the general proteomics workflow is very similar to when it was first described.

Proteomics Workflow

As stated above, intact proteins are difficult to prepare, separate, and analyze, gave rise to the method termed "bottom-up" or "shotgun" proteomics (**Figure 2**).[22] In a typical bottom-up scheme the cells are lysed releasing proteins, which are chemically unfolded to enable proteases such as trypsin or LysC to digest the protein. These enzymes are typically chosen as they cleave c-terminal to lysine (LysC) or lysine and arginine (trypsin) ensuring that each peptide that is generated will have a basic residue on the C-terminus of the peptide, in addition to the positive free amine group on the N-terminus. Samples are then purified by solid phase extraction, at which point they can be analyzed by LC-MS/MS directly or separated into fractions using chromatography methods such as high-pH reversed-phase or SCX which are orthogonal to the final low-pH reversed phase separation from which peptides are ionized as they elute.

Intact peptide masses are often determined by high-resolution mass analysis which offers mass accuracy routinely below 10 ppm and yields charge state information used to determine the correct monoisotopic mass. Charge state information is vital to limiting the search space when spectra are mapped to the proteome and ensuring that precursors that are selected are peptide species. To determine peptide sequences, many hybrid instruments offer a variety of peptide fragmentation methods –each with benefits and drawbacks. Resonant excitation collision activated dissociation (CAD) is commonly used to dissociate peptides in global protein identification experiments, whereas electron-based dissociation methods (electron transfer dissociation (ETD) or electron capture dissociation (ECD)) are well suited for the characterization of PTMs and intact proteins.[28-31] For example, ETD and ECD have been used to characterize combinations of PTMs on histones in human embryonic stem (ES) and cancer cells. [32,33] More recently, higher energy beam-type CAD led to increased phosphopeptide identifications in both human ES and induced pluripotent stem (iPS) cell types.[34,35] Unlike resonant excitation, HCD is not constrained by low mass cutoff and is therefore compatible with isobaric tag-based quantitation techniques. An increasingly common practice is to use multiple dissociation methods, in combination, to achieve increases

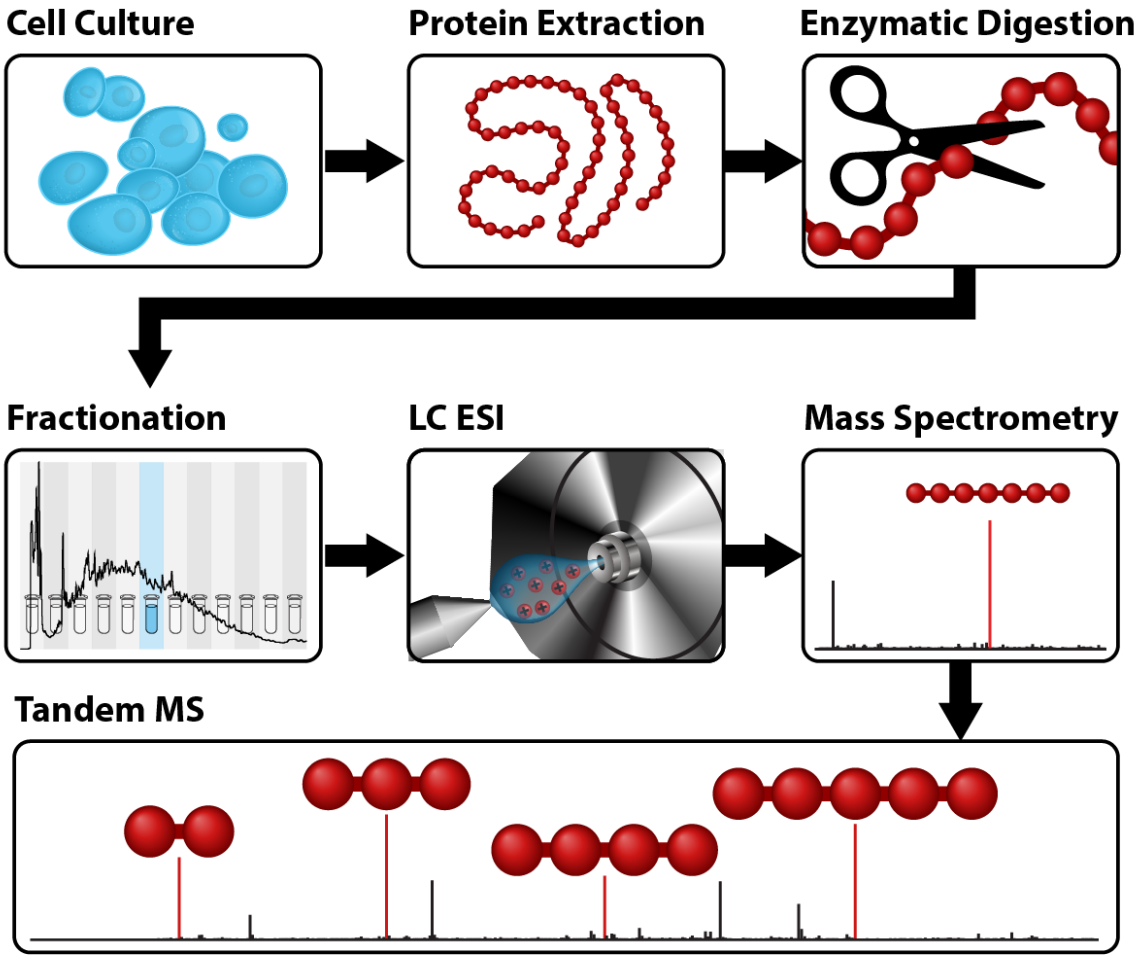


Figure 2. Schematic illustration of proteomic workflow
Proteins are isolated from the sample and peptides are produced *via* enzymatic digestion. Samples may be fractionated to decrease complexity before nano-HPLC-MS/MS analysis. Intact precursors are analyzed using high resolution mass spectrometry and fragmented to produce sequence informative fragment ions.

in identification and/or quantitative accuracy.[36-39] Thus, a number of options are available for tandem MS analysis and fragmentation can be tailored to address particular biological questions of interest.

Once the MS and MS/MS data is collected it is converted to a peaks list and searched against an *in silico* digestion of all annotated proteins for the given organism.[40,41] This workflow underscores the importance of annotated genomes as a basis for predicting protein products. More recently, large-scale analysis of mRNA has been used to make protein databases that are rooted in transcripts that are specific to the system of interest.[42-44] This is an exciting new technique that no longer relies on algorithms that predict alternative splicing events and may miss non-canonical state sequences. To control for false discovery, reverse protein sequences are also made and digested *in silico,* such that both target and decoy peptides are scored and a false discovery threshold can be determined.[45] Once *in silico* peptides are generated, peak lists from each MS/MS scan are matched to the proteome by first narrowing the search space to peptides with the same intact neutral mass. Each peptide on this shortened list is then fragmented *in silico* and the theoretical fragments are matched to the experimental peaks using complex scoring algorithms such as SEQUEST, MASCOT, or OMSSA.[23,46,47] These search engines return peptide sequences that match to a given MS/MS spectra with a confidence score. These results are then typically filtered to 1% FDR using target-decoy approaches.[48] The confidently identified peptides are then combined together as protein groups, such that you can explain all of the peptides with the fewest number of proteins, also at 1% FDR.[49] In addition to identifying proteins many mass spectrometry based methods have been developed to enable large-scale quantification of thousands of peptides within a single set of experiments.

Quantitative proteomic analyses of biological systems

Several mass spectrometry based protein quantitation methods are currently available, providing flexible solutions for a wide range of experimental conditions (**Figure 3**). Label-free methods allow quantitation across many samples and do not require additional steps during sample preparation, the

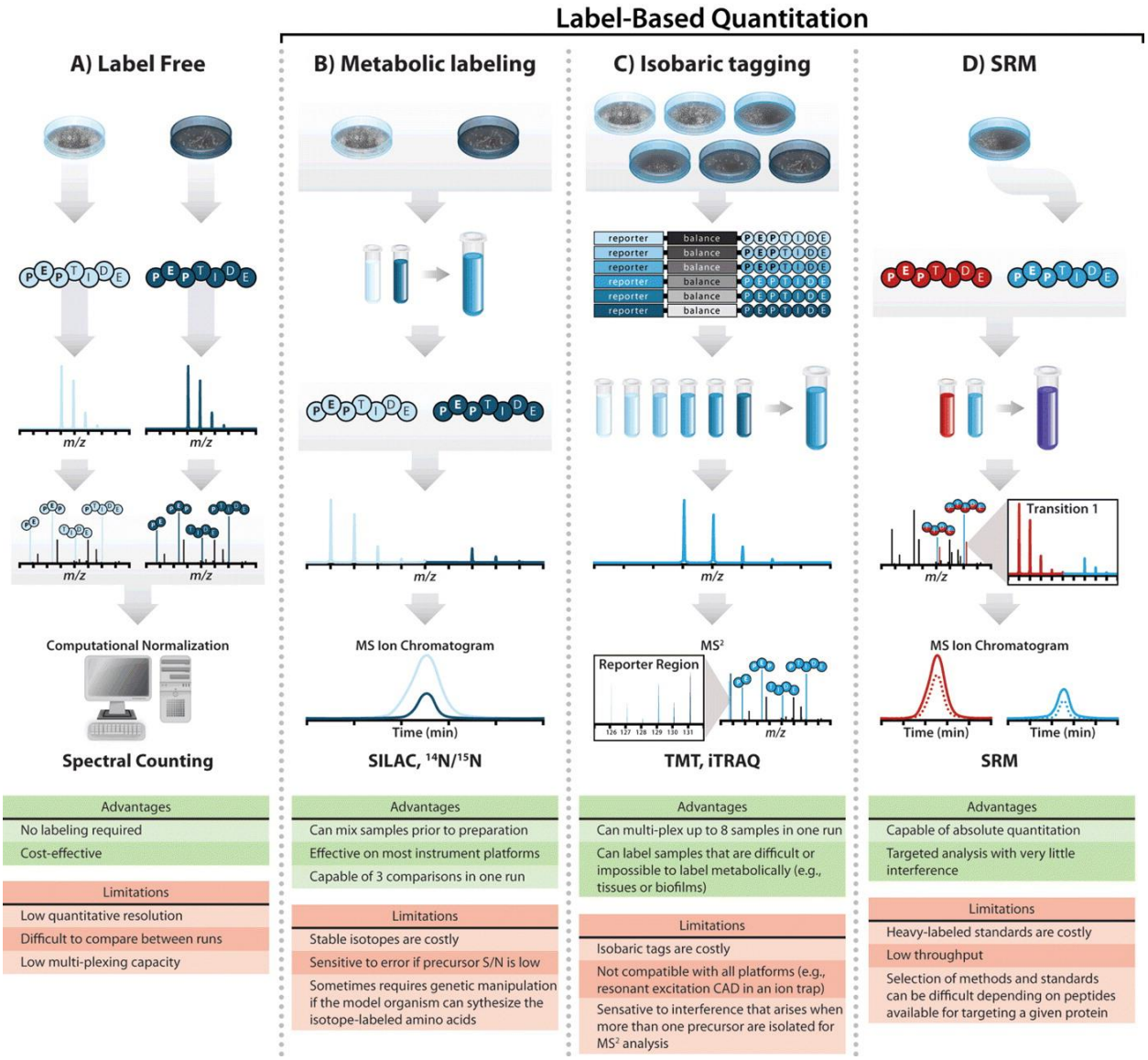


Figure 3. Quantitative proteomics strategies

A) Label-free quantitation is achieved through spectral counting of peptide spectral matches followed by normalization that accounts for protein size among other factors. This technique allows for relative quantitation without the use of isotopes. B) Isobaric labeling strategies, such as TMT and iTRAQ, are capable of comparing up to eight samples in a single run. Labeled peptides from different samples have the same nominal mass and co-elute with reversed-phase chromatography. Peptide fragmentation during tandem mass spectrometry produces both sequence ions for peptide identification and reporter tags for quantitation. C) Metabolic labeling incorporates heavy isotopes into proteins through amino acids or nutrients (i.e., ^{15}N or ^{13}C). These stable isotopes generate a shift in *m/z* values observed in MS^1 scans. The difference between extracted ion chromatograms for peaks corresponding to heavy and light samples is proportional to the relative abundance of each. D) Quantitation *via* single reaction monitoring (SRM) is capable of absolute quantitation and is often performed on triple quadruple mass spectrometers. Figure panels adapted from (Gerber *et al.*, 2007) and (Ross *et al.*, 2004).

samples of interest are merely prepared normally and analyzed. However, care must be taken when comparing results between samples as it may be difficult to replicate chromatography conditions that enable sampling between runs.[50,51] Label free quantification has experienced a renaissance coinciding with an increase in mass spectrometer scan speed that enables the identification of thousands of proteins without fractionation, enabling many replicate nLC-MS/MS experiments to be performed in succession. Label free quantitation (LFQ) algorithms have also matured such that they provide accurate quantitation, even if the sample is fractionated.[52] Generally, LFQ extracts MS^1 chromatograms for the intact peptide species and compares the areas between samples, being careful to normalize for overall sample loading amounts as indicated by the overall signal intensity throughout the run.[52-54] These algorithms can be quite complex and rely on accurately matching precursors between nLC-MS/MS experiments, even if the precursor was not selected for MS/MS analysis, but require high-resolution mass analysis of precursor peptides to limit the interference from co-eluting peptides. Lastly, the re-introduction of data-independent mass spectrometry methods (e.g., SWATH) has gained popularity due to the relative ease of performing such experiments, as well as the promise of collecting data that can be mined in the future.[55-58] It is possible that within the next few years LFQ may become the preferred method of protein quantification for many labs, especially those who do not have the resources to explore stable isotope labeling technologies.

Stable isotope labels, introduced metabolically or through peptide labeling, provide multiplexed relative quantitation in one analysis.[59] Metabolic labeling, in which stable isotopes are introduced via media or diet, can be used to quantify differences between two or three different samples in a single run.[60,61] A key advantage of metabolic labeling is that samples can be mixed prior to preparation, which reduces variation from sample processing and decreases instrument analysis time. Stable isotope labeling by amino acids in cell culture (SILAC) is a well-established form of this technique.[62,63] Typically, a SILAC experiment consists of two cultures with media containing natural lysine ("light culture") or isotopically labeled +8 Da lysine ("heavy culture"). Proteins from each culture are mixed, digested with LysC (ensuring each peptide will contain a lysine), and analyzed in one MS experiment.[14] This results in two MS^1 features

for each peptide, a heavy and light partner, and the MS^1 intensities of these two species is compared to yield relative quantitation. Quantifying from MS^1 spectra results in excellent accuracy, as precursor interference is now a minor issue; however this method suffers from poor quantitation when precursors are present at low signal-to-noise levels. Additionally, MS^1 based quantification has traditionally suffered from limited multiplexing ($\leq$3) due to the increased MS^1 spectral complexity with each additional sample.[64] Recently, we introduced a method to increase the multiplexing capacity of SILAC by using neutron encoded (NeuCode) lysine isotopologues that differ in mass by 6 to 36 mDa.[65-68] Metabolic introduction of NeuCode lysine reagents (NeuCode SILAC) enabled higher levels of multiplexing as compared to traditional SILAC, while providing more accurate measurements with a larger dynamic range when compared to isobaric labeling methods.[64] NeuCode utilizes isotopologues of amino acids that differ in mass by as little as 6 mDa and as much as ~ 40 mDa. These small mass differences are not detectable at mass resolutions lower than ~ 100,000, but are revealed under ultra-high resolution analysis that is achievable on modern FT-MS systems, including the Orbitrap analyzer.[69-75] Reducing partner spacing from 4 - 10 Da to < 40 mDa with NeuCode SILAC compresses the various precursor species in *m/z* space – space so small that only one form of the peptide is detected under typical mass resolution conditions. This approach lowers the number of precursor species in MS^1 spectra, eliminates redundant sampling of peptide partners, and increases overall protein identifications.[76] Metabolic labeling offers excellent accuracy for thousands of proteins; however, chemical isobaric labels offer an alternative when metabolic labeling is not feasible (e.g., human clinical samples).

Chemical labeling strategies such as isobaric tags for relative and absolute quantitation (iTRAQ) and tandem mass tags (TMT) allow for up to ten sample comparisons, further increasing throughput.[77-79] Like metabolic labeling, this approach has been used to quantify thousands of proteins and PTMs; however isobaric labeling is compatible with a number of model systems that cannot be metabolically labeled.[80,81] In a typical isobaric labeling experiment samples are lysed, reduced, alkylated, digested, and purified separately before the chemical label is attached. Isobaric labels are molecules that comprise a reporter,

balance, and linker regions. The linker region contains an NHS-ester molecule that will modify any primary amine group, including the N-terminus and lysine side chains. The reporter region of each isobaric label have unique mass signatures and utilize neutron encoded mass shifts to enable up to 10-plex analysis.[82-84] The balance regions for each unique label are isotopically encoded such that each flavor of isobaric label will have the same intact mass, a crucial step to ensure that the mass shift for each peptide is the same for all samples. The beauty of this technique is displayed upon peptide fragmentation, where the isobaric label is dissociated and the unique reporter ions are revealed enabling quantitation in the MS^2 spectra.

This quantitative technique enables near 100% quantification of the peptides that are identified, but suffers from a challenge known as precursor interference.[85] When comparing two samples an ideal experimental design would ensure that only a select number of key, biologically important proteins change in abundance. Due to this, most measurements made in proteomic analysis will yield a ratio of 1:1, but a small set of important proteins will change. Consider a peptide changing with a ratio of 10:1, this will be peptide A. Isobaric labeled peptides are quantified in MS^2 spectra, and are typically isolated with a width of 1.8 to 3 Th. In this case peptide A will be isolated along with other peptides whose *m/z* falls within the isolation window. These interfering peptides are most likely at a ratio of 1:1 and because reporter ion from these peptides will overlap with those from peptide A, the measured ratio of peptide A will be compressed – most likely to a ratio of 4:1. A clever experimental design can be used to prove this phenomenon. Mixing a sample of yeast at a ratio of 10:5:1:1:5:10 with human peptides at a ratio of 0:0:0:1:1:1 will enable modeling of human interference by analyzing only the yeast peptides. When this sample is analyzed the measured values will be close to 10:5:1:1:2:4, demonstrating that the human peptides mixed at 0:0:0:1:1:1 are interfering with the yeast peptides mixed at the ratio. Many methods were proposed to solve the interference problem. One method used proton transfer reaction to purify the precursor in the gas phase, restoring a compressed 4:1 ratio to a ratio of 9:1.[86] This method, deemed QuantMode, was developed in our lab and used in several biological studies; however, the reduced scan speed and the reliance on PTR reagents implemented on ETD-enabled mass spectrometers led to this method being abandoned. An alternative MS^3

method was introduced, where a peptide was fragmented by CAD and the fragment ions were re-isolated in the ion trap before being sent to the HCD cell for dissociation.[87] This method also improved the dynamic range of measurements made by isobaric labels, but came at a cost of data quality due to the isolation of a single fragment ion. Recently, this MS^3 method has been improved by selecting multiple fragment ions simultaneously, improving the reporter tag signal by 8-fold.[88] This method has been implemented on the newest generation mass spectrometers (Orbitrap Fusion), but also incurs a duty cycle penalty, as the mass spectrometer takes two scans for each precursor.

A different approach altogether, selected reaction monitoring (SRM), targets specific proteins and fragment ions to achieve highly accurate quantitation.[89,90] While the aforementioned tactics are largely used for relative quantitation, SRM can be used to determine absolute protein amount. Still, this targeted approach currently has much lower throughput than other quantitation methods. In a typical SRM experiment isotopically labeled peptide standards are synthesized and mixed with an experimental sample at a known concentration.[91] Great care is taken to design assays which analyze intense fragment ions and produce retention times that do not shift from analysis to analysis. Mass spectrometers such as triple quadrupole mass spectrometers are used to isolate the peptide of interest with the first quadrupole, fragmented in the second quadrupole, followed by selection of the fragment ion with the third quadrupole.[91] The analysis consists of oscillating between the standard peptide and the experimental peptide, a process that reduced the duty cycle and limits the number of peptides that can be analyzed in a given analysis. More recently, high resolution mass spectrometers have been used to perform parallel reaction monitoring (PRM), where peptides are fragmented and all of the fragment ions are analyzed.[92,93] The high resolution mass analysis slows duty cycle, but sensitivity can be regained by simultaneous analysis of all fragment ions. A comparison of PRM to traditional SRM reveals that both methods achieve similar levels of detection in a complex matrix. PRM represents a bright future for the targeted proteomic technologies, especially as high resolution mass spectrometers become more accessible.

Overview of projects

The following chapters contain a bevy of proteomic techniques aimed at the improvement of methods for the identification and quantification of peptide and protein species. **Chapter 2** describes the application of large scale quantitative proteomics utilizing isobaric labels to investigate the initiation of symbiosis between rhizobia and *Medicago truncatula* (published 2012).[80] **Chapter 3** details the implementation of real time peptide identification and its application to improving the quality and quantity of quantitative proteomics measurements utilizing both chemical and metabolic labels (published 2012).[94] **Chapter 4** describes the implementation of a modified multi-purpose dissociation cell (MDC) that enables the reaction of larger precursor ion populations and increases the ability to identify intact proteoforms[95] (published 2013).[96] **Chapter 5** discuss the implementation of activated ion ETD (AI-ETD) within the MDC and demonstrated increased peptide identification as compared to ion-trap ETD (published 2013).[97] **Chapter 6** details the implementation of normalized ETD reaction times that reduce scan time while ensuring that each ETD spectra results in the highest yield of product ion intensity (submitted 2014). **Chapter 7** described the use of NeuCode mass shifts of both lysine and leucine isotopologues in conjunction with GluC digestion enables the identification of peptides within MS/MS analysis (published 2013).[66] **Chapter 8** describes the application of NeuCode labeling to the quantification of whole-protein species (published 2014).[68] **Chapter 9** discusses the application of NeuCode labeling to the metabolic labeling of mammals, specifically mice, and describes how NeuCode isotopologues enable multiplex quantitative analysis after a short labeling period (manuscript in preparation). **Chapter 10** details multiplex targeted PRM analysis utilizing NeuCode reagents and an elution order prediction algorithm that when combined enables the quantitation of over 1,000 peptides in a single targeted analysis (manuscript in preparation). Conclusions and future directions relating to the content of this thesis are discussed in **Chapter 11.**

References

1. Watson JD & Crick FHC. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature*, 1953, 171(4356):737-738.

2. Gibson DG, Glass JI, Lartigue C, Noskov VN, Chuang R-Y, Algire MA, Benders GA, Montague MG, Ma L, Moodie MM, Merryman C, Vashee S, Krishnakumar R, Assad-Garcia N, Andrews-Pfannkoch C, Denisova EA, Young L, Qi Z-Q, Segall-Shapiro TH, Calvey CH, Parmar PP, Hutchison CA, Smith HO, & Venter JC. Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome. *Science*, 2010, 329(5987):52-56.

3. Malyshev DA, Dhami K, Lavergne T, Chen T, Dai N, Foster JM, Correa IR, & Romesberg FE. A semi-synthetic organism with an expanded genetic alphabet. *Nature*, 2014, 509(7500):385-388.

4. Brumbaugh J, Hou Z, Russell JD, Howden SE, Yu P, Ledvina AR, Coon JJ, & Thomson JA. Phosphorylation regulates human OCT4. *Proceedings of the National Academy of Sciences*, 2012, 109(19):7162-7168.

5. Jenuwein T & Allis CD. Translating the Histone Code. *Science*, 2001, 293(5532):1074-1080.

6. Kouzarides T. Chromatin Modifications and Their Function. *Cell*, 2007, 128(4):693-705.

7. Jabbari K & Bernardi G. Cytosine methylation and CpG, TpG (CpA) and TpA frequencies. *Gene*, 2004, 333(0):143-149.

8. Crick F. Central Dogma of Molecular Biology. *Nature*, 1970, 227(5258):561-563.

9. Anonymous. Initial sequencing and analysis of the human genome. *Nature*, 2001, 409(6822):860-921.

10. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng Z, Francesco VD, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan W, Ge W, Gong F, Gu Z, Guan P, Heiman TJ, Higgins ME, Ji R-R, Ke Z, Ketchum KA, Lai Z, Lei Y, Li Z, Li J, Liang Y, Lin X, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue B, Sun J, Wang ZY, Wang A, Wang X, Wang J, Wei M-H, Wides R, Xiao C, Yan C, Yao A, Ye J, Zhan M, Zhang W, Zhang H, Zhao Q, Zheng L, Zhong F, Zhong W, Zhu SC, Zhao S, Gilbert D, Baumhueter S, Spier G, Carter C, Cravchik A, Woodage T, Ali F, An H, Awe A, Baldwin D, Baden H, Barnstead M, Barrow I, Beeson K, Busam D, Carver A, Center A, Cheng ML, Curry L, Danaher S, Davenport L, Desilets R, Dietz S, Dodson K, Doup L, Ferriera S, Garg N, Gluecksmann A, Hart B, Haynes J, Haynes C, Heiner C, Hladun S, Hostin D, Houck J, Howland T, Ibegwam C, Johnson J, Kalush F, Kline L, Koduru S, Love A, Mann F, May D, McCawley S, McIntosh T, McMullen I, Moy M, Moy L, Murphy B, Nelson K, Pfannkoch C, Pratts E, Puri V, Qureshi H, Reardon M, Rodriguez R, Rogers Y-H, Romblad D, Ruhfel B, Scott R, Sitter C, Smallwood M, Stewart E, Strong R, Suh E, Thomas R, Tint NN, Tse S, Vech C, Wang G, Wetter J, Williams S, Williams M, Windsor S, Winn-Deen E, Wolfe K, Zaveri J, Zaveri K, Abril JF, Guigó R, Campbell MJ, Sjolander KV, Karlak B, Kejariwal A, Mi H, Lazareva B, Hatton T, Narechania A, Diemer K, Muruganujan A, Guo N, Sato S,

Bafna V, Istrail S, Lippert R, Schwartz R, Walenz B, Yooseph S, Allen D, Basu A, Baxendale J, Blick L, Caminha M, Carnes-Stine J, Caulk P, Chiang Y-H, Coyne M, Dahlke C, Mays AD, Dombroski M, Donnelly M, Ely D, Esparham S, Fosler C, Gire H, Glanowski S, Glasser K, Glodek A, Gorokhov M, Graham K, Gropman B, Harris M, Heil J, Henderson S, Hoover J, Jennings D, Jordan C, Jordan J, Kasha J, Kagan L, Kraft C, Levitsky A, Lewis M, Liu X, Lopez J, Ma D, Majoros W, McDaniel J, Murphy S, Newman M, Nguyen T, Nguyen N, Nodell M, Pan S, Peck J, Peterson M, Rowe W, Sanders R, Scott J, Simpson M, Smith T, Sprague A, Stockwell T, Turner R, Venter E, Wang M, Wen M, Wu D, Wu M, Xia A, Zandieh A & Zhu X. The Sequence of the Human Genome. *Science*, 2001, 291(5507):1304-1351.

11. Hayden EC. Technology: The $1,000 genome. *Nature*, 2014, 507:294-295.

12. Brenner S, Jacob F, & Meselson M. An Unstable Intermediate Carrying Information from Genes to Ribosomes for Protein Synthesis. *Nature*, 1961, 190(4776):576-581.

13. Hunter T. Protein kinases and phosphatases: The Yin and Yang of protein phosphorylation and signaling. *Cell*, 1995, 80(2):225-236.

14. Lee MV, Topper SE, Hubler SL, Hose J, Wenger CD, Coon JJ, & Gasch AP. A dynamic model of proteome changes reveals new roles for transcript alteration in yeast. *Molecular Systems Biology*, 2011, 7(1):n/a-n/a.

15. Washburn MP, Wolters D, & Yates JR. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotech*, 2001, 19(3):242-247.

16. Wilhelm M, Schlegl J, Hahne H, Gholami AM, Lieberenz M, Savitski MM, Ziegler E, Butzmann L, Gessulat S, Marx H, Mathieson T, Lemeer S, Schnatbaum K, Reimer U, Wenschuh H, Mollenhauer M, Slotta-Huspenina J, Boese J-H, Bantscheff M, Gerstmair A, Faerber F, & Kuster B. Mass-spectrometry-based draft of the human proteome. *Nature*, 2014, 509(7502):582-587.

17. Kim M-S, Pinto SM, Getnet D, Nirujogi RS, Manda SS, Chaerkady R, Madugundu AK, Kelkar DS, Isserlin R, Jain S, Thomas JK, Muthusamy B, Leal-Rojas P, Kumar P, Sahasrabuddhe NA, Balakrishnan L, Advani J, George B, Renuse S, Selvan LDN, Patil AH, Nanjappa V, Radhakrishnan A, Prasad S, Subbannayya T, Raju R, Kumar M, Sreenivasamurthy SK, Marimuthu A, Sathe GJ, Chavan S, Datta KK, Subbannayya Y, Sahu A, Yelamanchi SD, Jayaram S, Rajagopalan P, Sharma J, Murthy KR, Syed N, Goel R, Khan AA, Ahmad S, Dey G, Mudgal K, Chatterjee A, Huang T-C, Zhong J, Wu X, Shaw PG, Freed D, Zahari MS, Mukherjee KK, Shankar S, Mahadevan A, Lam H, Mitchell CJ, Shankar SK, Satishchandra P, Schroeder JT, Sirdeshmukh R, Maitra A, Leach SD, Drake CG, Halushka MK, Prasad TSK, Hruban RH, Kerr CL, Bader GD, Iacobuzio-Donahue CA, Gowda H, & Pandey A. A draft map of the human proteome. *Nature*, 2014, 509(7502):575-581.

18. Hebert AS, Richards AL, Bailey DJ, Ulbrich A, Coughlin EE, Westphall MS, & Coon JJ. The One Hour Yeast Proteome. *Molecular & Cellular Proteomics*, 2014, 13(1):339-347.

19. Fenn J, Mann M, Meng C, Wong S, & Whitehouse C. Electrospray ionization for mass spectrometry of large biomolecules. *Science*, 1989, 246(4926):64-71.

20. Karas M & Hillenkamp F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Analytical Chemistry*, 1988, 60(20):2299-2301.

21. Tran JC, Zamdborg L, Ahlf DR, Lee JE, Catherman AD, Durbin KR, Tipton JD, Vellaichamy A, Kellie JF, Li M, Wu C, Sweet SMM, Early BP, Siuti N, LeDuc RD, Compton PD, Thomas PM, & Kelleher NL. Mapping intact protein isoforms in discovery mode using top-down proteomics. *Nature*, 2011, 480(7376):254-258.

22. Yates JR. The Revolution and Evolution of Shotgun Proteomics for Large-Scale Proteome Analysis. *Journal of the American Chemical Society*, 2013, 135(5):1629-1640.

23. Eng JK, McCormack AL, & Yates Iii JR. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.*, 1994, 5(11):976-989.

24. Glish GL & Burinsky DJ. Hybrid mass spectrometers for tandem mass spectrometry. *J Am Soc Mass Spectrom*, 2008, 19(2):161-172.

25. Shevchenko A, Loboda A, Ens W, & Standing KG. MALDI quadrupole time-of-flight mass spectrometry: a powerful tool for proteomic research. *Anal Chem*, 2000, 72(9):2132-2141.

26. Syka JE, Marto JA, Bai DL, Horning S, Senko MW, Schwartz JC, Ueberheide B, Garcia B, Busby S, Muratore T, Shabanowitz J, & Hunt DF. Novel linear quadrupole ion trap/FT mass spectrometer: performance characterization and use in the comparative analysis of histone H3 post-translational modifications. *J Proteome Res*, 2004, 3(3):621-626.

27. Makarov A, Denisov E, Kholomeev A, Balschun W, Lange O, Strupat K, & Horning S. Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. *Anal Chem*, 2006, 78(7):2113-2120.

28. Syka JE, Coon JJ, Schroeder MJ, Shabanowitz J, & Hunt DF. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci U S A*, 2004, 101(26):9528-9533.

29. Zubarev RA, Horn DM, Fridriksson EK, Kelleher NL, Kruger NA, Lewis MA, Carpenter BK, & McLafferty FW. Electron capture dissociation for structural characterization of multiply charged protein cations. *Anal Chem*, 2000, 72(3):563-573.

30. Coon JJ, Syka JE, Shabanowitz J, & Hunt DF. Tandem mass spectrometry for peptide and protein sequence analysis. *Biotechniques*, 2005, 38(4):519, 521, 523.

31. Coon JJ, Syka JEP, Schwartz JC, Shabanowitz J, & Hunt DF. Anion dependence in the partitioning between proton and electron transfer in ion/ion reactions. *International Journal of Mass Spectrometry*, 2004, 236(1–3):33-42.

32. Phanstiel D, Brumbaugh J, Berggren WT, Conard K, Feng X, Levenstein ME, McAlister GC, Thomson JA, & Coon JJ. Mass spectrometry identifies and quantifies 74 unique histone H4 isoforms in differentiating human embryonic stem cells. *Proc Natl Acad Sci U S A*, 2008, 105(11):4093-4098.

33. Garcia BA, Pesavento JJ, Mizzen CA, & Kelleher NL. Pervasive combinatorial modification of histone H3 in human cells. *Nature Methods*, 2007, 4(6):487-489.

34. Olsen JV, Macek B, Lange O, Makarov A, Horning S, & Mann M. Higher-energy C-trap dissociation for peptide modification analysis. *Nature Methods*, 2007, 4(9):709-712.

35. Phanstiel DH, Brumbaugh J, Wenger CD, Tian S, Probasco MD, Bailey DJ, Swaney DL, Tervo MA, Bolin JM, Ruotti V, Stewart R, Thomson JA, & Coon JJ. Proteomic and phosphoproteomic comparison of human ES and iPS cells. *Nature Methods*, 2011.

36. Swaney DL, McAlister GC, & Coon JJ. Decision tree-driven tandem mass spectrometry for shotgun proteomics. *Nature Methods*, 2008, 5(11):959-964.

37. Guo T, Gan CS, Zhang H, Zhu Y, Kon OL, & Sze SK. Hybridization of pulsed-Q dissociation and collision-activated dissociation in linear ion trap mass spectrometer for iTRAQ quantitation. *J Proteome Res*, 2008, 7(11):4831-4840.

38. Phanstiel D, Unwin R, McAlister GC, & Coon JJ. Peptide quantification using 8-plex isobaric tags and electron transfer dissociation tandem mass spectrometry. *Anal Chem*, 2009, 81(4):1693-1698.

39. Frese CK, Altelaar AF, Hennrich ML, Nolting D, Zeller M, Griep-Raming J, Heck AJ, & Mohammed S. Improved peptide identification by targeted fragmentation using CID, HCD and ETD on an LTQ-Orbitrap Velos. *J Proteome Res*, 2011, 10(5):2377-2388.

40. Clauser KR, Baker P, & Burlingame AL. Role of Accurate Mass Measurement (±10 ppm) in Protein Identification Strategies Employing MS or MS/MS and Database Searching. *Analytical Chemistry*, 1999, 71(14):2871-2882.

41. Sadygov RG, Cociorva D, & Yates JR. Large-scale database searching using tandem mass spectra: Looking up the answer in the back of the book. *Nat Meth*, 2004, 1(3):195-202.

42. Sheynkman GM, Shortreed MR, Frey BL, Scalf M, & Smith LM. Large-Scale Mass Spectrometric Detection of Variant Peptides Resulting from Nonsynonymous Nucleotide Differences. *Journal of Proteome Research*, 2013, 13(1):228-240.

43. Sheynkman GM, Shortreed MR, Frey BL, & Smith LM. Discovery and Mass Spectrometric Analysis of Novel Splice-junction Peptides Using RNA-Seq. *Molecular & Cellular Proteomics*, 2013, 12(8):2341-2353.

44. Volkening JD, Bailey DJ, Rose CM, Grimsrud PA, Howes-Podoll M, Venkateshwaran M, Westphall MS, Ané J-M, Coon JJ, & Sussman MR. A Proteogenomic Survey of the Medicago truncatula Genome. *Molecular & Cellular Proteomics*, 2012, 11(10):933-944.

45. Elias JE & Gygi SP. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Meth*, 2007, 4(3):207-214.

46. Geer LY, Markey SP, Kowalak JA, Wagner L, Xu M, Maynard DM, Yang X, Shi W, & Bryant SH. Open Mass Spectrometry Search Algorithm. *Journal of Proteome Research*, 2004, 3(5):958-964.

47. Perkins DN, Pappin DJC, Creasy DM, & Cottrell JS. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 1999, 20(18):3551-3567.

48. Wenger CD, Phanstiel DH, Lee MV, Bailey DJ, & Coon JJ. COMPASS: A suite of pre- and post-search proteomics software tools for OMSSA. *Proteomics* 2011, 11(6):1064-1074.

49. Nesvizhskii AI & Aebersold R. Interpretation of Shotgun Proteomic Data: The Protein Inference Problem. *Molecular & Cellular Proteomics*, 2005, 4(10):1419-1440.

50. Wang W, Zhou H, Lin H, Roy S, Shaler TA, Hill LR, Norton S, Kumar P, Anderle M, & Becker CH. Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards. *Anal Chem*, 2003, 75(18):4818-4826.

51. Liu H, Sadygov RG, & Yates JR, 3rd. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem*, 2004, 76(14):4193-4201.

52. Cox J, Hein MY, Luber CA, Paron I, Nagaraj N, & Mann M. MaxLFQ allows accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction. *Molecular & Cellular Proteomics*, 2014.

53. Weisser H, Nahnsen S, Grossmann J, Nilse L, Quandt A, Brauer H, Sturm M, Kenar E, Kohlbacher O, Aebersold R, & Malmström L. An Automated Pipeline for High-Throughput Label-Free Quantitative Proteomics. *Journal of Proteome Research*, 2013, 12(4):1628-1644.

54. Rosenberger G, Ludwig C, Röst HL, Aebersold R, & Malmström L. aLFQ: An R-package for estimating absolute protein quantities from label-free LC-MS/MS proteomics data. *Bioinformatics*, 2014.

55. Panchaud A, Scherl A, Shaffer SA, von Haller PD, Kulasekara HD, Miller SI, & Goodlett DR. Precursor Acquisition Independent From Ion Count: How to Dive Deeper into the Proteomics Ocean. *Analytical Chemistry*, 2009, 81(15):6481-6488.

56. Venable JD, Dong M-Q, Wohlschlegel J, Dillin A, & Yates JR. Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nat Meth*, 2004, 1(1):39-45.

57. Gillet LC, Navarro P, Tate S, Röst H, Selevsek N, Reiter L, Bonner R, & Aebersold R. Targeted Data Extraction of the MS/MS Spectra Generated by Data-independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis. *Molecular & Cellular Proteomics*, 2012, 11(6).

58. Rost HL, Rosenberger G, Navarro P, Gillet L, Miladinovic SM, Schubert OT, Wolski W, Collins BC, Malmstrom J, Malmstrom L, & Aebersold R. OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. *Nat Biotech*, 2014, 32(3):219-223.

59. Merrill AE & Coon JJ. Quantifying proteomes and their post-translational modifications by stable isotope label-based mass spectrometry. *Current Opinion in Chemical Biology*, 2013, 17(5):779-786.

60. Jiang H & English AM. Quantitative analysis of the yeast proteome by incorporation of isotopically labeled leucine. *J Proteome Res*, 2002, 1(4):345-350.

61. Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, & Mann M. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics*, 2002, 1(5):376-386.

62. Ong SE, Foster LJ, & Mann M. Mass spectrometric-based approaches in quantitative proteomics. *Methods*, 2003, 29(2):124-130.

63. Pimienta G, Chaerkady R, & Pandey A. SILAC for global phosphoproteomic analysis. *Methods Mol Biol*, 2009, 527:107-116, x.

64. Merrill AE, Hebert AS, MacGilvray ME, Rose CM, Bailey DJ, Bradley JC, Wood WW, ElMasri M, Westphall MS, Gasch AP, & Coon JJ. NeuCode labels for relative protein quantification. *Molecular & Cellular Proteomics*, 2014.

65. Hebert AS, Merrill AE, Bailey DJ, Still AJ, Westphall MS, Strieter ER, Pagliarini DJ, & Coon JJ. Neutron-encoded mass signatures for multiplexed proteome quantification. *Nat Meth*, 2013, 10(4):332-334.

66. Rose CM, Merrill AE, Bailey DJ, Hebert AS, Westphall MS, & Coon JJ. Neutron Encoded Labeling for Peptide Identification. *Analytical Chemistry*, 2013, 85(10):5129-5137.

67. Richards AL, Vincent CE, Guthals A, Rose CM, Westphall MS, Bandeira N, & Coon JJ. Neutron-encoded Signatures Enable Product Ion Annotation From Tandem Mass Spectra. *Molecular & Cellular Proteomics*, 2013, 12(12):3812-3823.

68. Rhoads TW, Rose CM, Bailey DJ, Riley NM, Molden RC, Nestler AJ, Merrill AE, Smith LM, Hebert AS, Westphall MS, Pagliarini DJ, Garcia BA, & Coon JJ. Neutron-Encoded Mass Signatures for Quantitative Top-Down Proteomics. *Analytical Chemistry*, 2014, 86(5):2314-2319.

69. Makarov A, Denisov E, & Lange O. Performance Evaluation of a High-field Orbitrap Mass Analyzer. *Journal of the American Society for Mass Spectrometry*, 2009, 20(8):1391-1396.

70. Michalski A, Damoc E, Hauschild J-P, Lange O, Wieghaus A, Makarov A, Nagaraj N, Cox J, Mann M, & Horning S. Mass Spectrometry-based Proteomics Using Q Exactive, a High-performance Benchtop Quadrupole Orbitrap Mass Spectrometer. *Molecular & Cellular Proteomics*, 2011, 10(9).

71. Bruce JE, Anderson GA, Lin C-Y, Gorshkov M, Rockwood AL, & Smith RD. A novel high-performance Fourier transform ion cyclotron resonance cell for improved biopolymer characterization. *Journal of Mass Spectrometry*, 2000, 35(1):85-94.

72. Syka JEP, Marto JA, Bai DL, Horning S, Senko MW, Schwartz JC, Ueberheide B, Garcia B, Busby S, Muratore T, Shabanowitz J, & Hunt DF. Novel Linear Quadrupole Ion Trap/FT Mass Spectrometer: Performance Characterization and Use in the Comparative Analysis of Histone H3 Post-translational Modifications. *Journal of Proteome Research*, 2004, 3(3):621-626.

73. Kaiser N, Quinn J, Blakney G, Hendrickson C, & Marshall A. A Novel 9.4 Tesla FTICR Mass Spectrometer with Improved Sensitivity, Mass Resolution, and Mass Range. *Journal of the American Society for Mass Spectrometry*, 2011, 22(8):1343-1351.

74. Denisov E, Damoc E, Lange O, & Makarov A. Orbitrap mass spectrometry with resolving powers above 1,000,000. *International Journal of Mass Spectrometry*, 2012, 325–327(0):80-85.

75. Michalski A, Damoc E, Lange O, Denisov E, Nolting D, Müller M, Viner R, Schwartz J, Remes P, Belford M, Dunyach J-J, Cox J, Horning S, Mann M, & Makarov A. Ultra High Resolution Linear Ion Trap Orbitrap Mass Spectrometer (Orbitrap Elite) Facilitates Top Down LC MS/MS and Versatile Peptide Fragmentation Modes. *Molecular & Cellular Proteomics*, 2012, 11(3).

76. Hebert AS, Merrill AE, Bailey DJ, Still AJ, Westphall MS, Streiter ER, Pagliarini DJ, & Coon JJ. Neutron-encoded mass signatures for multi-plexed proteome quantification. *Nature Methods*, 2012:In Press.

77. Thompson A, Schafer J, Kuhn K, Kienle S, Schwarz J, Schmidt G, Neumann T, Johnstone R, Mohammed AK, & Hamon C. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem*, 2003, 75(8):1895-1904.

78. Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, Khainovski N, Pillai S, Dey S, Daniels S, Purkayastha S, Juhasz P, Martin S, Bartlet-Jones M, He F, Jacobson A, & Pappin DJ. Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics*, 2004, 3(12):1154-1169.

79. Choe L, D'Ascenzo M, Relkin NR, Pappin D, Ross P, Williamson B, Guertin S, Pribil P, & Lee KH. 8-plex quantitation of changes in cerebrospinal fluid protein expression in subjects undergoing intravenous immunoglobulin treatment for Alzheimer's disease. *Proteomics*, 2007, 7(20):3651-3660.

80. Rose CM, Venkateshwaran M, Volkening JD, Grimsrud PA, Maeda J, Bailey DJ, Park K, Howes-Podoll M, den Os D, Yeun LH, Westphall MS, Sussman MR, Ané J-M, & Coon JJ. Rapid Phosphoproteomic and Transcriptomic Changes in the Rhizobia-legume Symbiosis. *Molecular & Cellular Proteomics*, 2012, 11(9):724-744.

81. Honarpour N, Rose CM, Brumbaugh J, Anderson J, Graham RLJ, Sweredoski MJ, Hess S, Coon JJ, & Deshaies RJ. F-box Protein FBXL16 Binds PP2A-B55α and Regulates Differentiation of Embryonic Stem Cells along the FLK1+ Lineage. *Molecular & Cellular Proteomics*, 2014, 13(3):780-791.

82. McAlister GC, Huttlin EL, Haas W, Ting L, Jedrychowski MP, Rogers JC, Kuhn K, Pike I, Grothe RA, Blethrow JD, & Gygi SP. Increasing the Multiplexing Capacity of TMTs Using Reporter Ion Isotopologues with Isobaric Masses. *Analytical Chemistry*, 2012, 84(17):7469-7478.

83. Werner T, Becher I, Sweetman G, Doce C, Savitski MM, & Bantscheff M. High-Resolution Enabled TMT 8-plexing. *Analytical Chemistry*, 2012, 84(16):7188-7194.

84. Werner T, Sweetman G, Savitski MF, Mathieson T, Bantscheff M, & Savitski MM. Ion Coalescence of Neutron Encoded TMT 10-Plex Reporter Ions. *Analytical Chemistry*, 2014, 86(7):3594-3601.

85. Ow SY, Salim M, Noirel J, Evans C, Rehman I, & Wright PC. iTRAQ Underestimation in Simple and Complex Mixtures: "The Good, the Bad and the Ugly". *Journal of Proteome Research*, 2009, 8(11):5347-5355.

86. Wenger CD, Lee MV, Hebert AS, McAlister GC, Phanstiel DH, Westphall MS, & Coon JJ. Gas-phase purification enables accurate, multiplexed proteome quantification with isobaric tagging. *Nat Meth*, 2011, 8(11):933-935.

87. Ting L, Rad R, Gygi SP, & Haas W. MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. *Nat Meth*, 2011, 8(11):937-940.

88. McAlister GC, Nusinow DP, Jedrychowski MP, Wühr M, Huttlin EL, Erickson BK, Rad R, Haas W, & Gygi SP. MultiNotch MS3 Enables Accurate, Sensitive, and Multiplexed Detection of Differential Expression across Cancer Cell Line Proteomes. *Analytical Chemistry*, 2014, 86(14):7150-7158.

89. Gerber SA, Rush J, Stemman O, Kirschner MW, & Gygi SP. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc Natl Acad Sci U S A*, 2003, 100(12):6940-6945.

90. Gerber SA, Kettenbach AN, Rush J, & Gygi SP. The absolute quantification strategy: application to phosphorylation profiling of human separase serine 1126. *Methods Mol Biol*, 2007, 359:71-86.

91. Lange VPPDBAR. Selected reaction monitoring for quantitative proteomics: a tutorial. *Molecular Systems Biology*, 2008, 4(1).

92. Peterson AC, Russell JD, Bailey DJ, Westphall MS, & Coon JJ. Parallel Reaction Monitoring for High Resolution and High Mass Accuracy Quantitative, Targeted Proteomics. *Molecular & Cellular Proteomics*, 2012, 11(11):1475-1488.

93. Gallien S, Duriez E, Crone C, Kellmann M, Moehring T, & Domon B. Targeted Proteomic Quantification on Quadrupole-Orbitrap Mass Spectrometer. *Molecular & Cellular Proteomics*, 2012.

94. Bailey DJ, Rose CM, McAlister GC, Brumbaugh J, Yu P, Wenger CD, Westphall MS, Thomson JA, & Coon JJ. Instant spectral assignment for advanced decision tree-driven mass spectrometry. *Proceedings of the National Academy of Sciences*, 2012, 109(22):8411-8416.

95. Smith LM & Kelleher NL. Proteoform: a single term describing protein complexity. *Nat Meth*, 2013, 10(3):186-187.

96. Rose C, Russell J, Ledvina A, McAlister G, Westphall M, Griep-Raming J, Schwartz J, Coon J, & Syka JP. Multipurpose Dissociation Cell for Enhanced ETD of Intact Protein Species. *J. Am. Soc. Mass Spectrom.*, 2013, 24(6):816-827.

97. Ledvina A, Rose C, McAlister G, Syka JP, Westphall M, Griep-Raming J, Schwartz J, & Coon J. Activated Ion ETD Performed in a Modified Collision Cell on a Hybrid QLT-Oribtrap Mass Spectrometer. *J. Am. Soc. Mass Spectrom.*, 2013, 24(11):1623-1633.

Chapter 2

Rapid phosphoproteomic and transcriptomic changes in the rhizobia-legume symbiosis

CMR designed research, performed experiments, analyzed data, and wrote the paper.

This chapter has been published:

Rose C. M.*, Venkateshwaran M.*, Volkening J. D., Grimsrud P. A., Maeda J, Bailey D. J., Park K., Howes-Podoll M., Os D. D., Yeun H. L., Westphall M. S., Coon J. J., Sussman M. R., Ane J. M. "Rapid Phosphoproteomic and Transcriptomic Changes in the Rhizobia-Legume Symbiosis." *Molecular & Cellular Proteomics*, **2012**, (11):724-744.

*co-first author

Abstract

Symbiotic associations between legumes and rhizobia usually commence with the perception of bacterial lipochitooligosaccharides, known as Nod factors (NF), which triggers rapid cellular and molecular responses in host plants. We report here deep untargeted tandem mass spectrometry-based measurements of rapid NF-induced changes in the phosphorylation status of 13,506 phosphosites in 7,739 proteins from the model legume *Medicago truncatula*. To place these phosphorylation changes within a biological context, quantitative phosphoproteomic and RNA measurements in wild-type plants were compared with those observed in mutants, one defective in NF perception (*nfp*) and one defective in downstream signal transduction events (*dmi3*). Our study quantified the early phosphorylation and transcription dynamics that are specifically associated with NF-signaling, confirmed a *dmi3*-mediated feedback loop in the pathway, and suggested 'cryptic' NF-signaling pathways, some of them being also involved in the response to symbiotic arbuscular mycorrhizal fungi.

Introduction

Legumes have the ability to form a very efficient symbiotic association with rhizobia to meet their nitrogen demand. This results in root nodule formation, inside which the rhizobia can fix atmospheric nitrogen efficiently and transfer it to the plants, in exchange for a carbon source. This interaction is often characterized by a high level of host specificity and generally requires an exchange of diffusible signals between legumes and rhizobia. Flavonoids and isoflavonoids present in the legume root exudates induce the expression of *nod* genes in rhizobia, and are responsible for the production and secretion of bacterial lipochitooligosaccharides (LCOs), known as Nod factors (NF).[1-2] NF are generally required for rhizobial infection and nodule development. Mere application of purified NF at low concentrations (10^{-8} to 10^{-12} M) is sufficient to trigger responses in host plants similar to those elicited by the rhizobia themselves.[2-3] Responses are elicited in root cells within a few seconds to minutes after NF application and include changes in ion fluxes across the plasma membrane, as well as accumulation of reactive oxygen species.[4-5] After 15-20 minutes, oscillations of the nuclear and perinuclear calcium concentration (calcium spiking) are initiated and trigger the expression of early nodulin genes.[3] Within the first hour, legume root hairs undergo a cytoplasmic disorganization which leads to a transient swelling,[6] followed by root hair deformations.[7]

Forward and reverse genetic approaches, in the model legumes *Medicago truncatula* (Medicago) and *Lotus japonicus* (Lotus), have identified several components controlling these events **(Figure 1**). NF are perceived with high specificity by LysM receptor-like kinases residing on the plasma membrane. These kinases include Nod Factor Perception (NFP), an essential component of a signaling receptor necessary for early responses to NF and rhizobial infection.[8-9] Mutants in *NFP* are affected in all known cellular responses to NF reported so far, including calcium spiking, early nodulin genes expression and root hair deformations.[9] A leucine-rich repeats (LRR)-receptor kinase, Does not Make Infections 2 (DMI2/NORK), is also localized on the plasma membrane, and acts downstream of NFP. *Dmi2* mutants exhibit NF-induced

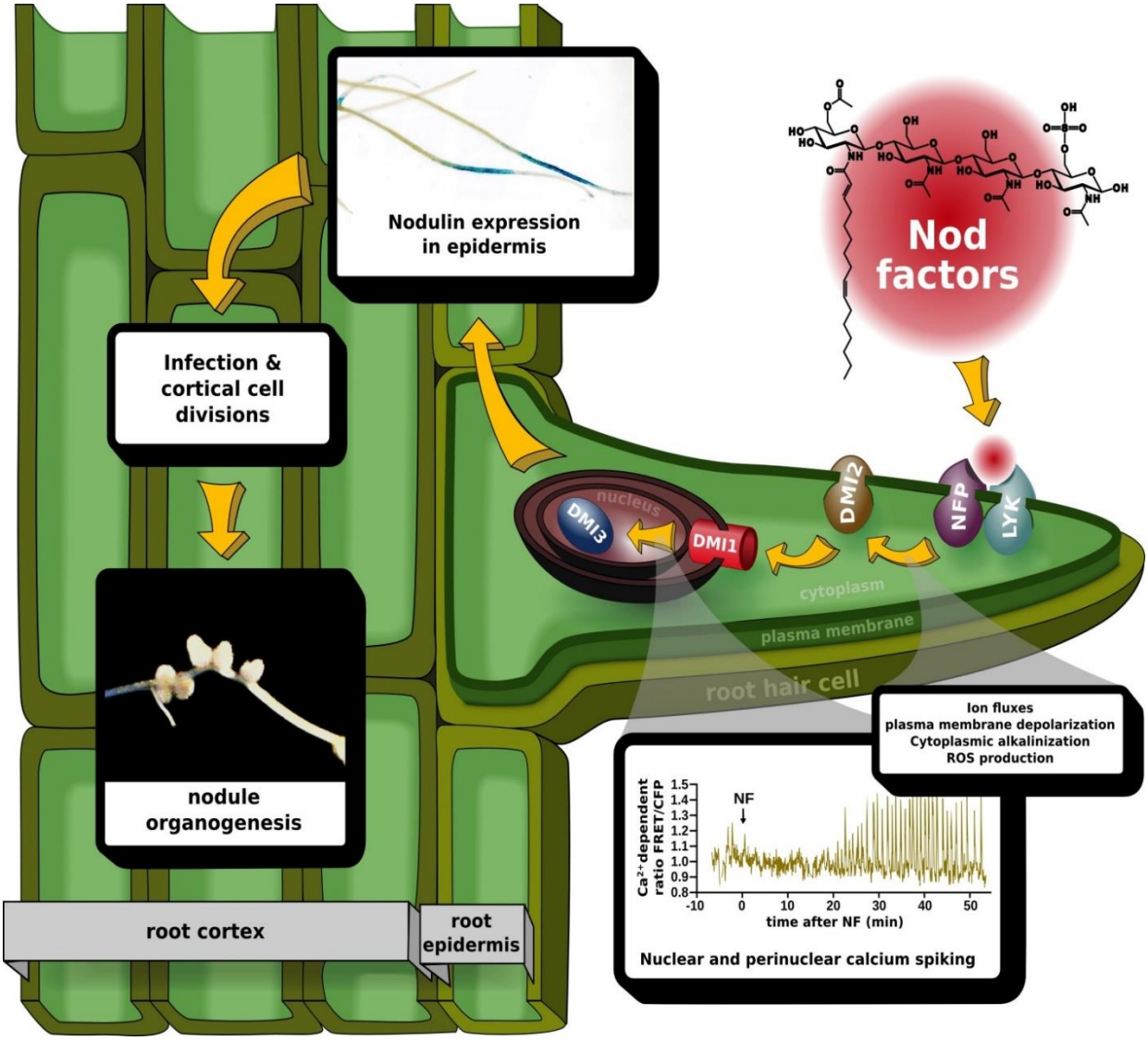


Figure 1. Schematic illustration of genetic components involved in legume-rhizobia symbiotic signaling. NF are perceived by a receptor complex containing NFP. Perception of NF triggers early responses, like ion fluxes, depolarization of plasma membrane and cytoplasmic alkalinization. The signals are transduced downstream to activate LRR-receptor kinase, DMI2 and DMI1. The nuclear ion channel DMI1 is required for NF-induced nuclear calcium spiking. Later these calcium signals are perceived and decoded by calcium- and calmodulin-dependent protein kinase, DMI3 which acts as a central player in the Nod factor signaling and coordinates the expression of symbiotic genes, including early nodulin genes. NF signaling culminates with the coordinated and synchronous progression of infection and cortical cell division, resulting in the formation of mature, infected and nitrogen-fixing nodules.

root hair deformations, such as root hair swelling and branching, but are defective in NF-induced calcium spiking and rhizobial infection.[10-11] Signals perceived at the plasma membrane level are transduced downstream to activate signaling components residing at the nuclear level. DMI1, an ion channel, and MCA8, a calcium pump, are localized on the nuclear envelope and control NF-induced calcium spiking.[11-14] These complex calcium signatures are decoded by a calcium/calmodulin-dependent protein kinase (CCaMK), called DMI3, which is localized inside the nucleus.[15-16] In response to NF, *dmi3* mutants are able to elicit calcium spiking and root hair deformations, but are unable to induce the expression of early nodulin genes and cortical cell divisions.[11, 15] Mutations in the auto-inhibitory domain of DMI3 result in constitutive activation of the protein and the spontaneous formation of root nodules even in the absence of rhizobia.[17] This indicates that DMI3 is a central regulator of the NF signaling pathway and coordinates downstream nodulation-specific transcriptional events. It has also been shown that *dmi3* mutants have increased sensitivity to low concentrations of NF, suggesting a negative feedback in the NF signal transduction pathway.[18] Downstream of DMI3, two transcription regulators of the GRAS family, Nodulation Signaling Pathway 1 and 2 (NSP1 and NSP2), control nodulin gene expression in NF-dependent manner.[19-20] NSP1 interacts with NSP2 and binds to the promoter region of *Early NODulin 11* (*ENOD11*) and *Nodule INception* (*NIN*), which are popular markers for the study of this signaling pathway.[21]

Several genes controlling NF signaling are also required for the establishment of arbuscular mycorrhization (AM), indicating the existence of a "common symbiotic pathway" between these two endosymbioses.[22] Similar to rhizobia, AM fungi release diffusible signals, known as Myc factors, which are also LCOs and can be easily collected in germinating spore exudates (GSE).[23-24] Myc factors trigger responses in host plants similar to those elicited by NF.[23] *Dmi1*, *dmi2* and *dmi3* mutants are unable to establish AM associations whereas *nfp*, *nsp1* and *nsp2* display a wild-type AM phenotype.[9,10,12,15] However, some role for NFP and NSP2 in Myc factor signaling was recently identified, suggesting that an overlap between the two LCO pathways may be more significant than previously thought.[23] Although Myc factor

receptors have not yet been identified, LYR1, a close homolog of NFP, is a potential candidate since its expression is strongly upregulated during AM symbiosis.[25-26]

Over the last decade, various large-scale approaches such as transcriptomics, proteomics, and metabolomics, have been utilized to dissect the legume-rhizobia symbiosis.[27-29] However, most of these studies looked at relatively late responses to rhizobia or NF and unfortunately were not integrated. The techniques used in the past for analyzing changes in the transcriptome during nodulation include serial analysis of gene expression (SAGE), custom-made macro- and micro-arrays, and Affymetrix GeneChip Genome Arrays.[25,28,30-33] Except for a few examples,[28,33] most of these studies explored transcriptome changes at late stages of the legume-rhizobia symbiosis (from 1 to 6 days post inoculation, dpi). Moreover, since, many of these studies were performed before the completion of the Medicago genome sequencing project,[26] the data are not readily comparable, as they use different probes and nomenclature. Several proteomic studies have been undertaken to explore changes in protein levels during the legume-rhizobia symbiosis, but they all analyzed late stages of the interaction and none of them were quantitative. Many protein kinases have been identified by genetic approaches in the NF signaling pathway and yet, only one phosphoproteomic study has been published so far.[34] This very interesting study used 2-D gel electrophoresis and radiolabeled phosphoryl groups, allowing the identification of differential phosphorylation events, but unfortunately the identification of the corresponding proteins was not possible.

The Medicago genome sequence enables us to utilize next generation RNA sequencing strategy, as well as MS-based methods to explore in depth the early phosphoproteomic, proteomic and transcriptional responses to NF. In this study, we report an integrated large-scale approach to investigate changes in the phosphoproteome, proteome, and transcriptome that occur one hour after NF treatment. To help place these results within a biological context, we compared results obtained with wild-type Medicago plants to those with two key mutants (*nfp* and *dmi3*). Our quantitative study sheds light on the NF-induced transcriptional, translational, and post-translational responses and reveals signaling pathways which previously remained undetected using conventional approaches.

Results

Large-scale detection and quantification of the Medicago proteome and phosphoproteome

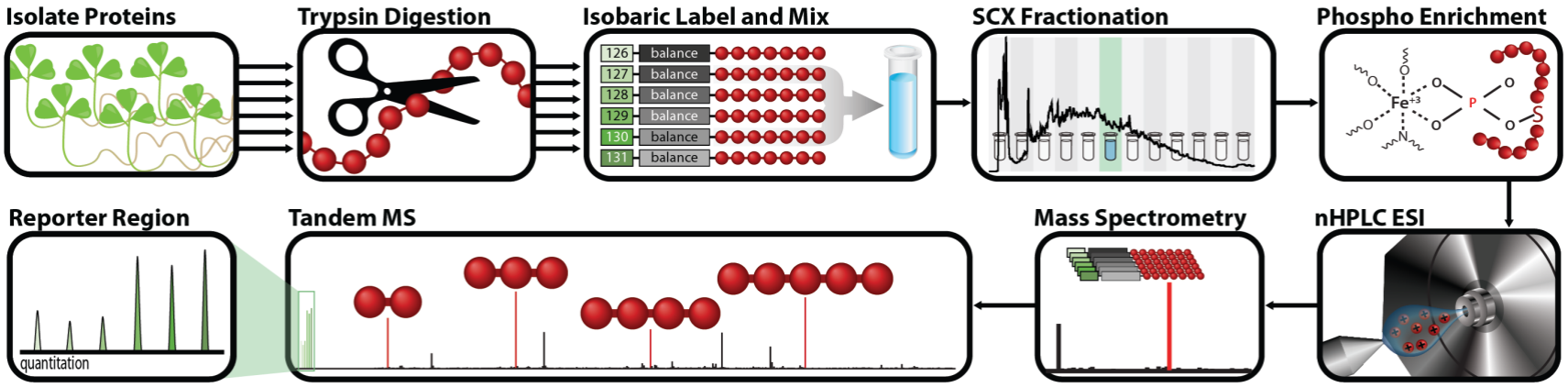
NF-induced dynamics in protein and phosphorylation levels were analyzed in Medicago by high mass accuracy tandem mass spectrometry. Data were collected from eleven quantitative isobaric tag-based[48-49] nano-HPLC-MS/MS experiments comprising various plant genotypes (wild-type, *nfp* and *dmi3*), time points (0, 10, 30 and 60 min), cellular extracts (whole cell, microsome and upper or lower membrane fractions) and growth conditions (aeroponic, hydroponic and plates). **Figure 2A** depicts the proteomic workflow used to generate the quantitative data. Proteins extracted from Medicago were subjected to tryptic digestion. To enable large-scale multiplexed quantification, peptides were labeled with isobaric tags (TMT or iTRAQ) and fractionated *via* strong cation exchange (SCX) to reduce sample complexity. All SCX fractions were enriched for phosphopeptides using immobilized metal affinity chromatography (IMAC). Both unmodified and phosphopeptide fractions were analyzed using an ETD-enabled Velos hybrid linear ion trap-orbitrap mass spectrometer (Thermo-Fisher Scientific).

Tandem mass spectra were searched against a concatenated target-decoy database comprising Medicago protein sequences. Across all experiments, more than 2,400 hours of instrument time generated ~10,000,000 high resolution MS/MS spectra of which 2,033,877 were confidently (1% FDR) assigned to 78,635 unique peptide sequences mapping to 7,739 proteins (**Figure 2B**). This includes 15,335 unique phosphopeptides, containing 13,506 non-redundant sites of phosphorylation localized to specific amino acids within 3,926 phosphoproteins. This discussion includes analysis of protein and phosphoprotein identification overlap between experiments, as well as analysis of protein and phosphoprotein identification increases due to cellular fractionation.

The Medicago proteome is largely unaltered during the first hour of response to NF

Quantitative proteome analysis revealed few changes in protein expression in response to NF (0.16%). Specifically, only 13 proteins are significantly altered ($p<0.05$, Student's t-test, assuming equal

A



B

Experiment	1	2	3	4	5	6	7	8	9	10	11	Overall
Growth	Aeroponic	Hydroponic	Hydroponic	Hydroponic	Plate	Plate	Plate	Plate	Plate	Plate	Plate	-
Cellular Fraciton	Whole Cell	Whole Cell	Whole Cell	Whole Cell	Upper Membrane	Lower Membrane	Microsome	Whole Cell	Whole Cell	Upper Membrane	Lower Membrane	-
Genotype	WT, *dmi3*	WT, *dmi3*	WT, *nfp*	WT, *nfp*	*nfp*	*nfp*	WT, *nfp*, *dmi3*	WT	WT	WT	WT	-
Biological Replicates	1	1	1	1	3	3	1	3	3	3	3	-
Time Points	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	10/30 min	N/A	N/A	-
Unique Peptides*	34195	37436	28976	27723	11129	9152	13073	12218	8581	7430	10309	63290
Unique Phosphopeptides**	5975	9694	8231	5471	2976	4069	2055	4129	2782	1032	1898	15335
Proteins	6315	6983	6290	6521	3996	4071	3956	4256	3603	2865	3604	7739
Phosphoproteins	2392	3076	2800	2382	1351	1706	1000	1807	1320	565	977	3926
Localized Phosphosites	5588	7988	6982	4936	2656	3643	1242	2627	1879	829	1582	13506

*Unique based on amino acid sequence, does not include phosphopeptides
**Unique based on amino acid sequence and number of phosphosites

Figure 2. Proteomic and phosphoproteomic workflow, experimental design, and results

(A) Schematic illustration of proteomic and phosphoproteomic workflow. (B) Summary of proteomic and phosphoproteomic experiments. Peptides were determined to be unique based on amino acid sequence alone (i.e., variable modifications such as methionine oxidation are not taken into account), while phosphopeptides were determine to be unique based on both amino acid sequence and number of phosphosites.

variances) more than 1.35-fold in wild-type plants. In contrast, transcriptomic data revealed 136 genes significantly altered in wild-type plants (see below). The disparity between proteomic and transcriptomic changes is expected, as changes in RNA expression require time to affect protein level changes. The largely unchanged proteome suggests rapid cellular responses (i.e, calcium spiking, membrane depolarization, etc.) result from NF-induced post-translational modifications (i.e., phosphorylation) which are independent of protein translations.

The Medicago proteome is largely unaltered during the first hour of response to NF

Quantitative proteome analysis revealed few changes in protein expression in response to NF (0.16%). Specifically, only 13 proteins are significantly altered ($p<0.05$, Student's t-test, assuming equal variances) more than 1.35-fold in wild-type plants. In contrast, transcriptomic data revealed 136 genes significantly altered in wild-type plants (see below). The disparity between proteomic and transcriptomic changes is expected, as changes in RNA expression require time to affect protein level changes. The largely unchanged proteome suggests rapid cellular responses (i.e, calcium spiking, membrane depolarization, etc.) result from NF-induced post-translational modifications (i.e., phosphorylation) which are independent of protein translations.

NF induce rapid changes in Medicago phosphoproteome

The phosphoproteomic response of Medicago plants to NF treatment was quantified for 13,185 unique phosphoisoforms. 454 phosphoisoforms exhibited a significant change in phosphorylation upon NF treatment ($p<0.05$, Student's t-test, assuming equal variances). Considering only phosphorylation changes greater than 1.35-fold results in 98 candidate phosphoisoforms, several of these targets are described in detail in the discussion.

To elucidate the phosphorylation dynamics within the first hour of NF treatment we designed an 8-Plex time course experiment, including three biological replicates for the 0 and 60 min time points and one biological replicate for the 10 and 30 min time points. This experiment, listed as experiment 9 in **Figure 2**,

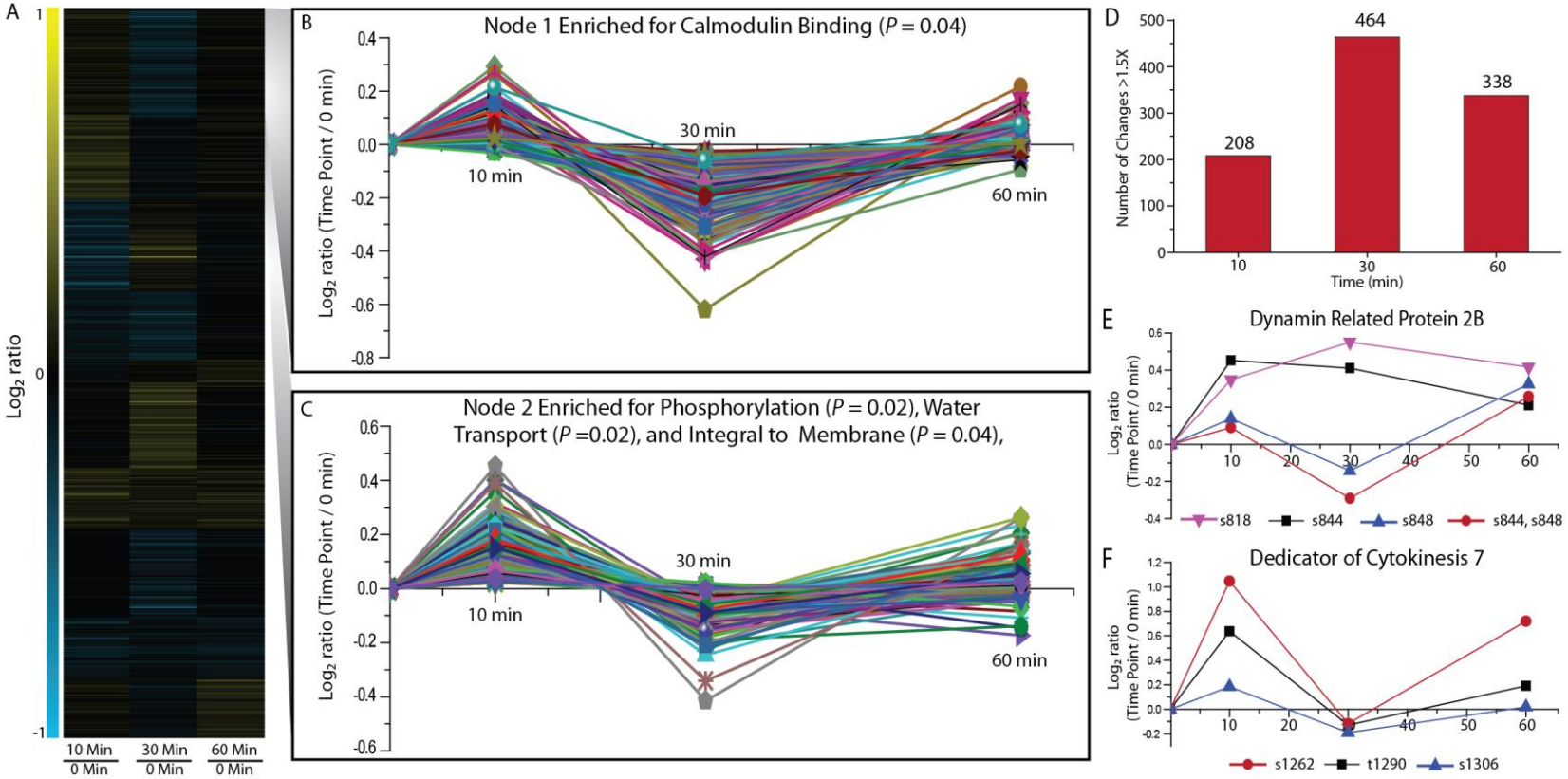


Figure 3. Time Course Analysis of NF Induced Phosphorylation Dynamics

(A) Heat map representation of phosphorylation changes at 10, 30, and 60 min post NF treatment. (C) Functional analysis of the phosphoisoforms comprising node 2 identified the GO terms phosphorylation, water transport, and integral to membrane as significantly enriched ($P < 0.05$, Fisher exact test with Benjimini-Hochberg correction for multiple comparisons). (D) Global phosphorylation changes for each time point. To investigate global phosphorylation activity, number of 1.5 fold phosphorylation changes are plotted for each time point. Phosphorylation changes occur rapidly, 10 min post NF treatment and continue to intensify until 30 min. At 60 min, the number of changes is still large with 338 phosphoisoforms displaying a 1.5 fold change. (E) Dynamin 2B phosphorylation dynamics within the first hour post NF treatment. (F) DOCK7 phosphoryation dynamics within the first hour post NF treatment.

produced 1,468 localized and quantified phosphoisoforms. The $\log_2$ ratio of each time point compared to the 0 min time point was calculated (using the average for both the 0 and 60 min time points) and grouped into 10 nodes using K-means clustering (**Figure 3A**). Functional analysis determined node 1 was enriched for the gene ontology (GO) term calmodulin binding (**Figure 3B**), while node 2 was enriched for the GO terms phosphorylation, water transport, and integral to membrane (**Figure 3C**). No significant enrichment of GO terms was determined for the remaining nodes of the time course. Calmodulin binding proteins are an interesting group of candidates as calcium signatures both at the plasma membrane and nuclear levels play an essential role in NF signaling. To further investigate the phosphorylation dynamics within the time course, the number of phosphorylation events altered greater than 1.5 fold was counted for each time point (**Figure 3D**). The greatest number of 1.5 fold changes occurs at 30 min, corresponding to the amount of time before calcium spiking is initiated within the cell.

Two proteins highlighted the importance of time course data. Dynamin-related protein 2B (DRP2B, Medtr4g030140) belongs to a GTPase family of proteins responsible for membrane dynamics, vesicle trafficking, and is likely a key player in the NF signaling cascade (**Figure 3E**). Two separate phosphosites (s818 and s848) demonstrate a significant increase ($p<0.05$) in phosphorylation in the time course experiment (**Figure 3E**). Two additional DRP2B phosphoisoforms, s844 and the doubly phosphorylated s844/s848, are also plotted in **Figure 3E**. For DRP2B, s848 and the doubly phosphorylated s844/s848 share a distinct expression pattern, suggesting the removal and addition of s848 from these isoforms. While the s844 and s818 phosphoisoform exhibit a rapid and sustained increase in phosphorylation over the course of the experiment, further implicating DRP2B in the NF signaling cascade. Similarly, time course data reveals three phosphosites at s1263, t1290 and s1306 for Dedicator of Cytokinesis 7 (DOCK7, Medtr8g056900). The s1306 phosphosite belongs to the group of 98 proteins significantly altered in the overall analysis, *vide supa*, while s1263 and t1290 fall short of statistical significance, but follow a similar phosphorylation pattern over the time course (**Figure 3F**). These data provide evidence of global phosphorylation changes

on DOCK7 in response to NF treatment. While it escapes the scope of the current paper, the cases of DRP2B and DOCK7 demonstrate the power of analyzing time course data for each individual protein.

Symbiosis-defective mutants display altered phosphoproteome upon NF treatment

To investigate the dependence of NF-induced phosphorylation events on NFP- and DMI3-induced signaling, the phosphoproteomes of *nfp* and *dmi3* mutants were analyzed and global phosphorylation dynamics compared to the wild-type response. 13,993 unique localized phosphoisoforms were detected in at least one genotype, with individual genotypes containing 13,185, 10,198 and 10,062 isoforms for wild-type, *nfp* and *dmi3* plants, respectively. The number of biological replicates across genotypes was not equal, making a direct comparison of significant changes difficult. Therefore, we compared the number of phosphorylation events changing greater than 1.5-fold post-NF treatment. Wild-type plants readily respond NF application, as the phosphorylation level of 6.3% of isoforms changed more than 1.5-fold. *nfp* plants demonstrate a reduced phosphorylation response with only 4.0% of isoforms altered in the same manner (**Figure 4A**). In addition, *nfp* mutants contain a greater percentage of isoforms altered less than 1.2-fold (81% for *nfp*, 72% for wild-type), implicating NFP as a major contributor to the initiation of NF signaling. However, the presence of differential phosphorylation exhibited in *nfp* provides clear evidence for the involvement of additional NF receptors, prompting our transcriptomic approach discussed below. Interestingly, *dmi3* appears hypersensitive to NF, as the phosphorylation status of 7.8% of isoforms was altered more than 1.5-fold after treatment. The increased phosphorylation response in *dmi3* confirms its role in a negative feedback mechanism.[18] In contrast, NF-induced transcriptional changes in *dmi3* showed only one regulated gene, confirming that DMI3 is a key regulator of NF-induced transcription. Altogether, our data suggest that this negative feedback loop probably involves post-translational, but not transcriptional, mechanisms.

To analyze phosphorylation changes directly related to NF signal transduction, we focused on phosphoisoforms that displayed a fold change greater than 1.35-fold in wild-type plants and less than 1.25-fold in *nfp*, indicating that these targets are downstream of the NFP receptor in the signaling cascade (**Figure**

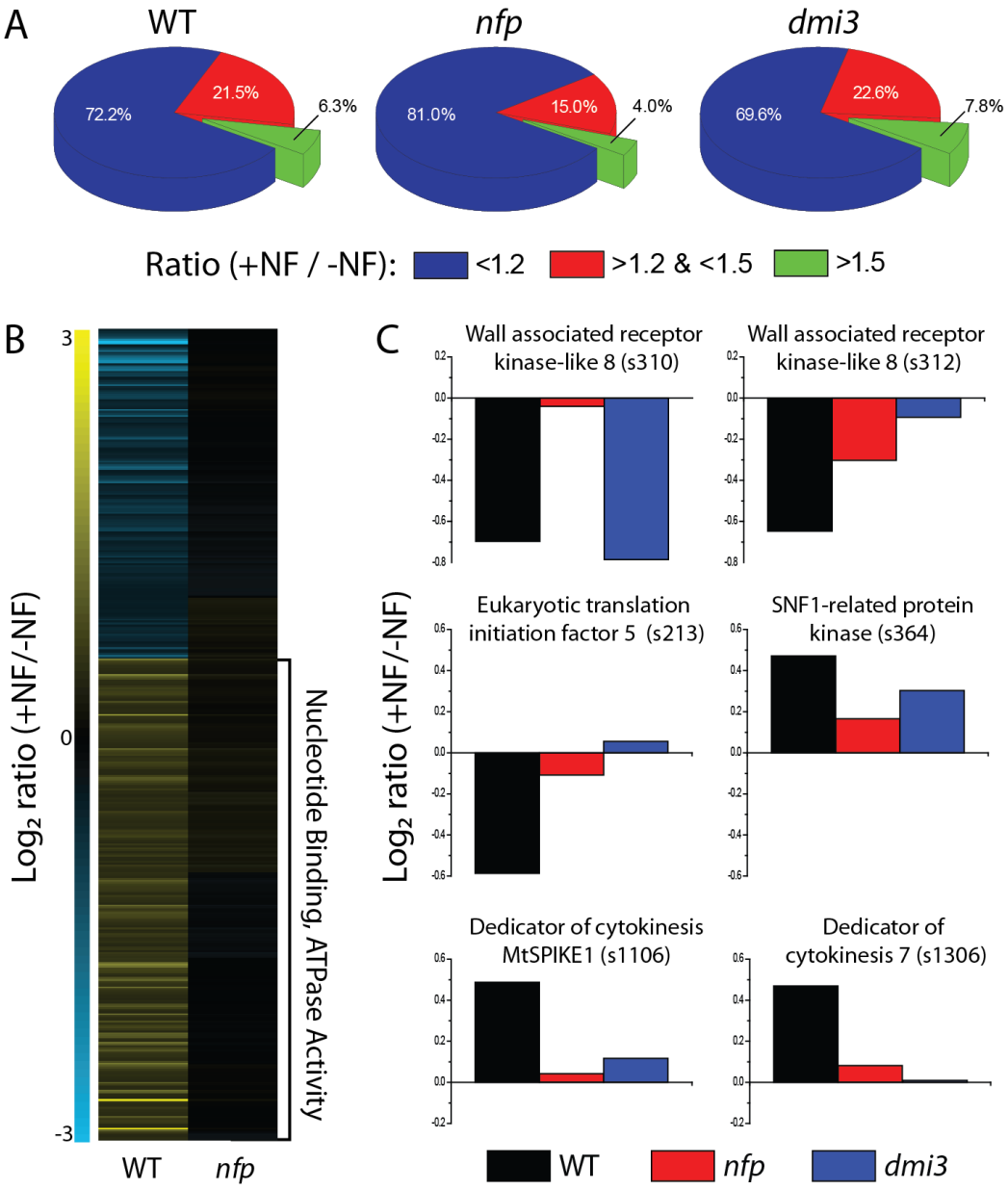


Figure 4. Global view of changes in the proteome and phosphoproteome from all the experiments combined.

(A) Pie charts displaying the distribution of phosphorylation changes for wild-type, *nfp* and *dmi3*. (B) Heatmap of phosphoisoforms altered in wild-type and not in *nfp*. Phophoisofroms exhibiting a fold change more than 1.35 in wild-type plants and less then 1.25 in *nfp* were grouped *via* hierarchical clustering. (C) Representative proteins significantly altered in wild-type and not in *nfp*. Six pohsphoisoforms that were significantly ($p<0.05$, Student's t-test, assuming equal variances) altered more than 1.35 fold in wild-type plants and less than 1.25 fold in *nfp* demonstrate the utility of comparing wild-type and mutant measurements. These examples contain phosphoisoforms which display both *nfp* dependent regulation and *dmi3* dependent regulation.

4B). Functional analysis of phosphorylation responses to NF revealed that phosphoisoforms in wild-type plants exhibiting increased phosphorylation were enriched for the GO terms nucleotide binding, ATPase activity and nucleotide triphosphate activity, while those phosphorylation changes which exhibited down-regulation in wild-type plants revealed no significant enrichment (**Figure 4B**). To identify specific proteins involved in NF signaling, 64 phosphoisoforms significantly altered ($p<0.05$) more than 1.35-fold in wild-type plants and less than 1.25-fold in *nfp* were analyzed. From these 64, 24 displayed similar fold changes in both wild-type and *dmi3* plants, indicating that these are most likely affected before DMI3 activation. The other 40 isoforms were either not detected or exhibited expression in the *dmi3* mutant different from wild-type plants, suggesting that these NF-induced phosphorylation changes are dependent on DMI3 activity. The phosphorylation response to NF for all three genotypes for six of the 64 candidates is displayed in **Figure 4C**. Phosphorylation sites in two different amino acids within the same tryptic peptide, wall-associated receptor kinase-like 8 (WAKL8, Medtr5g085790.1), demonstrates two separate patterns of phosphorylation dynamics between mutants (**Figure 4C**). Phosphorylation of s310 in WAKL8 is down-regulated in both wild-type and *dmi3* plants upon NF treatment, while relatively unaltered in *nfp*, suggesting phosphorylation of this site is dependent on NFP, but not DMI3. s312 in WAKL8 displays increased phosphorylation in wild-type plants when compared to *nfp* and *dmi3*, suggesting this phosphorylation is DMI3-dependent. In addition, in both *nfp* and *dmi3* mutants, the s213 phosphoisoform of eIF5, and two phosphoisoforms of DOCK-family proteins (MtSPIKE1 s1106 and DOCK7 s1306) display reduced phosphorylation levels upon NF treatment and in comparison with wild-type plants, suggesting regulation of these phosphosites downstream of DMI3. While the s364 phosphoisoform of SNF1-related protein kinase is down-regulated upon NF treatment in both *nfp* and *dmi3*, it is reduced less in *dmi3*, suggesting NFP-dependent regulation.

NF induce rapid changes in the Medicago transcriptome

To monitor changes in gene expression one hour post-NF treatment, next generation RNA sequencing was performed in wild-type, *nfp* and *dmi3*. A total of 489.8 million 100 bp single-end reads (49

gigabases) were generated from the mRNA libraries. Individual sample sets ranged from 20.4 – 35.0 million reads (mean of 27.2 ±3.8 M reads). Of these, an average of 76.3% ±3.5% mapped to a unique location on the genome – 69.1% ±3.3% to annotated exons and 7.2% ±0.3% to non-exonic locations. The remaining reads mapped to multiple locations in the genome (9.1% ±3.1%) or did not map at all (14.6% ±0.7%) The variation in multireads was due primarily to per-sample variation in rRNA content, which ranged from 0.5 – 10.5% of total reads. A total of 25,237 genes from Mt3.5v4 averaged 10 or more mapped reads per kb transcript over all normalized samples as evidence of transcription, with an overall dynamic range of roughly 10^4 across this group of genes.

In wild-type plants, ~136 genes were differentially regulated in response to NF (pooled dispersion, 5% FDR). Of these, 116 genes were significantly upregulated and 20 were down-regulated. The major groups of genes differentially regulated in wild-type plants encode transcription factors, protein kinases, cell wall-modifying enzymes, defense-related proteins, hormone signaling proteins, transporters and flavonoid biosynthesis enzymes (**Figure 5B**). As preliminary validation of our RNA sequencing data, we checked the expression of genes with known roles in NF signaling. *NIN* (*Medtr5g099060*) is an early marker for the activation of the NF signaling pathway in Medicago and Lotus, encoding a transcription factor required for both infection thread formation and cortical cell divisions.[50-51] Our data confirmed up-regulation of *NIN* expression one hour post-NF treatment. Transcriptional regulators from the GRAS family (e.g., NSP1 and NSP2), as well as AP2/ERF-like transcription factors (e.g., ERN1, 2 and 3), also have known roles in early symbiotic signaling.[19,20,52-53] Our data confirms prior data that *ERN1* (*Medtr7g085810*) is upregulated one hour post-NF treatment.[53] Genes encoding AP2/ERF-like transcription factors (*Medtr7g085810, contig_82104, contig_52379, contig_16415, contig_71245* and *contig_75657*) were also upregulated upon NF treatment. We validated our next generation RNA sequencing approach using quantitative RT-PCR (qRT-PCR) for several additional candidates. Expression levels of genes encoding ethylene-responsive transcription factor 1B (*contig_82104*), ethylene-responsive

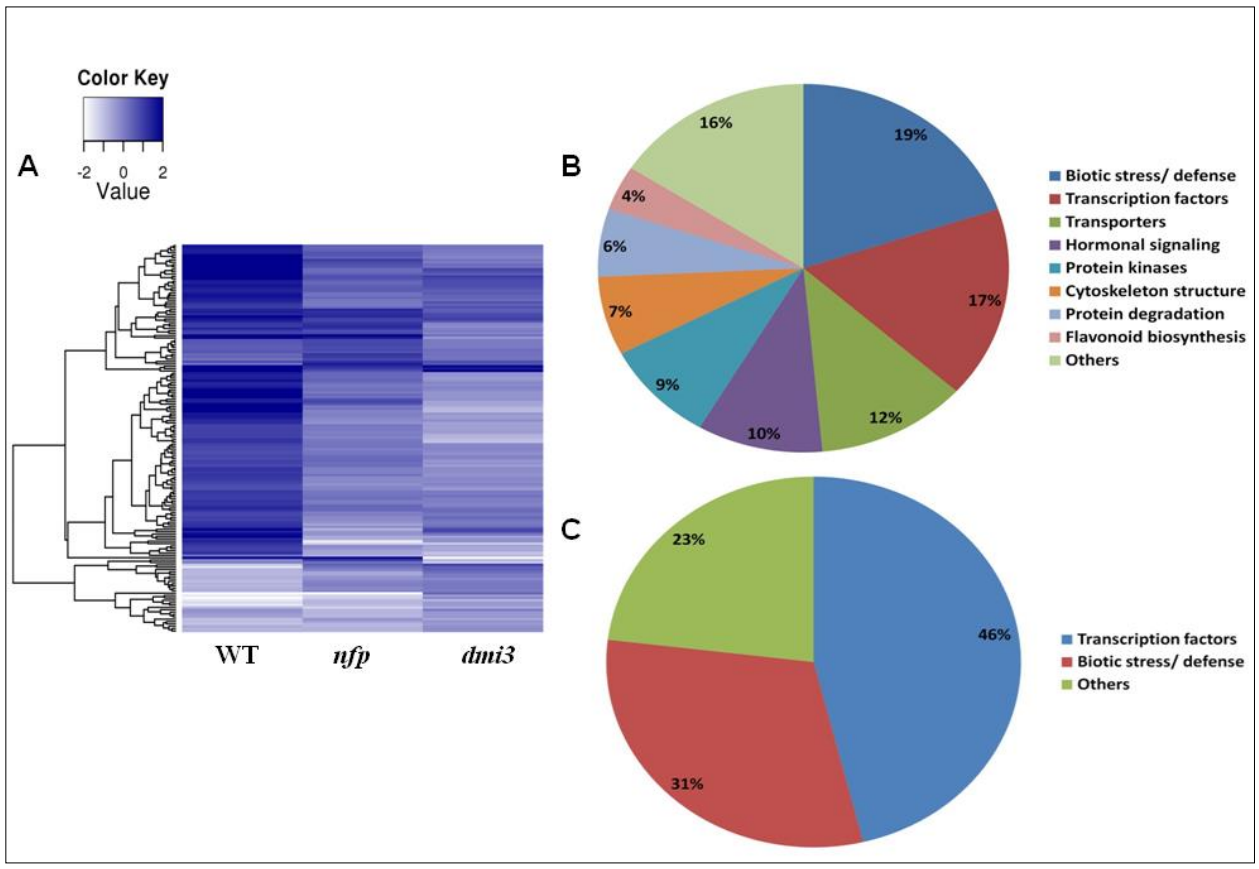


Figure 5. NF-induced transcriptional dynamics within 1 hr after NF-treatment in Medicago (A) Heat map of differentially regulated candidates in wild-type plants, *nfp* and *dmi3* mutants identified in the transcriptomic data (B) Functional group of genes which are differentially regulated in the roots of wild-type plants within one hour of NF-treatment. The major group of genes which were differentially regulated comprise Biotic stress/defense, transcriptional factors, transporters, hormonal signaling, protein kinases etc. (C) Functional group of genes which are differentially regulated in the roots of *nfp* mutants within one hour of NF-treatment. The two major group of genes which were differentially regulated in *nfp* mutant include transcription factors and genes related to biotic stress/ defense response.

transcription factor 1A (*Medtr4g100380*), tubulin beta chain (*Medtr8g107250*), nuclear transcription factor Y subunit B-3 (*Medtr8g091720*), pectinestarase (contig_73581), inorganic phosphate transporter 1, *MtPT1* (*Medtr1g043220*), and TIR-NBS disease resistance protein (contig_72999) displayed the direction of changes (up- or down-regulated) similar to RNA sequencing measurements.

NF-induced transcriptional changes in *nfp* indicate the existence of cryptic NF receptors

To our knowledge, no molecular or cellular response to NF has been observed in the *nfp* mutant and NFP was believed to be absolutely necessary for all the responses to NF in Medicago. However, our quantitative phosphoproteomics clearly revealed some responses in the *nfp* mutant which prompted us to test this result by RNA sequencing. NF-induced transcriptional changes observed in wild-type plants were largely reduced but not abolished in the *nfp* mutant (**Figure 5A**). While 136 genes were differentially regulated in wild-type, 31 genes were differentially regulated in *nfp*, of which 20 were upregulated. The two major groups of differentially-regulated genes in *nfp* include transcription factors and defense-related genes (**Figure 5C**). Among the transcription factors, ethylene-responsive transcription factors (*contig_61090*, *contig_82104*, *Medtr4g100380* and *Medtr1g040430*), bZIP transcription factor (*AC146721_1015*) and bHLH transcription factor (*contig_14899*) were significantly upregulated. Defense-related genes coding for TIR-NBS disease resistance protein (*contig_72999*), disease resistance protein RPS4 (*contig_71516*) and glutaredoxin (*contig_239949*) were also upregulated, while a gene encoding a Kunitz-type trypsin inhibitor alpha chain probably involved in defense (*Medtr6g065460*) was down-regulated. Since these NF-induced transcriptional changes in *nfp* were unexpected, we validated the expression of several candidate genes by qRT-PCR. Our results for *Ctg_82104*, *Ctg_72999*, *Medtr4g100380* and *Medtr4g112440* confirm regulation in wild-type and *nfp* observed by RNA sequencing upon NF treatment. These results confirm that NFP is responsible for transducing the majority of the NF signal, but also reveal clearly the existence of other NFP-independent NF receptor complexes.

Given the structural similarities between Nod and Myc factors, we speculated that some of NF signal may be transduced by the mycorrhizal receptors. Therefore, we explored the transcriptional changes

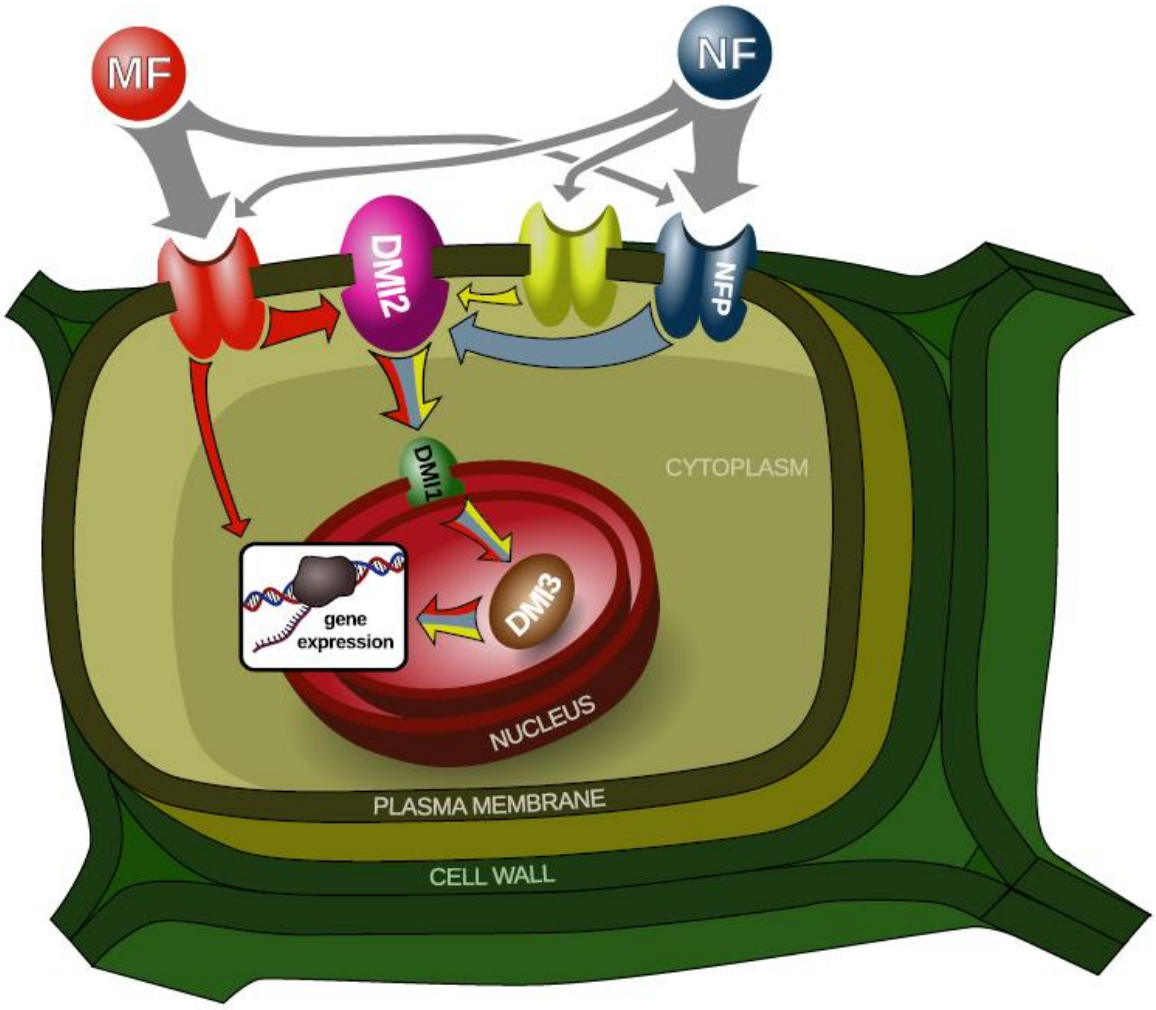


Figure 6. Proposed model depicting putative LCO-receptor kinases on the plasma membrane and the cross-talks between different symbiotic signaling pathways in Medicago.
NF are perceived with high stringency by NF receptor (NFP) and its putative heteromultimeric partner, an NFR1 homolog. This accounts for majority of signal transduction events. However, NF-induced, but NFP–independent signaling responses suggest the existence of other receptor complexes. Differential transcriptional regulation of these candidate genes by both NF and GSE in NFP-independent manner suggests the existence of Myc factor receptor, which is capable of perceiving NF. Interestingly, NFP-independent, but NF-specific transcriptional responses suggest the existence of NF-specific 'cryptic' signaling receptor in Medicago. While majority of these transcriptional responses are dependent on DMI3, GSE-induced *DMI3*-independent transcriptional regulation of candidate genes, suggest the existence of parallel signaling pathway, which is independent of well established signaling pathway comprising *DMI* genes.

of genes that were regulated in *nfp* upon treatment with mycorrhizal signals (GSE). C*ontig_82104* and *Medtr4g100380* were upregulated in wild-type and *nfp* plants upon GSE treatment, indicating NFP-independent gene activation by the mycorrhizal pathway. However, GSE treatment did not result in up-regulation of *contig_72999*, which was upregulated in *nfp* by NF-treatment in our RNA sequencing and qRT-PCR data. This indicates that some of the NFP-independent signal is not transduced by the Myc factor receptor, suggesting the existence of a third Myc factor- and NFP-independent NF receptor (**Figure 6**).

LYR1, an NFP homolog, upregulated during AM symbiosis and has been suggested as potential candidate for Myc factor perception.[8,26,54] We took a candidate reverse genetics approach to explore the possibility that LYR1 could mediate the transcriptional responses in *nfp* upon NF treatment. Utilizing RNA interference, we silenced LYR1 in wild-type and *nfp* plants as described previously.[55,56] Transgenic roots containing the RNA interference construct were analyzed for transcriptional changes one hour post-NF treatment by qRT-PCR. Silencing LYR1 in wild-type and *nfp* plants did not abolish NF-induced up-regulation of c*ontig_82104* transcript, indicating that differential regulation of this gene in response to NF is not mediated by LYR1. However, a negative result in these experiments may be misleading; expression analyses in a total *lyr1* knock-out might provide more conclusive evidence of its potential role in NF- and/or Myc factor-induced transcriptional changes in Medicago.

DMI3 is a key regulator of NF-induced transcriptional changes in Medicago

Downstream of plasma membrane NF receptors, nuclear events play are essential for the transduction of NF signals. The nuclear envelope is the main calcium store for NF-induced calcium spiking and several key proteins (e.g., DMI1 and MCA8) are localized on nuclear membranes.[13] DMI3 decodes calcium spiking signatures through its EF hands and transduces the signal downstream, probably by phosphorylation of specific targets.[15,57] We monitored NF-induced transcriptional changes in the *dmi3* mutant. In contrast to the significant transcriptome changes observed in wild-type plants, the *dmi3* mutant exhibited almost no change one hour post-NF treatment, confirming its central role in controlling NF-dependent transcriptional events (**Figure 5A**). Interestingly, the only genes showing a significant regulation

in the *dmi3* mutant encode well-characterized inorganic phosphate transporters. Medicago phosphate transporters *MtPT1*, *MtPT2* and *MtPT3* share ~96% sequence identify at the nucleic acid level and their encoded proteins share 98% amino acid sequence identity,[58-59] making them impossible to distinguish by RNA sequencing. Our data indicate a significant down-regulation of these genes. A similar down-regulation of these genes and proteins was observed during AM symbiosis, suggesting that the NF-induced down-regulation may result from AM pathway activation.[58,60,61] Using qRT-PCR, we confirmed that GSE treatment decreases *Medtr1g043220* (*MtPT1*) expression similar to that observed with NF treatment. We validated this NF- and GSE-induced down-regulation of *Medtr1g043220* in wild-type and *nfp* as well. Our data indicate that this event is DMI3- and NFP-independent, confirming this regulation as a probable consequence of AM pathway activation.

Data-sharing *via* the Medicago-Omics Repository (MORE)

The variety of experimental conditions used enables multiple levels of analysis, many of which escape the current scope of this article. To allow the research community to explore data not mentioned here, we created the Medicago-Omics Repository (MORE; http://coongrp.ad.biotech.wisc.edu/more/), a web-based resource containing transcriptomic, proteomic, and phosphoproteomic data collected during the course of this project. MORE allows users to view quantitative information for transcript, protein and post-translational modifications from multiple experiments simultaneously. Researchers are able to search for genes of interest and determine if gene expression, protein expression or post-translational modification levels are altered due to sub-cellular location, time or growth conditions. MORE is also dynamic in that users will be assisted in uploading their data to the existing database. To allow offline analysis, all entries into the database are freely available for download.

Discussion

Combining transcriptomics, proteomics and phosphoproteomics approaches we have performed a comprehensive systems analysis of NF-induced symbiotic signaling response of legume roots. The collection of phosphoproteomic analyses presented here is a database representing high-throughput and large scale mass spectrometric experiments requiring many hours of instrument time on a high resolution tandem mass spectrometer (ESI-LTQ-orbitrap mass spectrometer). In the past, many proteomic studies in plants have utilized off-gel mass spectrometry analysis,[62-67] and large-scale shotgun (untargeted) proteomic and phosphoproteomic technologies used shotgun mass spectrometry analysis to quantify phosphoproteome changes in response to pathogen elicitors in *Arabidopsis thaliana* (Arabidopsis),[68-72] yielding 1,172 measurements. In 2010, we described the identification of 3,457 phosphopeptides in Medicago roots,[37] as well as the quantification of 4,675 phosphopeptides in Arabidopsis.[73] Prior to the results reported here, the largest phosphoproteomic study in plants identified 6,919 phosphopeptides from rice.[74] Our current report presents 15,335 unique phosphopeptides and demonstrates the potential value of large scale quantitative phosphoproteomic analysis in plants.

While transcriptomic approaches have been used in the past to identify potential candidates regulating legume-rhizobia symbiosis, most of these studies explored transcriptional regulations occurring during later stages in the legume-rhizobia interaction, *i.e.* infection, nodule development and symbiotic nitrogen fixation spanning 1-32 days after inoculation with rhizobia.[31,32,75-77] In contrast, our study sheds light on the early transcriptional dynamics which occur with-in the first hour of NF perception. The genes identified in our study play a role very early in the nodulation symbiotic signaling and may regulate the initial NF-induced cellular responses in host plants. Our current study utilized the deep coverage of next generation RNA sequencing to identify the Medicago transcripts regulated rapidly in response to NF. In addition, we took advantage of Medicago genetics by using mutants affected in NF perception (*nfp*) and transduction (*dmi3*), RNA interference in transgenic roots (*LYR1*) and additional symbiotic signals (GSE)

in order to identify the different pathways regulating transcription in response to NF and their relative importance.

This large dataset allowed for the identification of many candidates which may control the early molecular and cellular responses observed in response to NF and, ultimately, the establishment of legume nodulation. Our transcriptomics and phosphoproteomic data suggested the presence of at least two NF receptors independent of NFP as well as molecular markers that allow us to track down the corresponding pathways (**Figure 6**). One of them is a receptor that also transduces mycorrhizal signals (GSE) but is probably not LYR1, and the other one is so far completely unknown. Our quantitative data also allowed for estimating the relative contribution of these pathways. NFP-dependent responses to NF represent about 80% of transcriptomic changes observed in wild-type plants, confirming that NFP is still the major signaling receptor in Medicago and that the other receptors play a less important role in NF perception (**Figure 6**). Our integrated approach also helped us identify proteins potentially controlling downstream stages of NF signal transduction. Many differentially regulated transcripts, and proteins which are phosphorylated in response to NF belong to families involved in cell signaling, cell cycle, cell morphology, defense reactions, hormone signaling and cell wall remodeling. Many differentially regulated transcripts, and proteins which are phosphorylated in response to NF in the current study belong to families involved in cell signaling, cell cycle, cell morphology, defense reactions, hormone signaling and cell wall remodeling.

Symbiotic signaling and cell cycle regulation

Protein kinases

Receptor-like kinases play a major role in the perception and perpetuation of external environmental signals. Several receptor-like kinases are required for NF perception (NFP and LYK3) and signal transduction (DMI2/NORK). We identified sites differently phosphorylated on several receptor-like kinases (Medtr2g098910, Medtr5g085790, Medtr4g115630, Medtr3g076990, Medtr4g113100, and Medtr5g075630). Medtr2g098910 is, for instance, a close homolog of Arabidopsis ACR4, which is thought

to control the initiation of cell divisions in the pericycle, for lateral root development and in the root tip meristem.[78] This is a very interesting candidate for the control of NF-induced cell divisions in the cortex and pericycle which give rise to the nodule primordium. Cell wall-associated receptor kinases (WAKs) belong to another group of plant receptor-like kinases, which play important roles in cell expansion, plant defense against pathogens, and tolerance to abiotic stresses.[79-80] WAKs physically link the plasma membrane to the cell wall matrix through the extracellular domain and mediate cellular events through their cytoplasmic kinase domain, hence acting as one of the most likely candidates participating in cell wall-cytoplasm signaling in plant response to external stimuli. In response to NF, WAK-like 8 (WAKL8, Medtr5g085790) seem to be differentially regulated at a significant level suggesting its potential role in NF signaling in legumes.

As mentioned previously, DMI3 is a calcium- / calmodulin-dependent protein kinase (CCaMK) that probably senses the nuclear calcium spikes through its EF hands and transduces this signal into phosphorylation changes in other nuclear proteins, such as transcription factors.[15,57] DMI3 was found to interact with and phosphorylate a nuclear protein with predicted coiled-coil domains called IPD3/CYCLOPS.[16,81,82] We found a serine residue (s50) in IPD3 to have its phosphorylation level increased *in vivo* in the presence of NF, suggesting that this residue might be a direct target of DMI3 phosphorylation. This residue falls within the region of IPD3 identified to be phosphorylated *in vitro* by DMI3.[81] We had already highlighted this residue on IPD3 in our previous non-quantitative study surveying the Medicago phosphoproteome.[37] We also identified a calcium-dependent protein kinase (CDPK, Medtr1g101630) and a calcium binding protein (Medtr8g107110) with sites differentially phosphorylated in response to NF. CDPKs have been identified as playing important roles in Medicago nodulation and root development.[83-84]

Mitogen-activated protein kinases (MAPKs) are involved in many aspects of plant development, in the response of plants to changes in their environment and in particular to pathogens.[85-86] MAPKs have been implicated in ethylene signaling[87] and ethylene is a potent inhibitor of NF signaling and, in particular, of calcium spiking.[88-89] More recently, MAPKs have been shown to be activated in response to exudates (Myc

factors) of AM fungi.[90] In our dataset, MAPK-like proteins (Medtr8g093730 and Medtr5g094390) are significantly regulated in response to NF.

SNF1-related proteins, also called AMP-activated protein kinases (AMPK), are master regulators of energy metabolism and one of them shows a significantly increased level of phosphorylation in response to NF (Medtr6g0129900). These proteins have been shown, in particular, to inhibit the activity of 3-hydroxy-3-methylglutaryl-CoA reductases (HMGRs) by phosphorylation.[91-92] An HMGR was found to interact with DMI2/NORK and is required for early NF signaling[55] although the function of this interaction is still unclear. In a separate study with Arabidopsis, we have also recently identified SNF1-related proteins whose phosphorylation levels are altered in response to abscisic acid (ABA).[73] Since ABA is a major regulator of NF signaling, nodule development and nitrogen fixation,[94-95] our observation suggests that this protein may be involved in the regulation of NF signal transduction by ABA. Several other protein kinases have also been found to be differentially phosphorylated in response to NF (Medtr4g128650, Medtr5g088400 and Medtr4g078290) but more targeted work is required to interpret these changes in the context of the NF signaling pathway.

Protein phosphatases

Dephosphorylation of proteins plays essential roles in plant signaling pathways including ABA and auxin signaling but also in symbiotic signaling.[73,96] Several protein phosphatases 2C (PP2C) isoforms are themselves significantly phosphorylated (Medtr5g080680) or dephosphorylated (Medtr6g086970) after addition of NF. The expression of the dominant negative allele of a PP2C (*abi1-1*) from Arabidopsis in Medicago roots dominantly suppressed ABA signaling and enhanced nodulation.[94] In *L. japonicus*, the expression of PP2C genes are induced in nodules and may play important roles at early and late stages of nodule development.[96] Given the relationships between PP2C and ABA signaling, these PP2C proteins may also be involved in the crosstalk between ABA and NF signaling.

Transcription and translation factors

Several transcription factors have been shown to be regulated by NF. As mentioned previously, NSP1 and NSP2 are transcription factors of the GRAS family that are required for NF signaling. They interact together and bind to the promoters of early nodulin genes such as *ENOD11*, *NIN*, and *ERN1*.[19-21] ERN1 itself is an AP2-ERF transcription factor that also binds to a different region of the *ENOD11* promoter along with two other AP2-ERF transcription factors, ERN2 and ERN3.[52-53] Our phosphoproteomic study identified several transcription factors with significant increase (Medtr8g077920, Medtr3g089910, AC233675_22.1, Medtr5g038620, ABN08601.1, Medtr2g086140, Medtr1g023690) or decreased (Medtr2g060650, Medtr7g100790, AC234952_22.1) phosphorylation levels. Several of these transcription factors (Medtr5g038620, Medtr7g100790) contain IQ domains, which are reported to bind calmodulin and are involved in calcium signaling.[97] These proteins may thus be activated in response to the calcium influx at the plasma membrane level or by the nuclear calcium spiking. The regulation of Medtr7g100790 by phosphorylation seems relatively complex since the same protein possesses sites more (s301) and less (s411, s413) phosphorylated in response to NF. Another of these transcription factors is a close homolog of Arabidopsis CIP7. This nuclear protein is an interactor of COP1 (Constitutive Photomorphogeneic 1) that possesses transcriptional activation activity without any obvious DNA binding motif and acts as positive regulator of light-regulated genes. CIP7 could provide a molecular basis for the inhibition of nodulation by light.[98]

Similarly, after the perception of NF, several transcriptional regulators were differentially regulated at the transcript level in the root tissues. In addition to *NIN*, *ERN1* and other ethylene responsive transcription factors mentioned earlier, genes encoding Myb family transcription factors (*Medtr1g087540, contig_103831*), bHLH transcription factors (*Medtr4g087920, Medtr4g097920, Medtr5g014600*), a NAC domain containing transcription factor (*Medtr4g035590*) and a bZIP transcription factor (*Medtr3g117120*) show significant differential regulation at the gene expression level. Two other bZIP transcription factors (MtATB2 in Medicago and ASTRAY in Lotus) have been found to play an important role in nodulation but probably at later stages of nodule development and senescence.[99-100]

Proteins involved in translation are also differentially phosphorylated (Medtr7g082940, Medtr6g021800, Medtr3g109550) in response to NF. For instance, Medtr7g082940 is a eukaryotic translation initiation factor 5 (eIF5) and possesses five phosphorylation sites (s207,s213, s431, s434 and s436), distributed in two different tryptic peptides, that are consistently less phosphorylated in presence of NF. In yeast, mammalian and plant cells, the casein kinase 2 (CK2) controls cell cycle progression. This protein was shown to phosphorylate *in vitro* wheat and Arabidopsis eIF2a, eIF2b, eIF5, and wheat eIF3c.[101-102] Phosphosites s207, s431, s434 and s436 on Medicago eIF5 (Medtr7g082940) are conserved in Arabidopsis and wheat and these four sites correspond to those phosphorylated by CK2 *in vitro*.[102] These phosphorylation sites on eIF5 are highly conserved from monocots to eudicots but are absent in yeast and mammalian proteins suggesting that they could be a plant-specific innovation. Interestingly, the fifth Medicago site (s213) is also conserved in Arabidopsis and wheat but this site was not found in the *in vitro* studies. All of this suggests that the reduction of phosphorylation on eIF5 that we observed might be related to the regulation of cell cycle. NF have been shown to regulate the cell cycle for the formation of both pre-infection threads and nodule primordium.[103-104] Our phosphoproteomic data suggest that eIF5 could be involved in these processes. Other phosphorylation sites found on eIF2ab and eIF3c *in vitro* by were also quantified in our study but did not change significantly in response to NF treatment.[102]

Other protein post-translational modifications

Several proteins involved in protein ubiquitination have been identified recently in investigations of early and late stages of legume nodulation.[105] Our study identified a RING finger RHF2A-like E3 ubiquitin-protein ligase (Medtr2g087820) with significantly decreased levels of phosphorylation. We also identified a RING finger SIZ1-like E3 SUMO-protein ligase with decreased levels of phosphorylation on a specific site. Nothing is known yet about the role of sumoylation in symbiotic signaling. These two types of protein modifications may be regulating the degradation, stability or sub-cellular localization of proteins involved in early responses to NF.[106] However, our observation of little or no changes in protein levels of ca. 8,000 proteins within the first hour suggests that either the ubiquitin/sumoyl-mediated protease systems

do not become significantly stimulated until later in the response, or that the affected proteins are not within the 8,000 proteins we measured (e.g. lower abundance transcription factors). At the transcript level, several members involved in protein degradation, such as RING finger family protein (contig_164276), F-box family protein (contig_48455 and contig_13936), U-box domain containing protein (Medtr5g083030) and several proteinases were also upregulated by NF perception. Hence, these differentially regulated members involved in proteolytic activities may be involved in early NF signaling.

Cell growth and root hair deformations

Plant root hairs exhibit a characteristic polarized cell growth similar to the one observed in pollen tubes and fungal hyphae.[107] Upon perception of NF, legume root hairs undergo a rapid calcium influx, an alkalinization of the cytoplasm, and a transient disorganization of the cytoskeleton followed by a growth re-orientation leading to characteristic root hair deformations.[4,108-110] When NF are applied locally on the root hair surface or by rhizobia, these deformations lead to a root hair curl that entraps the rhizobia.[7] All of these mechanisms require a very dynamic reorganization of the cytoskeleton and vesicle trafficking. Many proteins involved in these processes displayed differential phosphorylation upon NF treatment. As mentioned previously, Dynamin 2B (DRP2B, Medtr4g030140) was found to present interesting changes in phosphorylation patterns (**Figure 3E**) and probably plays a role in these vesicle trafficking processes.

In addition, a P-type plasma membrane proton pump ATPase (Medtr2g036650) shows a significant increase in phosphorylation level suggesting that this protein could be responsible for the observed alkalinization of the cytoplasm,[108] especially since this was the only one of twelve members of the proton pump gene family whose protein was affected in the first hour. Another protein whose phosphorylation status was changed in the first hour (Medtr1g044570.1) is a close homolog or Arabidopsis CAX1-interacting protein 4. This nuclear protein was shown to regulate the activity of the CAX1 H^+/Ca^{2+} antiporter.[111] These proteins could be involved in the proton and calcium fluxes observed in the early response to NF.[4,109] Phosphorylation sites on a phosphoinositide phospholipase C (PLC) (Medtr3g070720) are significantly increased in presence of NF. PLCs catalyze the hydrolysis of phosphatidylinositol 4,5-

bisphosphate in a calcium-dependent manner. The role of PLC activity was clearly shown in early responses to NF.[112] For instance, PLC antagonists, neomycin and the aminosteroid PLC inhibitor U73122 inhibited NF-induced *ENOD12* expression in Medicago. Medtr3g070720 seems therefore an excellent candidate to be involved in this NF-dependent pathway.

Small G proteins (guanine nucleotide-binding proteins) are known to be involved in the regulation of the phosphoinositide pathway, vesicle trafficking, root hair growth, and the early steps of legume nodulation.[56,107,113-114] The role of heterotrimeric G proteins in early symbiotic signaling was also proposed. In mammalian systems, Mastoparan is a well-characterized agonist of heterotrimeric G proteins. In Medicago, Mastoparan and its synthetic analog Mas7 was shown to induce calcium spiking and early nodulin expression.[115] Even if the targets of this drug in plants remain unclear, a role for heterotrimeric G proteins in regulating calcium spiking remain a possibility. We identified an increased phosphorylation on a site of the alpha-2 subunit of heterotrimeric G protein (Medtr5g018510). The activity of these small and heteromeric G proteins is regulated by guanine nucleotide exchange factors (GEFs), GTPase-Activating Proteins (GAPs), guanine nucleotide dissociation inhibitors (GDIs), and other interacting proteins. We identified differential phosphorylation on many GEF and GAP proteins such as the RopGEF SPIKE1, a homolog of BIG1/2 (Brefeldin A-inhibited GEF) (Medtr4g124430), a GEF homologous to DOCK7 (Dedicator of cytokinesis 7) (Medtr8g056900) as mentioned previously (**Figure 3F**), a putative RabGEF (Medtr3g073360), a putative RasGAP (Medtr2g037930), and a RasGAP-binding protein (Medtr4g083150). A protein potentially interacting with these small GTPases (Medtr5g007350) displays a decrease of phosphorylation in presence of NF.

Proteins probably involved in F-actin polymerization and belonging to the SCAR/WAVE (suppressor of cyclic AMP receptor/Wiskott-Aldrich syndrome Verprolin-homologous protein) complex were also regulated by phosphorylation (Medtr8g086300 and Medtr7g071440). This complex is known to regulate polar growth but also infection thread formation during later stages of this symbiosis.[116-117] Phosphorylation of proteins from this complex as well as interactions with small GTPases and their

regulators regulates actin polymerization.[118] A protein from the gelsolin superfamily and similar to villins (Medtr7g091460) displays a significant increase in phosphorylation level. Villins have been reported to regulate the organization of the actin cytoskeleton, cytoplasmic streaming and bundling of actin filaments in root hairs.[119] Activation of heterotrimeric and small G proteins dissociate these proteins from the barbed end of actin filaments.[120] Medtr7g091460 may therefore be a key player connecting G protein activity to cytoskeleton reorganization in response to NF.

Proteins involved in vesicle trafficking were found to have their phosphorylation levels change, such as a vesicle-fusing ATPase (AC233659_1.1), a SNARE-interacting protein KEULE (Medtr4g102120), an AP-3 complex subunit delta (Medtr8g104380), a neurobeachin-like protein (Medtr7g075660), synaptogamins (Medtr8g035590 and Medtr1g094810), and a component of the exocyst complex (Medtr4g103540). Synaptogamins have been reported to be phosphorylated by the cell cycle regulator casein kinase 2 (CK2) like eIF5.[121] All of these observations support the presence of an extensive remodeling of cell cytoskeleton and vesicle trafficking in response to NF and mediated by phosphorylation.

Defense reactions and plant hormones

The establishment of symbiotic associations requires a tight control of plant defense reactions. Similarities exist between responses to NF and responses to pathogen elicitors including the production of reactive oxygen species or proteins closely related to defense proteins.[122-123] However, the kinetic and the intensity of these responses to NF are different from those observed in response to pathogens.[124] Two lipoxygenases (LOX, Medtr8g018520 and Medtr2g099570) and a putative Kunitz-type trypsin inhibitor (Medtr6g059730) display decreased phosphorylation levels in response to NF. Using antibodies, LOX have been detected in the nodule cortex and in the cytosol of uninfected nodule cells of the central tissue, but were absent in infected nodule cells and in vascular tissues.[125] These proteins could be an indication of an early control of defense reactions mediated by NF. These lipoxygenases could also be related to jasmonic

acid (JA) synthesis. JA is a potent inhibitor of responses to NF including calcium spiking and nodulin gene expression and its production might be decreased to ensure the transduction of NF signals. In general, however, little is known about the role of these molecules during early NF signaling.[126]

Similarly, stress and defense-related genes were also differentially regulated, at the transcript level in the first hour of NF-perception. In line with our observations,[28] reported the differential regulation of stress/defense-related genes one hour post inoculation with rhizobia. Such defense/disease-responses may be mediated not only by NF released by the rhizobia, but also by other bacterial signals. It is interesting to observe such defense-related responses when purified NF were applied at 10^{-8} M, a concentration which is known to activate symbiotic signaling but not an obvious defense reaction in legumes. Pathogen elicitors-induced defense responses and NF-induced symbiotic responses share certain common features, such as production of reactive oxygen species, nitric oxide production, and associated activation of redox balance machinery.[124] In our transcriptomic study, a gene encoding a peroxidase (*Medtr4g133800*) is induced by NF but is distinct from the well-characterized *Rhizobium meliloti*-induced peroxidase (RIP; Peng et al., 1996). Among the defense-related genes, the upregulation in the expression levels of salicylic acid carboxyl methyltransferase (SAMT, *contig_54253* and *contig_20507*) are striking (16.61 and 14.41 fold increase respectively). SAMT catalyzes the formation of methylsalicylate, a major defense signal in plants. Similarly, genes encoding a TIR-NBS disease resistance-like protein (*Medtr5g037700*), pathogenesis-related protein 1 (PR1) (*Medtr4g128750*) and a disease resistance response protein (*Medtr7g021300*) were induced rapidly after NF perception. Induction of *MtN1* (a homolog of cystein-rich pathogen inducible protein in pea) and *PR10* (*MtN13*) have already been reported during early infection and root nodule development in Medicago but these are later stages of legume nodulation.[28,122] Our data indicate that application of NF can trigger very rapid defense-like responses in the roots. These results highlight, once again, the similarities between responses to pathogens and symbionts.[34]

Another major group of genes which are differentially regulated by NF-treatment include genes which are involved in hormone biosynthesis or hormonal responses. Auxin transport and auxin/cytokinin

ratio play a major role in NF signaling and nodule formation.[127] A gene encoding an auxin responsive SAUR protein (*Medtr4g124700*) is significantly upregulated by NF. Similarly, genes encoding a cytokinin receptor histidine kinase (*Medtr1g013360*) and cytokinin-O-glucosyltransferases (*Medtr7g070740* and *Medtr5g072860*) were also upregulated by NF, confirming the involvement of cytokinin perception and transport during nodule organogenesis. In Medicago, RNA interference (RNAi) based-silencing of another cytokinin receptor MtCRE1 resulted in a strong reduction in rhizobia infection and nodule primordial formation.[128] As mentioned previously, ethylene is a well-known negative regulator of symbiotic signaling and in particular, of calcium spiking.[89] Medicago ethylene-insensitive *skl* mutant displays a 10-fold increase in nodule number.[88] Several genes encoding 1-aminocyclopropane-1-carboxylate oxidases (*contig_54169*, *contig_90027*, *contig_62339*, and *contig_74946*) which catalyze the final step of ethylene biosynthesis were differentially regulated. Similarly, the role of several ethylene responsive transcription factors (ERN1, 2, 3 and EFD) in early NF signaling is also well-established and in our data *ERN1* and several other ethylene responsive transcription factors were significantly upregulated. Altogether our data confirm that the regulation of ethylene signaling is critical during the early hours of NF signaling.

Cell wall remodeling

Cell wall degrading enzymes secreted by pathogens often act as pathogenesis factors in disease establishment. A similar strategy is employed, in the legume-rhizobia symbiosis, to facilitate bacterial entry. Cell wall degrading enzymes produced by the rhizobia play crucial roles in facilitating the entry of rhizobia through root hairs but also possibly infection thread progression and the release of bacteria into symbiosomes.[129] In contrary to plant defense reactions, the plant itself loosens its cell wall in preparation for bacterial entry.[130] Application of NF to legume root hairs induce a temporary swelling and it was shown that, during that phase, the cell wall exhibits a mottled aspect similar to the one observed in epidermal cells during root hair bulging.[6] This indicates a transient relaxation of the cell wall induced by NF, even in the absence of rhizobia. In Medicago, expression of early nodulins, *ENOD11* and *ENOD12* is induced within 3-6 hours following rhizobia inoculation.[35] These genes are among the best characterized markers for the

activation of the NF signaling pathway and were used as controls in our current study. *ENOD11* and *ENOD12* encode hydroxyproline-rich glycoproteins (HyPRPs) with relatively few tyrosine residues. The expression of *ENOD11* and *ENOD12* probably results in enhanced cell wall plasticity or in components of the infection thread matrix.[35] Recently, a pectate lyase was shown to play a major role in root infection by the rhizobia. *L. japonicus Nodulation Pectate Lyase* (*LjNPL*) is induced in roots and root hairs by NF *via* the NF signaling pathway and the NIN transcription factor. In our study, a close homolog of *MtNPL* (*Medtr3g086310*) was upregulated within one hour of NF perception. Other cell wall degrading enzymes such as polygalacturonases (contig_67431 and contig_83719) and pectinestarases (Medtr3g033690, contig_73581, Medtr3g010770, contig_51305) were also upregulated. Hence, very rapidly and even before contact, legume root cells are paving the way for the accommodation of their symbiotic partner by remodeling the cell wall barrier. In conclusion, the combination of large-scale analyses of transcriptional, translational, and post-translational events has enabled a multi-faceted examination of cellular responses to NF within one hour of treatment.

In conclusion, the combination of large-scale analyses of transcriptional, translational, and post-translational events has enabled a multi-faceted examination of cellular responses to NF within one hour of treatment. Our transcriptomics data demonstrate that Medicago shows a limited but significant rapid response in changes of gene expression within one hour after NF-perception, and that symbiotic deficient mutants are lacking much of the NF-induced transcriptional dynamics, albeit not completely. The NF-induced transcriptional responses in *nfp*, combined with supporting evidence utilizing Myc factor signaling perturbation suggest the presence of additional signaling receptors in Medicago other than NFP. Large-scale quantitative proteomic analyses indicate that protein levels remain relatively unaltered after NF treatment, but protein phosphorylation is actively regulated in all three genotypes within the first hour of NF-perception, with many changes in phosphorylation events occurring as early as 10 minutes after NF treatment. This integration of multiple-‘omic’ approaches has thus yielded an extensive list of candidate genes implicated in the NF signaling cascade. Few of these targets have documented roles in NF signaling

pathway, while many novel candidates identified in our current study warrant targeted proteomic and biochemical validation. To enable the scientific community to explore further the data reported here, we have created the Medicago-Omics Repository (MORE; http://coongrp.ad.biotech.wisc.edu/more/), an online resource to obtain quantitative transcriptomic, proteomic, and phosphoproteomic data. To build a comprehensive resource, we invite researchers to upload quantitative and non-quantitative data relating to Medicago transcriptomics, proteomics, and phosphoproteomics. Hopefully, MORE will act as a central resource enabling researchers to query large scale data sets, facilitating future systems level studies in Medicago and other legumes.

Experimental Procedures

Plant materials, growth and treatment with NF and GSE

Medicago truncatula Jemalong A17 (wild-type), C31 (*nfp-1*) (9) and TRV25 (*dmi3-1*)[15] plants were used for transcriptomics, proteomics and phosphoproteomics analyses. Plants were grown under three different growth conditions (aeroponic, hydroponic and plates) which facilitate collection of large quantities of NF-responsive root tissues for RNA/protein extraction. The different growth conditions used in the current study were optimized for efficient response to NF with the help of p*ENOD11*:GUS plants.[35] Seeds of Medicago were harvested, acid-scarified, surface sterilized and germinated[36] before growing in different growth conditions. For growth in the aeroponic system, 3-day old seedlings were placed in the aeroponic system[37] and grown in nitrogen-free modified Fahraeus medium[36] for 14 to 15 d at 22°C and 24 h of 130 to 200 umol m^{-2} s^{-1} light. The roots, however, grew in the dark. For uniform treatment with NF, the modified Fahraeus medium without NF was replaced by the one with 10^{-8} M NF obtained from *SinoRhizobium meliloti* strain Rm1021 pRmE43 (pTE3:*nodD1*) as indicated.[36] One hour after treatment, the root tissues were harvested and flash frozen for phosphoprotein extraction. For growth in the hydroponic system, 2-day old Medicago seedlings were transferred to conical flasks containing liquid modified Fahraeus medium

supplemented with 1% sucrose and incubated in the dark at room temperature with constant shaking at 50 rotations per minute. The seedlings were treated with 10^{-8} M NF, incubated in the medium for one hour and the root tissues were flash frozen for phosphoprotein extraction. For growth in the plate system, 2-day old germinated seedlings were grown on 23 cm x 23 cm square plates containing Fahraeus medium overlaid with moist sterile germination paper. The seedlings were grown in the dark at room temperature for five days and treated with 10^{-8} M NF by flood inoculation to ensure that roots of all the seedlings grown on the plates are treated. One hour after the NF treatment, the root tissues were rapidly harvested and flash-frozen in liquid nitrogen. Proteins were isolated from whole-cell lysate of root tissue or membrane-enriched fractions for phosphoproteomics with the addition of a variety of phosphatase inhibitors as described previously.[37] For transcriptomics studies seedlings grown in the plate system were used.

GSE from *Glomus intraradices* was obtained as described previously.[24] Each Medicago seedling (wild-type plant, *nfp* and *dmi3*) was treated with GSE obtained from 250 spores dissolved in 1 ml of distilled-sterile water. As a mock inoculation, control seedlings were treated with 1 ml of sterile distilled water. One hour after the treatment, the root tissues were excised and flash-frozen for RNA extraction.

RNA extraction and TruSeq RNA sample preparation

Total RNA was isolated from the root tissues using QIAGEN RNeasy Plant mini kit. RNA samples were run on a Bioanalyzer 2100 RNA Pico chip to test the quality of the preparations. RNA from nine samples showing signs of degradation was re-isolated from the same plant material and used in place of the first samples. The final eighteen samples submitted for sequencing had RIN values ranging from 6.2–10.0 (mean 7.7, median 7.7). Sequencing libraries were prepared at the University of Wisconsin Gene Expression Center using the Illumina TruSeq RNA Sample Preparation Kit (mRNA protocol rev. A). Libraries were sequenced at the University of Wisconsin DNA Sequencing Facility on an Illumina HiSeq 2000 system. Samples were multiplexed on three lanes of the flow cell with one replicate of each genotype/treatment on each lane to minimize the effects of technical variation. Data was collected for 100 bp single-end reads.

Sequencing reads were checked for quality using FastQC v0.9.4. Reads were mapped to the publicly available Mt3.5 genome sequence using TopHat v1.3.2.[26,38] Minimum intron size, maximum intron size, and microexon search parameters were set to 10, 20000, and true, respectively, and default settings were used for all other parameters. Reads mapping to multiple locations on the genome were filtered out using samtools v0.1.18.[39] The remaining data were used to generate gene-level exon coverage counts based on the published Mt3.5v4 gene annotations using htseq-count from the HTSeq package v0.5.1p2 in non-stranded union mode. Gene-level read counts for the eighteen samples were normalized and analyzed for differential expression using the R package DESeq v1.6.0 with pooled dispersion estimates.[40] Test statistics were adjusted for multiple testing using the method of Benjamini and Hochberg and the adjusted p-values were used to select differentially expressed genes at a false discovery rate of 5%.[41]

Silencing Medicago *LYR1*

To silence Medicago *LYR1*, RNAi fragment of about 252 base pairs from *LYR1* coding sequences was amplified using 'ATGTCGACCAAAGGACATAATCACAGC' and 'TTGATATCCTCACAAGCCTTTCTCTTCC' primers and cloned into pENTR1A entry vector. The *LYR1*-RNAi fragment was cloned into pK7GWIWG2-R hairpin RNAi destination vector carrying *Ds*Red1 as visible reporter by LR-recombination. Medicago Jemalong A17 seedlings were transformed with *LYR1*-RNAi cassette by *Agrobacterium rhizogenes* (strain MSU440) -mediated hairy root transformation (Riely et al., 2011). Transgenic roots expressing *LYR1*-RNAi cassette were identified by *Ds*Red1 expression, and the validation of reduction in *LYR1* expression was confirmed by semi-quantitative RT-PCR using the primers 'GGCAGTGAATTGCATAGAAG' and 'CCATTTCAGCAACCTCTACA'.

Validation of RNA sequencing data by qRT-PCR

The validation of the expression level of candidate genes by qRT-PCR was performed with RNA samples prepared from three biological replicates each with 2 technical replicates. The RNA samples were

treated with DNAse, DNA-free™ (Ambion) to remove any residual DNA contamination present in the samples. RNA samples were quantified using NanoDrop 1000 for precise quantification. For cDNA synthesis, 0.5 µg of RNA was used from all the samples to standardize the RNA concentration. cDNA was synthesized using RevertAid™ First Strand cDNA Synthesis Kit (Fermentas). Quantitative RT-PCR was performed using SYBR® Advantage® qPCR Premix (Clontech) with respective primer set (given in a separate section) optimized to have an efficiency of 100 ± 5 %. The expression levels of candidate genes were normalized using two references genes, *MtActin* and *MtEF1a*. The qRT-PCR data were analyzed using GenEx software from Bio-Rad. Primers used for quantitative RT-PCR validation of RNA sequencing data are listed here: Ctg_72999 (GTATAGAACCTTCACACGTCCG and CTTTGCCTTGTTTCTTTTGTTTCG); Ctg_73581 (CAAATCGCCAATTCTACCGC and CCTTGTGCTGTTATTGTGTTCTG); Ctg_82104 (GTCACCAAAAGAAATTCGATACCG and GCGGTGAGAAATGTGCCTAG); Mt1g043220 (TCCAGCTGCATCACTGTTAG and GGGCTGTTTCTGGCATTTTC); Mt4g100380 (AAGAACGGAGCTAGAGTTTGG and ACCCTCTCATCCTATACGCAG); Mt4g112440 (AACAAGCTGGTACTATCCGC and AGCAGTGACCATATCAATCCC); Mt8g091720 (ACATCAGGAAACATCTCAGACAG and GAGACACATTCTTGAACCGTTTC); Mt8g107250 (AGGTTGAGAATTGTGATTGCTTG and TGGAGATGGAAAGACAGAGAAAG); *MtActin* (GCAGATGCTGAGGATATTAACCC and CGACCACTTGCATAGAGGGAGAGG); *MtEF1a* (GTCAAAACATGGTTGCTGCACAAGC and TTAGGTCACAAGGCAGATTGCAGG).

Microsome and Membrane Preparation

Plants were grown as described for RNA extraction with the exception that the tissue was harvested immediately after one hour of NF treatment with no freeze step. All subsequent stages of membrane isolation were carried out in a 4°C cold room with chilled equipment. Plant tissue was ground in 2x v/w (e.g. 20 ml buffer per 10 g fresh weight tissue) ice-cold Homogenization Buffer (230 mM sorbitol, 50 mM TrisHCl, 10 mM KCl, 3 mM EGTA, pH 7.5) containing freshly added protease inhibitors (1 mM PMSF,

0.7 μg/ml leupeptin, 1.0 μg/ml pepstatin, 1 mM potassium metabisulfite) and phosphatase inhibitors (10 mM NaF, 2 mM Na pyrophosphate, 1 mM ammonium molybdate). Homogenization was performed in a Waring blender using two 20-second pulses. Homogenate was filtered through four layers of Miracloth and centrifuged at 6000xG for 10 min at 4°C. The supernatant containing the microsome fraction was transferred to a cold ultracentrifuge tube and spun at 65,000xG for 30 min at 4°C. The supernatant was discarded and the microsomal pellet was resuspended in ice-cold Resuspension Buffer (250 mM sucrose, 10 mM KCl, 10 mM HEPES, pH to 7.0 with KOH) with the aid of a Potter teflon homogenizer. The microsomal fraction was either stored at - 80°C for subsequent analysis or further processed as follows to enrich for plasma membrane proteins.

One gram (approximated using 1.0 ml volume) of microsomal sample was added to a four gram (3 g pre-made + 1 g microsomal) 6% PEG/Dextran two-phase system prepared the previous day in disposable glass culture tubes and stored in the cold room overnight (final concentrations: 6% w/w Dextran T500, 6% w/w PEG-3350, 333 mM sucrose, 5 mM potassium phosphate, 5 mM potassium chloride, 0.1 mM EDTA). Dithiothreitol was added fresh immediately before use to a concentration of 1 mM. After addition of the microsomal sample, the two-phase system was mixed thoroughly by inversion and incubated at 4°C for 15 min followed by a 15 min spin at 900xG in a 4°C centrifuge. The upper phase (enriched for plasma membrane) was removed slowly using a Pasteur pipet, leaving the interface with the lower phase, and transferred to an ultracentrifuge tube. The lower phase was transferred to a separate ultracentrifuge tube. Samples were diluted to 24 ml using Resuspension Buffer and centrifuged at 100,000 ×g for 35 min. at 4°C. The supernatant was discarded and the pellets were suspended in Resuspension Buffer with the aid of a Potter teflon homogenizer. Protein concentrations in microsomal, upper, and lower phase fractions were quantified using a Bradford assay with a BSA standard curve.

Protein Extraction

For experiments analyzing whole plant extract, protein was precipitated from plant extract *via* chloroform/methanol extraction or acetone precipitation. Chloroform/methanol extraction was performed

by adding four volumes of methanol to one volume of plant extract. The mixture was vortexed and one volume of chloroform was added before additional vortexing. Three volumes of water was added, the solution was vortexed, and subsequently centrifuged for 5 min (4696 × g, 4°C). The top layer was removed and discarded *via* serological pipette. Then, three volumes of methanol were added and the solution was vortexed before centrifugation for 15 min (4696 × g, 4°C). The resulting pellet of protein was washed three times by vortexing with ice cold 80% acetone and centrifuging at 10,000 × g for 5 min. Once washed, the pellet was dried on ice for 30 min and stored at -80°C.

Acetone precipitation protein extraction consisted of adding four volumes of ice cold acetone to one volume of plant extract. The solution was then vortexed and stored at -20°C overnight. The sample was then centrifuged for 20 min (4696 × g, 4°C). The resulting protein pellet was washed three times with ice cold 80% acetone. Once washed, the pellet was dried on ice for 30 min and stored at -80°C.

Protein Digestion and Isobaric Labeling

For whole plant extract samples, protein pellets were resuspended in ice-cold 8 M urea, 30 mM NaCl, 40 mM tris (pH 8), 2 mM MgCl2, 50 mM b-glycero phosphate, 1 mM sodium orthovanadate, 10 mM sodium pyrophosphate, 1× mini EDTA-free protease inhibitor (Roche Diagnostics), and 1× phosSTOP phosphatase inhibitor (Roche Diagnostics). Total protein was then quantified using a BCA protein assay kit (Thermo Scientific Pierce). For analysis, 1 mg of protein from each sample was reduced by adding DTT to a final concentration of 5 mM, and alkylated with 15 mM iodoacetamide before final capping with 5 mM DTT. Digestion was carried out by adding LysC (Wako Chemicals) at a 1:100 enzyme-to-protein ratio and incubating at 37°C for 2 hours. At this time, the lysate was diluted with 25 mM tris (pH 8) to a final urea concentration of 1.5 M and further digested for 12 hours at 37°C with trypsin (Promega) at a 1:100 enzyme-to-protein ratio.

For membrane enriched samples, seven volumes of re-suspension buffer (from above) were added to one volume of membrane-enriched solution (membrane samples contained either 0.5 mg or 1 mg

protein). Samples were then reduced as described above. Digestion was carried out by adding LysC (Wako Chemicals) at a 1:120 enzyme-to-protein ratio and incubating at 37°C for 2 hours. At this time, an additional aliquot of LysC was added at a 1:240 enzyme-to-protein ratio before incubation for another 1.5 hours at 37°C. Finally, the lysate was diluted with 25 mM tris (pH 8) to a final urea concentration of 1.5 M and further digested for 12 hours at 37°C with trypsin (Promega) at a 1:100 enzyme-to-protein ratio.

For all experiments, peptides were then acidified with TFA to quench the reaction and de-salted using C-18 solid phase extraction (SPE) columns (Waters). TMT[48] or iTRAQ[49] labeling was carried out per manufacturer's directions (Thermo Scientific Pierce, TMT; Applied Biosystems, iTRAQ). To confirm BCA measurements, aliquots of total protein were taken from each sample, combined in a 1:1:1:1, 1:1:1:1:1:1, or 1:1:1:1:1:1:1:1 ratio (4-Plex, 6-Plex, and 8-Plex experiments, respectively), and analyzed *via* mass spectrometry. Summed reporter ion ratios from this experiment were used to adjust mixing ratios of the remaining labeled digests, after which the mixed samples were again de-salted using SPE.

Strong Cation Exchange Fractionation

Strong cation exchange (SCX) was carried out using polysulfoethylaspartamide column (9.4 × 200 mm; PolyLC) on a Surveyor LC quaternary pump (Thermo Scientific; flow rate of 3.0 ml/min.). The mixed and dried isobaric label sample was resuspended in buffer A before separation. The following gradient was used for separation: 0-2 min, 100% buffer A, 2-5 min, 0-15% buffer B, 5-35 min, 15-100% buffer B. Buffer B was held at 100% for 10 minutes. The column was washed extensively with buffer C and water prior to recalibration. Buffer compositions were as follows: buffer A [5 mM KH2PO4, 30% acetonitrile (pH 2.65)], buffer B [5 mM KH2PO4, 30% acetonitrile, 350 mM KCl (pH 2.65)], buffer C [50 mM KH2PO4, 500 mM KCl (pH 7.5)]. Typically 8 - 14 fractions were collected over a 50 min elution period. Samples were collected by hand, frozen, lyophilized, and desalted by SPE.

Phosphopeptide Sample Preparation

Following SCX fractionation, phosphopeptides were enriched *via* immobilized metal affinity chromatography (IMAC) using magnetic beads (Qiagen). Following equilibration with water, the beads were treated with 40 mM EDTA (pH 8.0) for 30 minutes with shaking, and washed three times with water again. The beads were then incubated with 100 mM FeCl3 for 30 minutes with shaking and finally were washed three times with 80% acetonitrile/0.1% TFA. Samples were likewise resuspended in 80% acetonitrile/0.15% TFA and incubated with beads for 45 minutes with shaking. The resultant mixture was washed three times with 1 ml 80% acetonitrile/0.1% TFA, and eluted using 1:1 acetonitrile: 0.7% NH4OH in water. Eluted phosphopeptides were acidified immediately with 4% formic acid and lyophilized to ~5 µL.

Nano-High Performance Liquid Chromatography

For all samples, online reverse-phase chromatography was performed using a NanoAcquity UPLC system (Waters). Peptides were loaded onto a pre-column (75 µm ID, packed with 7 cm Magic C18 particles, Bruker-Michrom) for 10 min at a flow rate of 1 µm/min. Samples were then eluted over an analytical column (50 µm ID, packed with 15 cm Magic C18 particles, Bruker-Michrom) using either a 90 or 120 min linear gradient from 2% to 35% acetonitrile with 0.2% formic acid and a flow rate of 300 nL/min.

Mass Spectrometry

All experiments were performed on an ETD-enabled LTQ Orbitrap Velos mass spectrometer (Thermo Fisher Scientific). High resolution MS1 scans in the orbitrap were used to guide data-dependent MS/MS scans that utilized electron transfer dissociation (ETD)[42], collisionally activated dissociation (CAD), or high-energy CAD (HCD)[43] to produce sequence informative ions analyzed in the orbitrap. In 5 of the 11 experiments, HCD alone was used to produce reporter tags, the remaining experiments utilized QuantMode (QM) MS/MS scans. QM is a recently described method that utilizes gas-phase purification to improve quantitative accuracy and dynamic range.[44]

The LTQ Orbitrap Velos used firmware version 2.6.0.1065 SP3. MS1 scans were performed in the Orbitrap at 30,000 or 60,000 resolution at a max injection time of 500 ms and a target value of 1,000,000. Dynamic exclusion duration was set to 30 or 60 seconds, with a max exclusion list of 500, and an exclusion width of 0.55 Th below and 2.55 Th above the selected average mass. HCD MS2 scans were also performed in the Orbitrap at a resolution of 7,500 and with HCD normalize collision energy (NCE) of 45 or 50%, a max inject time of 200 or 500 ms, and an AGC target of 50000. ETD MS2 scans were performed in the Orbitrap at a resolution of 7,500 and with a charge independent reaction time of 70 ms, a max inject time of 200 ms, an AGC target of 50000, and a reagent ion target of 400000. QuantMode (QM) scans were performed in the Orbitrap at a resolution of 7,500 and with a charge dependent PTR reaction time to enable gas-phase purification, HCD NCE of 65–80% to produce isobaric reporter tags, and CAD NCE of 35% to produce sequence informative ions.[44] For QM scans an AGC target of 400000, a reagent ion AGC target of 400000, and max injection time of 500 ms was used.

Database searching and FDR estimation

MS/MS data were analyzed using the Coon OMSSA Proteomics Software Suite (COMPASS).[45] The Open Mass Spectrometry Search Algorithm (OMSSA; version 2.1.8)[46] was used to search spectra against a concatenated target-decoy database consisting of Medicago protein sequences from the following databases: MT3.5, Uniprot, NCBI, and an internal Wisconsin Medicago Group database. This composite database contained 51,426 entries and is available for download at *more.Medicago.wisc.edu*. We have previously demonstrated a combined database returns a greater number of identifications.[37] For all searches, tryptic were created *in silico* allowing up to three missed cleavages. The precursor mass tolerance was set to ±4.5 Da and monoisotopic mass tolerance was set to ±0.015 Da for fragments ions. Carbamidomethylation of cysteines, isobaric label (TMT or iTRAQ) on the N-terminus, and isobaric label on lysines were included as a fixed modifications, while oxidation of methionines and isobaric label on tyrosines was added as a variable modification. Results from each experiment were then filtered to a 1% FDR using high resolution batch *FDROptimizer*.

Protein and Peptide Quantification and Phosphosite Localization

For each experiment, isobaric label quantification was performed using *TagQuant*. Peptides from all experiments were then combined into protein groups and quantified at the protein level using *ProteinHoarder*. For phosphopeptides, the *Phosphinator* software was used to localize phosphorylation sites and combine quantitative data for phosphoisoforms within experiments.[47] Isobaric reporter tag data from all eleven experiments were consolidated in Microsoft Excel. Experimental ratios and p-values (Student's t-test assuming equal variance) were then determined using Microsoft Excel.

Protein and Phosphoisoform Functional Analysis

To determine functional analysis enrichment, we developed a program, *MedicaGO*, which links gene ontology (GO) terms available from interPro to corresponding MT3.5 accessions. The GO terms are linked to unique gene identifies, however protein grouping can result in many gene identifiers being combined into a larger compilation of proteins. To account for this, we assigned GO terms from any gene identifier within a protein group to the group as a whole. Any GO term that was significantly represented in a subgroup as compared to the entire proteome/phosphoproteome was considered enriched ($P<0.05$, Fisher's exact test with Benjimini-Hochberg correction for multiple comparisons).

Building Medicago-Omics Repository (MORE)

All data presented here (transcriptomic, proteomic, phosphoproteomic) will be freely available on a newly created Medicago-Omics Repository (MORE; http://coongrp.ad.biotech.wisc.edu/more/). By entering Medicago gene identification number or protein name (e.g., Medtr4g030140 or MtDRP2B) users will obtain RNA, protein, and phosphorylation quantitative data. Overall wild-type, *nfp*, and *dmi3* ratios (+NF/-NF) are displayed for each protein entry. Expandable displays will show the measurements for each individual experiment for each protein and phosphorylation site, allowing users to compare the relative response between different experimental conditions. MORE also houses raw nHPLC-MS/MS data (Thermo

.RAW files), and downloadable spreadsheets including all RNA, protein, and phosphorylation data collected during the course of this experiment.

Acknowledgments

The authors thank Thomas W. Bryan for technical support and Kari L. Forshey for critical reading of the manuscript. This work was supported by a grant from the National Science Foundation (NSF#0701846) to M.R.S., J.J.C., and J.M.A. C.M.R was funded by an NSF Graduate Research Fellowship and NIH Traineeship (T32GM008505).

References

1. Hartwig, U. A., Maxwell, C. A., Joseph, C. M., and Phillips, D. A. (1990) *Chrysoeriol and luteolin released from alfalfa seeds induce nod genes in Rhizobium meliloti.* Plant Physiol. 92, 116-122

2. Dénarié, J., Debellé, F., and Promé, J. C. (1996) *Rhizobium lipo-chitooligosaccharide nodulation factors: Signaling molecules mediating recognition and morphogenesis.* Annu. Rev. Biochem. 65, 503-535

3. Ehrhardt, D. W., Wais, R., and Long, S. R. (1996) *Calcium spiking in plant root hairs responding to Rhizobium nodulation signals.* Cell 85, 673-681

4. Felle, H. H., Kondorosi, E., Kondorosi, A., and Schultze, M. (1998) *The role of ion fluxes in nod factor signalling in Medicago sativa.* The Plant Journal 13, 455-463

5. Cardenas, L., Holdaway-Clarke, T. L., Sanchez, F., Quinto, C., Feijo, J. A., Kunkel, J. G., and Hepler, P. K. (2000) *Ion changes in legume root hairs responding to nod factors.* Plant Physiol. 123, 443-452

6. Miller, D. D., Leferink-ten Klooster, H. B., and Emons, A. M. (2000) *Lipochito-oligosaccharide nodulation factors stimulate cytoplasmic polarity with longitudinal endoplasmic reticulum and vesicles at the tip in vetch root hairs.* Mol. Plant Microbe Interact. 13, 1385-1390

7. Esseling, J. J., Lhuissier, F. G., and Emons, A. M. (2003) *Nod factors-induced root hair curling: Continuous polar growth towards the point of nod factor application.* Plant Physiol. 132, 1982-1988

8. Arrighi, J. F., Barre, A., Ben Amor, B., Bersoult, A., Soriano, L. C., Mirabella, R., de Carvalho-Niebel, F., Journet, E. P., Gherardi, M., Huguet, T., Geurts, R., Dénarié, J., Rouge, P., and Gough, C. (2006) *The Medicago truncatula lysin [corrected] motif-receptor-like kinase gene family includes NFP and new nodule-expressed genes.* Plant Physiol. 142, 265-279

9. Ben Amor, B., Shaw, S. L., Oldroyd, G. E. D., Maillet, F., Penmetsa, R. V., Cook, D., Long, S. R., Dénarié, J., and Gough, C. (2003) *The NFP locus of Medicago truncatula controls an early step of nod factor signal transduction upstream of a rapid calcium flux and root hair deformation.* Plant J. 34, 495-506

10. Endre, G., Kereszt, A., Kevei, Z., Mihacea, S., Kaló, P., and Kiss, G. B. (2002) *A receptor kinase gene regulating symbiotic nodule development.* Nature 417, 962-966

11. Wais, R. J., Galera, C., Oldroyd, G., Catoira, R., Penmetsa, R. V., Cook, D., Gough, C., Dénarié, J., and Long, S. R. (2000) *Genetic analysis of calcium spiking responses in nodulation mutants of Medicago truncatula.* Proc. Natl. Acad. Sci. U. S. A. 97, 13407-13412

12. Ané, J. M., Kiss, G. B., Riely, B. K., Penmetsa, R. V., Oldroyd, G. E., Ayax, C., Lévy, J., Debellé, F., Baek, J. M., Kaló, P., Rosenberg, C., Roe, B. A., Long, S. R., Dénarié, J., and Cook, D. R. (2004) *Medicago truncatula DMI1 required for bacterial and fungal symbioses in legumes.* Science 303, 1364-1367

13. Capoen, W., Sun, J., Wysham, D., Otegui, M. S., Venkateshwaran, M., Hirsch, S., Miwa, H., Downie, J. A., Morris, R. J., Ane, J. M., and Oldroyd, G. E. (2011) *Nuclear membranes control symbiotic calcium signaling of legumes.* Proc. Natl. Acad. Sci. U. S. A. 108, 14348-14353

14. Peiter, E., Sun, J., Heckmann, A. B., Venkateshwaran, M., Riely, B. K., Otegui, M. S., Edwards, A., Freshour, G., Hahn, M. G., Cook, D. R., Sanders, D., Oldroyd, G. E., Downie, J. A., and Ané, J. M. (2007) *The Medicago truncatula DMI1 protein modulates cytosolic calcium signaling.* Plant Physiol. 145, 192-203

15. Lévy, J., Bres, C., Geurts, R., Chalhoub, B., Kulikova, O., Duc, G., Journet, E. P., Ané, J. M., Lauber, E., Bisseling, T., Dénarié, J., Rosenberg, C., and Debellé, F. (2004) *A putative Ca^{2+} and calmodulin-dependent protein kinase required for bacterial and fungal symbioses.* Science 303, 1361-1364

16. Messinese, E., Mun, J. H., Yeun, L. H., Jayaraman, D., Rouge, P., Barre, A., Lougnon, G., Schornack, S., Bono, J. J., Cook, D. R., and Ané, J. M. (2007) *A novel nuclear protein interacts with the symbiotic DMI3 calcium- and calmodulin-dependent protein kinase of Medicago truncatula.* Mol. Plant-Microbe Interact. 20, 912-921

17. Gleason, C., Chaudhuri, S., Yang, T., Muñoz, A., Poovaiah, B. W., and Oldroyd, G. E. (2006) *Nodulation independent of rhizobia induced by a calcium-activated kinase lacking autoinhibition.* Nature 441, 1149-1152

18. Oldroyd, G. E., Mitra, R. M., Wais, R. J., and Long, S. R. (2001) *Evidence for structurally specific negative feedback in the nod factor signal transduction pathway.* Plant J. 28, 191-199

19. Kaló, P., Gleason, C., Edwards, A., Marsh, J., Mitra, R. M., Hirsch, S., Jakab, J., Sims, S., Long, S. R., Rogers, J., Kiss, G. B., Downie, J. A., and Oldroyd, G. E. (2005) *Nodulation signaling in legumes requires NSP2, a member of the GRAS family of transcriptional regulators.* Science 308, 1786-1789

20. Smit, P., Raedts, J., Portyanko, V., Debellé, F., Gough, C., Bisseling, T., and Geurts, R. (2005) *NSP1 of the GRAS protein family is essential for rhizobial nod factor-induced transcription.* Science 308, 1789-1791

21. Hirsch, S., Kim, J., Muñoz, A., Heckmann, A. B., Downie, J. A., and Oldroyd, G. E. D. (2009) *GRAS proteins form a DNA binding complex to induce gene expression during nodulation signaling in Medicago truncatula.* Plant Cell 21, 545-557

22. Gough, C. and Cullimore, J. (2011) *Lipo-chitooligosaccharide signaling in endosymbiotic plant-microbe interactions.* Mol. Plant Microbe Interact. 24, 867-878

23. Maillet, F., Poinsot, V., André, O., Puech-Pagès, V., Haouy, A., Gueunier, M., Cromer, L., Giraudet, D., Formey, D., Niebel, A., Martinez, E. A., Driguez, H., Bécard, G., and Dénarié, J. (2011) *Fungal lipochitooligosaccharide symbiotic signals in arbuscular mycorrhiza.* Nature 469, 58-63

24. Mukherjee, A. and Ane, J. M. (2011) *Germinating spore exudates from arbuscular mycorrhizal fungi: Molecular and developmental responses in plants and their regulation by ethylene.* Mol. Plant Microbe Interact. 24, 260-270

25. Benedito, V. A., Torres-Jerez, I., Murray, J. D., Andriankaja, A., Allen, S., Kakar, K., Wandrey, M., Verdier, J., Zuber, H., Ott, T., Moreau, S., Niebel, A., Frickey, T., Weiller, G., He, J., Dai, X., Zhao, P. X., Tang, Y., and Udvardi, M. K. (2008) *A gene expression atlas of the model legume Medicago truncatula.* Plant J. 55, 504-513

26. Young, N. D., Debelle, F., Oldroyd, G. E., Geurts, R., Cannon, S. B., Udvardi, M. K., Benedito, V. A., Mayer, K. F., Gouzy, J., Schoof, H., Van de Peer, Y., Proost, S., Cook, D. R., Meyers, B. C., Spannagl, M., Cheung, F., De Mita, S., Krishnakumar, V., Gundlach, H., Zhou, S., Mudge, J., Bharti, A. K., Murray, J. D., Naoumkina, M. A., Rosen, B., Silverstein, K. A., Tang, H., Rombauts, S., Zhao, P. X., Zhou, P., Barbe, V., Bardou, P., Bechner, M., Bellec, A., Berger, A., Berges, H., Bidwell, S., Bisseling, T., Choisne, N., Couloux, A., Denny, R., Deshpande, S., Dai, X., Doyle, J. J., Dudez, A. M., Farmer, A. D., Fouteau, S., Franken, C., Gibelin, C., Gish, J., Goldstein, S., Gonzalez, A. J., Green, P. J., Hallab, A., Hartog, M., Hua, A., Humphray, S. J., Jeong, D. H., Jing, Y., Jocker, A., Kenton, S. M., Kim, D. J., Klee, K., Lai, H., Lang, C., Lin, S., Macmil, S. L., Magdelenat, G., Matthews, L., McCorrison, J., Monaghan, E. L., Mun, J. H., Najar, F. Z., Nicholson, C., Noirot, C., O'Bleness, M., Paule, C. R., Poulain, J., Prion, F., Qin, B., Qu, C., Retzel, E. F., Riddle, C., Sallet, E., Samain, S., Samson, N., Sanders, I., Saurat, O., Scarpelli, C., Schiex, T., Segurens, B., Severin, A. J., Sherrier, D. J., Shi, R., Sims, S., Singer, S. R., Sinharoy, S., Sterck, L., Viollet, A., Wang, B. B., Wang, K., Wang, M., Wang, X., Warfsmann, J., Weissenbach, J., White, D. D., White, J. D., Wiley, G. B., Wincker, P., Xing, Y., Yang, L., Yao, Z., Ying, F., Zhai, J., Zhou, L., Zuber, A., Denarie, J., Dixon, R. A., May, G. D., Schwartz, D. C., Rogers, J., Quetier, F., Town, C. D., and Roe, B. A. (2011) *The Medicago genome provides insight into the evolution of rhizobial symbioses.* Nature 480, 520-524

27. Brechenmacher, L., Lei, Z., Libault, M., Findley, S., Sugawara, M., Sadowsky, M. J., Sumner, L. W., and Stacey, G. (2010) *Soybean metabolites regulated in root hairs in response to the symbiotic bacterium bradyRhizobium japonicum.* Plant Physiol. 153, 1808-1822

28. Lohar, D. P., Sharopova, N., Endre, G., Penuela, S., Samac, D., Town, C., Silverstein, K. A., and VandenBosch, K. A. (2006) *Transcript analysis of early nodulation events in Medicago truncatula.* Plant Physiol. 140, 221-234

29. Wan, J., Torres, M., Ganapathy, A., Thelen, J., DaGue, B. B., Mooney, B., Xu, D., and Stacey, G. (2005) *Proteomic analysis of soybean root hairs after infection by bradyRhizobium japonicum.* Mol. Plant Microbe Interact. 18, 458-467

30. Asamizu, E., Nakamura, Y., Sato, S., and Tabata, S. (2005) *Comparison of the transcript profiles from the root and the nodulating root of the model legume Lotus japonicus by serial analysis of gene expression.* Mol. Plant Microbe Interact. 18, 487-498

31. Hogslund, N., Radutoiu, S., Krusell, L., Voroshilova, V., Hannah, M. A., Goffard, N., Sanchez, D. H., Lippold, F., Ott, T., Sato, S., Tabata, S., Liboriussen, P., Lohmann, G. V., Schauser, L., Weiller, G. F., Udvardi, M. K., and Stougaard, J. (2009) *Dissection of symbiosis and organ development by integrated transcriptome analysis of Lotus japonicus mutant and wild-type plants.* PLoS One 4, e6556

32. Maunoury, N., Redondo-Nieto, M., Bourcy, M., Van de Velde, W., Alunni, B., Laporte, P., Durand, P., Agier, N., Marisa, L., Vaubert, D., Delacroix, H., Duc, G., Ratet, P., Aggerbeck, L., Kondorosi, E., and Mergaert, P. (2010) *Differentiation of symbiotic cells and endosymbionts in Medicago truncatula nodulation are coupled to two transcriptome-switches.* PLoS One 5, e9519

33. Mitra, R. M. and Long, S. R. (2004) *Plant and bacterial symbiotic mutants define three transcriptionally distinct stages in the development of the Medicago truncatula/SinoRhizobium meliloti symbiosis.* Plant Physiol. 134, 595-604

34. Serna-Sanz, A., Parniske, M., and Peck, S. C. (2011) *Phosphoproteome analysis of Lotus japonicus roots reveals shared and distinct components of symbiosis and defense.* Mol. Plant Microbe Interact. 24, 932-937

35. Journet, E. P., El-Gachtouli, N., Vernoud, V., de Billy, F., Pichon, M., Dedieu, A., Arnould, C., Morandi, D., Barker, D. G., and Gianinazzi-Pearson, V. (2001) *Medicago truncatula ENOD11: A novel RPRP-encoding early nodulin gene expressed during mycorrhization in arbuscule-containing cells.* Mol. Plant-Microbe Interact. 14, 737-748

36. Catoira, R., Galera, C., de Billy, F., Penmetsa, R. V., Journet, E. P., Maillet, F., Rosenberg, C., Cook, D., Gough, C., and Dénarié, J. (2000) *Four genes of Medicago truncatula controlling components of a nod factor transduction pathway.* Plant Cell 12, 1647-1666

37. Grimsrud, P. A., den Os, D., Wenger, C. D., Swaney, D. L., Schwartz, D., Sussman, M. R., Ane, J. M., and Coon, J. J. (2010) *Large-scale phosphoprotein analysis in Medicago truncatula roots provides insight into in vivo kinase activity in legumes.* Plant Physiol. 152, 19-28

38. Trapnell, C., Pachter, L., and Salzberg, S. L. (2009) *TopHat: Discovering splice junctions with RNA-seq.* Bioinformatics 25, 1105-1111

39. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup. (2009) *The sequence Alignment/Map format and SAMtools.* Bioinformatics 25, 2078-2079

40. Anders, S. and Huber, W. (2010) *Differential expression analysis for sequence count data.* Genome Biol. 11, R106

41. Benjamini, Y. and Hochberg, Y. (1995) *Controlling the false discovery rate: A practical and powerful approach to multiple testing.* Journal of the Royal Statistical Society. Series B (Methodological) 57, 289-300

42. Syka, J. E., Coon, J. J., Schroeder, M. J., Shabanowitz, J., and Hunt, D. F. (2004) *Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry.* Proc. Natl. Acad. Sci. U. S. A. 101, 9528-9533

43. Olsen, J. V., Macek, B., Lange, O., Makarov, A., Horning, S., and Mann, M. (2007) *Higher-energy C-trap dissociation for peptide modification analysis.* Nat. Methods 4, 709-712

44. Wenger, C. D., Lee, M. V., Hebert, A. S., McAlister, G. C., Phanstiel, D. H., Westphall, M. S., and Coon, J. J. (2011) *Gas-phase purification enables accurate, multiplexed proteome quantification with isobaric tagging.* Nat. Methods 8, 933-935

45. Wenger, C. D., Phanstiel, D. H., Lee, M. V., Bailey, D. J., and Coon, J. J. (2011) *COMPASS: A suite of pre- and post-search proteomics software tools for OMSSA.* Proteomics 11, 1064-1074

46. Geer, L. Y., Markey, S. P., Kowalak, J. A., Wagner, L., Xu, M., Maynard, D. M., Yang, X., Shi, W., and Bryant, S. H. (2004) *Open mass spectrometry search algorithm.* J. Proteome Res. 3, 958-964

47. Phanstiel, D. H., Brumbaugh, J., Wenger, C. D., Tian, S., Probasco, M. D., Bailey, D. J., Swaney, D. L., Tervo, M. A., Bolin, J. M., Ruotti, V., Stewart, R., Thomson, J. A., and Coon, J. J. (2011) *Proteomic and phosphoproteomic comparison of human ES and iPS cells.* Nat. Methods 8, 821-827

48. Thompson, A., Schafer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., Neumann, T., Johnstone, R., Mohammed, A. K., and Hamon, C. (2003) *Tandem mass tags: A novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS.* Anal. Chem. 75, 1895-1904

49. Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A., and Pappin, D. J. (2004) *Multiplexed protein quantitation in saccharomyces cerevisiae using amine-reactive isobaric tagging reagents.* Mol. Cell. Proteomics 3, 1154-1169

50. Marsh, J. F., Rakocevic, A., Mitra, R. M., Brocard, L., Sun, J., Eschstruth, A., Long, S. R., Schultze, M., Ratet, P., and Oldroyd, G. E. (2007) *Medicago truncatula NIN is essential for rhizobial-independent nodule organogenesis induced by autoactive calcium/calmodulin-dependent protein kinase.* Plant Physiol. 144, 324-335

51. Murakami, Y., Miwa, H., Imaizumi-Anraku, H., Kouchi, H., Downie, J. A., Kawaguchi, M., and Kawasaki, S. (2006) *Positional cloning identifies Lotus japonicus NSP2, a putative transcription factor of the GRAS family, required for NIN and ENOD40 gene expression in nodule initiation.* DNA Research 13, 255-265

52. Andriankaja, A., Boisson-Dernier, A., Frances, L., Sauviac, L., Jauneau, A., Barker, D. G., and de Carvalho-Niebel, F. (2007) *AP2-ERF transcription factors mediate nod factor dependent mt ENOD11 activation in root hairs via a novel cis-regulatory motif.* Plant Cell 19, 2866-2885

53. Middleton, P. H., Jakab, J., Penmetsa, R. V., Starker, C. G., Doll, J., Kaló, P., Prabhu, R., Marsh, J. F., Mitra, R. M., Kereszt, A., Dudas, B., and VandenBosch, K. (2007) *An ERF transcription factor in Medicago truncatula that is essential for nod factor signal transduction.* Plant Cell 19, 1221-1234

54. Gomez, S. K., Javot, H., Deewatthanawong, P., Torres-Jerez, I., Tang, Y., Blancaflor, E. B., Udvardi, M. K., and Harrison, M. J. (2009) *Medicago truncatula and glomus intraradices gene expression in cortical cells harboring arbuscules in the arbuscular mycorrhizal symbiosis.* BMC Plant Biology 9, 10

55. Kevei, Z., Lougnon, G., Mergaert, P., Horvath, G. V., Kereszt, A., Jayaraman, D., Zaman, N., Marcel, F., Regulski, K., Kiss, G. B., Kondorosi, A., Endre, G., Kondorosi, E., and Ané, J. M. (2007) *3-hydroxy-3-methylglutaryl coenzyme a reductase 1 interacts with NORK and is crucial for nodulation in Medicago truncatula.* Plant Cell; Plant Cell 19, 3974-3989

56. Riely, B. K., He, H., Venkateshwaran, M., Sarma, B., Schraiber, J., Ane, J. M., and Cook, D. R. (2011) *Identification of legume RopGEF gene families and characterization of a Medicago truncatula RopGEF mediating polar growth of root hairs.* Plant J. 65, 230-243

57. Shimoda, Y., Han, L., Yamazaki, T., Suzuki, R., Hayashi, M., and Imaizumi-Anraku, H. (2012) *Rhizobial and fungal symbioses show different requirements for calmodulin binding to calcium calmodulin-dependent protein kinase in Lotus japonicus.* Plant Cell

58. Liu, H., Trieu, A. T., Blaylock, L. A., and Harrison, M. J. (1998) *Cloning and characterization of two phosphate transporters from Medicago truncatula roots: Regulation in response to phosphate and to colonization by arbuscular mycorrhizal (AM) fungi.* Mol. Plant Microbe Interact. 11, 14-22

59. Liu, J., Versaw, W. K., Pumplin, N., Gomez, S. K., Blaylock, L. A., and Harrison, M. J. (2008) *Closely related members of the Medicago truncatula PHT1 phosphate transporter gene family encode phosphate transporters with distinct biochemical activities.* J. Biol. Chem. 283, 24673-24681

60. Chiou, T. J., Liu, H., and Harrison, M. J. (2001) *The spatial expression patterns of a phosphate transporter (MtPT1) from Medicago truncatula indicate a role in phosphate transport at the root/soil interface.* Plant J. 25, 281-293

61. Grunwald, U., Guo, W., Fischer, K., Isayenkov, S., Ludwig-Muller, J., Hause, B., Yan, X., Kuster, H., and Franken, P. (2009) *Overlapping expression patterns and differential transcript levels of phosphate transporter genes in arbuscular mycorrhizal, pi-fertilised and phytohormone-treated Medicago truncatula roots.* Planta 229, 1023-1034

62. Lim, C. W., Park, J. Y., Lee, S. H., and Hwang, C. H. (2010) *Comparative proteomic analysis of soybean nodulation using a supernodulation mutant, SS2-2.* Biosci. Biotechnol. Biochem. 74, 2396-2404

63. Lee, M. V., Topper, S. E., Hubler, S. L., Hose, J., Wenger, C. D., Coon, J. J., and Gasch, A. P. (2011) *A dynamic model of proteome changes reveals new roles for transcript alteration in yeast.* Mol. Syst. Biol. 7, 514

64. Santelia, D., Kotting, O., Seung, D., Schubert, M., Thalmann, M., Bischof, S., Meekins, D. A., Lutz, A., Patron, N., Gentry, M. S., Allain, F. H., and Zeeman, S. C. (2011) *The phosphoglucan phosphatase like sex Four2 dephosphorylates starch at the C3-position in arabidopsis.* Plant Cell 23, 4096-4111

65. Bischof, S., Baerenfaller, K., Wildhaber, T., Troesch, R., Vidi, P. A., Roschitzki, B., Hirsch-Hoffmann, M., Hennig, L., Kessler, F., Gruissem, W., and Baginsky, S. (2011) *Plastid proteome assembly without Toc159: Photosynthetic protein import and accumulation of N-acetylated plastid precursor proteins.* Plant Cell 23, 3911-3928

66. Elmore, J. M., Liu, J., Smith, B., Phinney, B., and Coaker, G. (2012) *Quantitative proteomics reveals dynamic changes in the plasma membrane proteome during arabidopsis immune signaling.* Mol. Cell. Proteomics

67. Nuhse, T. S., Stensballe, A., Jensen, O. N., and Peck, S. C. (2004) *Phosphoproteomics of the arabidopsis plasma membrane and a new phosphorylation site database.* Plant Cell 16, 2394-2405

68. Chen, Y., Hoehenwarter, W., and Weckwerth, W. (2010) *Comparative analysis of phytohormone-responsive phosphoproteins in arabidopsis thaliana using TiO2-phosphopeptide enrichment and mass accuracy precursor alignment.* Plant J. 63, 1-17

69. Schutz, W., Hausmann, N., Krug, K., Hampp, R., and Macek, B. (2011) *Extending SILAC to proteomics of plant cell lines.* Plant Cell 23, 1701-1705

70. Reiland, S., Finazzi, G., Endler, A., Willig, A., Baerenfaller, K., Grossmann, J., Gerrits, B., Rutishauser, D., Gruissem, W., Rochaix, J. D., and Baginsky, S. (2011) *Comparative phosphoproteome profiling reveals a function of the STN8 kinase in fine-tuning of cyclic electron flow (CEF).* Proc. Natl. Acad. Sci. U. S. A. 108, 12955-12960

71. Nakagami, H., Sugiyama, N., Ishihama, Y., and Shirasu, K. (2012) *Shotguns in the front line: Phosphoproteomics in plants.* Plant Cell Physiol. 53, 118-124

72. Benschop, J. J., Mohammed, S., O'Flaherty, M., Heck, A. J., Slijper, M., and Menke, F. L. (2007) *Quantitative phosphoproteomics of early elicitor signaling in arabidopsis.* Mol. Cell. Proteomics 6, 1198-1214

73. Kline, K. G., Barrett-Wilt, G. A., and Sussman, M. R. (2010) *In planta changes in protein phosphorylation induced by the plant hormone abscisic acid.* Proc. Natl. Acad. Sci. U. S. A. 107, 15986-15991

74. Nakagami, H., Sugiyama, N., Mochida, K., Daudi, A., Yoshida, Y., Toyoda, T., Tomita, M., Ishihama, Y., and Shirasu, K. (2010) *Large-scale comparative phosphoproteomics identifies conserved phosphorylation sites in plants.* Plant Physiol. 153, 1161-1174

75. El Yahyaoui, F., Kuster, H., Ben Amor, B., Hohnjec, N., Puhler, A., Becker, A., Gouzy, J., Vernie, T., Gough, C., Niebel, A., Godiard, L., and Gamas, P. (2004) *Expression profiling in Medicago truncatula identifies more than 750 genes differentially expressed during nodulation, including many potential regulators of the symbiotic program.* Plant Physiol. 136, 3159-3176

76. Hernandez, G., Valdes-Lopez, O., Ramirez, M., Goffard, N., Weiller, G., Aparicio-Fabre, R., Fuentes, S. I., Erban, A., Kopka, J., Udvardi, M. K., and Vance, C. P. (2009) *Global changes in the transcript and metabolic profiles during symbiotic nitrogen fixation in phosphorus-stressed common bean plants.* Plant Physiol. 151, 1221-1238

77. Libault, M., Farmer, A., Joshi, T., Takahashi, K., Langley, R. J., Franklin, L. D., He, J., Xu, D., May, G., and Stacey, G. (2010) *An integrated transcriptome atlas of the crop model glycine max, and its use in comparative analyses in plants.* Plant J. 63, 86-99

78. De Smet, I., Vassileva, V., De Rybel, B., Levesque, M. P., Grunewald, W., Van Damme, D., Van Noorden, G., Naudts, M., Van Isterdael, G., De Clercq, R., Wang, J. Y., Meuli, N., Vanneste, S., Friml, J., Hilson, P., Jurgens, G., Ingram, G. C., Inze, D., Benfey, P. N., and Beeckman, T. (2008) *Receptor-like kinase ACR4 restricts formative cell divisions in the arabidopsis root.* Science 322, 594-597

79. He, Z. H., Fujiki, M., and Kohorn, B. D. (1996) *A cell wall-associated, receptor-like protein kinase.* J. Biol. Chem. 271, 19789-19793

80. Li, H., Zhou, S. Y., Zhao, W. S., Su, S. C., and Peng, Y. L. (2009) *A novel wall-associated receptor-like protein kinase gene, OsWAK1, plays important roles in rice blast disease resistance.* Plant Mol. Biol. 69, 337-346

81. Yano, K., Yoshida, S., Muller, J., Singh, S., Banba, M., Vickers, K., Markmann, K., White, C., Schuller, B., Sato, S., Asamizu, E., Tabata, S., Murooka, Y., Perry, J., Wang, T., Kawagachi, M., Imaizumi-Anraku, H., Hayashi, M., and Parniske, M. (2008) *CYCLOPS, a mediator of symbiotic intracellular accommodation.* Proc. Natl. Acad. Sci. U. S. A. 105, 20540-20545

82. Horvath, B., Yeun, L. H., Domonkos, A., Halasz, G., Gobbato, E., Ayaydin, F., Miro, K., Hirsch, S., Sun, J., Tadege, M., Ratet, P., Mysore, K. S., Ane, J. M., Oldroyd, G. E., and Kalo, P. (2011) *Medicago truncatula IPD3 is a member of the common symbiotic signaling pathway required for rhizobial and mycorrhizal symbioses.* Mol. Plant Microbe Interact. 24, 1345-1358

83. Ivashuta, S., Liu, J., Liu, J., Lohar, D. P., Haridas, S., Bucciarelli, B., VandenBosch, K. A., Vance, C. P., Harrison, M. J., and Gantt, J. S. (2005) *RNA interference identifies a calcium-dependent protein kinase involved in Medicago truncatula root development.* Plant Cell; Plant Cell 17, 2911-2921

84. Gargantini, P. R., Gonzalez-Rizzo, S., Chinchilla, D., Raices, M., Giammaria, V., Ulloa, R. M., Frugier, F., and Crespi, M. D. (2006) *A CDPK isoform participates in the regulation of nodule number in Medicago truncatula.* Plant Journal 48, 843-856

85. Tena, G., Asai, T., Chiu, W. L., and Sheen, J. (2001) *Plant mitogen-activated protein kinase signaling cascades.* Curr. Opin. Plant Biol. 4, 392-400

86. Wan, J., Zhang, S., and Stacey, G. (2004) *Activation of a mitogen-activated protein kinase pathway in arabidopsis by chitin.* Mol. Plant. Pathol. 5, 125-135

87. Ouaked, F., Rozhon, W., Lecourieux, D., and Hirt, H. (2003) *A MAPK pathway mediates ethylene signaling in plants.* EMBO J. 22, 1282-1288

88. Penmetsa, R. V. and Cook, D. (1997) *A legume ethylene-insensitive mutant hyperinfected by its rhizobial symbiont.* Science 275, 527-530

89. Oldroyd, G. E. D., Engstrom, E. M., and Long, S. R. (2001) *Ethylene inhibits the nod factor signal trasnduction pathway of Medicago truncatula.* Plant Cell 13, 1835-1849

90. Francia, D., Chiltz, A., Lo Schiavo, F., Pugin, A., Bonfante, P., and Cardinale, F. (2011) *AM fungal exudates activate MAP kinases in plant cells in dependence from cytosolic ca(2+) increase.* Plant Physiol. Biochem. 49, 963-969

91. Dale, S., Arro, M., Becerra, B., Morrice, N. G., Boronat, A., Hardie, D. G., and Ferrer, A. (1995) *Bacterial expression of the catalytic domain of 3-hydroxy-3-methylglutaryl-coa reductase (isoform Hmgr1) from arabidopsis-thaliana, and its inactivation by phosphorylation at Ser577 by brassica-oleracea 3-hydroxy-3-methylglutaryl-coa reductase kinase.* European Journal of Biochemistry 233, 506-513

92. Halford, N. G., Hey, S., Jhurreea, D., Laurie, S., McKibbin, R. S., Paul, M., and Zhang, Y. (2003) *Metabolic signalling and carbon partitioning: Role of Snf1-related (SnRK1) protein kinase.* J. Exp. Bot. 54, 467-475

93. Ikeda, Y., Koizumi, N., Kusano, T., and Sano, H. (2000) *Specific binding of a 14-3-3 protein to autophosphorylated WPK4, an SNF1-related wheat protein kinase, and to WPK4-phosphorylated nitrate reductase.* J. Biol. Chem. 275, 41528

94. Ding, Y., Kaló, P., Yendrek, C., Sun, J., Liang, Y., Marsh, J. F., Harris, J. M., and Oldroyd, G. E. D. (2008) *Abscisic acid coordinates nod factor and cytokinin signaling during the regulation of nodulation in Medicago truncatula.* Plant Cell; Plant Cell 20, 2681-2695

95. Tominaga, A., Nagata, M., Futsuki, K., Abe, H., Uchiumi, T., Abe, M., Kucho, K., Hashiguchi, M., Akashi, R., Hirsch, A. M., Arima, S., and Suzuki, A. (2009) *Enhanced nodulation and nitrogen fixation in the abscisic acid low-sensitive mutant enhanced nitrogen fixation1 of Lotus japonicus.* Plant Physiol. 151, 1965-1976

96. Kapranov, P., Jensen, T. J., Poulsen, C., de Bruijn, F. J., and Szczyglowski, K. (1999) *A protein phosphatase 2C gene, LjNPP2C1, from Lotus japonicus induced during root nodule development.* Proc. Natl. Acad. Sci. U. S. A. 96, 1738-1743

97. Black, D. J. and Persechini, A. (2011) *In calmodulin-IQ domain complexes, the ca(2+)-free and ca(2+)-bound forms of the calmodulin C-lobe direct the N-lobe to different binding sites.* Biochemistry 50, 10061-10068

98. Lee, K. H. and Larue, T. A. (1992) *Ethylene as a possible mediator of light- and nitrate-induced inhibition of nodulation of pisum sativum L. cv sparkle.* Plant Physiol. 100, 1334-1338

99. Nishimura, R., Ohmori, M., Fujita, H., and Kawaguchi, M. (2002) *A lotus basic leucine zipper protein with a RING-finger motif negatively regulates the developmental program of nodulation.* Proc. Natl. Acad. Sci. U. S. A. 99, 15206-15210

100. D'haeseleer, K., De Keyser, A., Goormachtig, S., and Holsters, M. (2010) *Transcription factor MtATB2: About nodulation, sucrose and senescence.* Plant Cell Physiol. 51, 1416-1424

101. Homma, M. K., Wada, I., Suzuki, T., Yamaki, J., Krebs, E. G., and Homma, Y. (2005) *CK2 phosphorylation of eukaryotic translation initiation factor 5 potentiates cell cycle progression.* Proc. Natl. Acad. Sci. U. S. A. 102, 15688-15693

102. Dennis, M. D., Person, M. D., and Browning, K. S. (2009) *Phosphorylation of plant translation initiation factors by CK2 enhances the in vitro interaction of multifactor complex components.* J. Biol. Chem. 284, 20615-20628

103. Yang, W. C., de Blank, C., Meskiene, I., Hirt, H., Bakker, J., van Kammen, A., Franssen, H., and Bisseling, T. (1994) *Rhizobium nod factors reactivate the cell cycle during infection and nodule primordium formation, but the cycle is only completed in primordium formation.* Plant Cell 6, 1415-1426

104. Bauer, P., Ratet, P., Crespi, M. D., Schultze, M., and Kondorosi, A. (1996) *Nod factors and cytokinins induce similar cortical cell division, amyloplast deposition and MsENOD12A expression patterns in alfalfa roots.* Plant J 10, 91-105

105. Herve, C., Lefebvre, B., and Cullimore, J. (2011) *How many E3 ubiquitin ligase are involved in the regulation of nodulation?* Plant. Signal. Behav. 6, 660-664

106. Miura, K., Jin, J. B., and Hasegawa, P. M. (2007) *Sumoylation, a post-translational regulatory process in plants.* Curr. Opin. Plant Biol. 10, 495-502

107. Lee, Y. J. and Yang, Z. (2008) *Tip growth: Signaling in the apical dome.* Curr. Opin. Plant Biol. 11, 662-671

108. Felle, H. H., Kondorosi, A., Kondorosi, E., and Schultze, M. (1996) *Rapid alkaliniztion in alfalfa root hairs in response to rhizobial lipochitooligosaccharide signals.* Plant J. 10, 295-301

109. Shaw, S. L. and Long, S. R. (2003) *Nod factor elicits two separable calcium responses in Medicago truncatula root hair cells.* Plant Physiol. ; Plant Physiol. 131, 976-984

110. Timmers, A. C., Vallotton, P., Heym, C., and Menzel, D. (2007) *Microtubule dynamics in root hairs of Medicago truncatula.* Eur. J. Cell Biol. 86, 69-83

111. Cheng, N. H., Liu, J. Z., Nelson, R. S., and Hirschi, K. D. (2004) *Characterization of CXIP4, a novel arabidopsis protein that activates the H+/Ca2+ antiporter, CAX1.* FEBS Lett. 559, 99-106

112. den Hartog, M., Musgrave, A., and Munnik, T. (2001) *Nod factor-induced phosphatidic acid and diacylglycerol pyrophosphate formation: A role for phospholipase C and D in root hair deformation.* Plant J. 25, 55-65

113. Yuksel, B. and Memon, A. R. (2009) *Legume small GTPases and their role in the establishment of symbiotic associations with Rhizobium spp.* Plant. Signal. Behav. 4, 257-260

114. Liu, W., Chen, A. M., Luo, L., Sun, J., Cao, L. P., Yu, G. Q., Zhu, J. B., and Wang, Y. Z. (2010) *Characterization and expression analysis of Medicago truncatula ROP GTPase family during the early stage of symbiosis.* J. Integr. Plant. Biol. 52, 639-652

115. Sun, J., Miwa, H., Downie, J. A., and Oldroyd, G. E. (2007) *Mastoparan activates calcium spiking analogous to nod factor-induced responses in Medicago truncatula root hair cells.* Plant Physiol. ; Plant Physiol. 144, 695-702

116. Yokota, K., Fukai, E., Madsen, L. H., Jurkiewicz, A., Rueda, P., Radutoiu, S., Held, M., Hossain, M. S., Szczyglowski, K., Morieri, G., Oldroyd, G. E., Downie, J. A., Nielsen, M. W., Rusek, A. M., Sato, S., Tabata, S., James, E. K., Oyaizu, H., Sandal, N., and Stougaard, J. (2009) *Rearrangement of actin cytoskeleton mediates invasion of Lotus japonicus roots by mesoRhizobium loti.* Plant Cell; Plant Cell

117. Miyahara, A., Richens, J., Starker, C., Morieri, G., Smith, L., Long, S., Downie, J. A., and Oldroyd, G. E. (2010) *Conservation in function of a SCAR/WAVE component during infection thread and root hair growth in Medicago truncatula.* Mol. Plant Microbe Interact. 23, 1553-1562

118. Kim, Y., Sung, J. Y., Ceglia, I., Lee, K. W., Ahn, J. H., Halford, J. M., Kim, A. M., Kwak, S. P., Park, J. B., Ho Ryu, S., Schenck, A., Bardoni, B., Scott, J. D., Nairn, A. C., and Greengard, P. (2006) *Phosphorylation of WAVE1 regulates actin polymerization and dendritic spine morphology.* Nature 442, 814-817

119. Tominaga, M., Yokota, E., Vidali, L., Sonobe, S., Hepler, P. K., and Shimmen, T. (2000) *The role of plant villin in the organization of the actin cytoskeleton, cytoplasmic streaming and the architecture of the transvacuolar strand in root hair cells of hydrocharis.* Planta 210, 836-843

120. Cooper, J. A. and Schafer, D. A. (2000) *Control of actin assembly and disassembly at filament ends.* Curr. Opin. Cell Biol. 12, 97-103

121. Davletov, B., Sontag, J. M., Hata, Y., Petrenko, A. G., Fykse, E. M., Jahn, R., and Sudhof, T. C. (1993) *Phosphorylation of synaptotagmin I by casein kinase II.* J. Biol. Chem. 268, 6816-6822

122. Gamas, P., de Billy, F., and Truchet, G. (1998) *Symbiosis-specific expression of two Medicago truncatula nodulin genes, MtN1 and MtN13, encoding products homologous to plant defense proteins.* Mol. Plant Microbe Interact. 11, 393-403

123. Peng, H. M., Dreyer, D. A., VandenBosch, K. A., and Cook, D. (1996) *Gene structure and differential regulation of the Rhizobium-induced peroxidase gene rip1.* Plant Physiol. 112, 1437-1446

124. Cardenas, L., Martinez, A., Sanchez, F., and Quinto, C. (2008) *Fast, transient and specific intracellular ROS changes in living root hair cells responding to nod factors (NFs).* Plant J. 56, 802-813

125. Demchenko, K., Zdyb, A., Feussner, I., and Pawlowski, K. (2012) *Analysis of the subcellular localisation of lipoxygenase in legume and actinorhizal nodules.* Plant. Biol. (Stuttg) 14, 56-63

126. Sun, J., Cardoza, V., Mitchell, D. M., Bright, L., Oldroyd, G. E. D., and Harris, J. M. (2006) *Crosstalk between jasmonic acid, ethylene and nod factor signaling allows integration of diverse inputs for regulation of nodulation.* Plant Journal 46, 961-970

127. Mathesius, U. (2008) *Auxin: At the root of nodule development?* Functional Plant Biology 35, 651-668

128. Gonzalez-Rizzo, S., Crespi, M., and Frugier, F. (2006) *The Medicago truncatula CRE1 cytokinin receptor regulates lateral root development and early symbiotic interaction with Sinorhizobium meliloti.* Plant Cell 18, 2680-2693

129. Mateos, P. F., Jimenez-Zurdo, J. I., Chen, J., Squartini, A. S., Haack, S. K., Martinez-Molina, E., Hubbell, D. H., and Dazzo, F. B. (1992) *Cell-associated pectinolytic and cellulolytic enzymes in Rhizobium leguminosarum biovar trifolii.* Appl. Environ. Microbiol. 58, 1816-1822

130. van Spronsen, P. C., Bakhuizen, R., van Brussel, A. A., and Kijne, J. W. (1994) *Cell wall degradation during infection thread formation by the root nodule bacterium Rhizobium leguminosarum is a two-step process.* Eur. J. Cell Biol. 64, 88-94

Chapter 3

Instant spectral assignment for advanced decision tree-driven mass spectrometry

CMR designed research, performed experiments, analyzed data, and wrote the paper.

This chapter has been published:

Bailey D. J.*, **Rose C. M.***, McAlister G. C., Brumbaugh J., Yu P., Westphall M. S., Thomson J. A., Coon J. J. "Instant Spectral Assignment for Advanced Decision Tree-Driven Mass Spectrometry." *Proceedings of the National Academy of Scientists*, **2012**, 109(22):8411-6.

*co-first author

Abstract

We have developed and implemented a sequence identification algorithm (*inSeq*) that processes tandem mass spectra in real-time using the mass spectrometer's (MS) on board processors. *inSeq* relies on accurate mass tandem MS data for swift spectral matching with high accuracy. The instant spectral processing technology takes ~ 16 ms to execute and provides information to enable autonomous, real-time decision making by the MS system. Using *inSeq*, and its advanced decision tree (DT) logic, we demonstrate: (1) real-time prediction of peptide elution windows *en masse* (~ 3 minute width, 3,000 targets), (2) significant improvement of quantitative precision and accuracy (~ 3X boost in detected protein differences), and (3) boosted rates of post-translation modification (PTM) site localization (90% agreement in real-time vs. offline localization rate and a ~25% gain in localized sites). The DT logic enabled by *inSeq* promises to circumvent longstanding problems with the conventional data-dependent acquisition paradigm and provides a direct route to streamlined and expedient targeted protein analysis.

Introduction

The shotgun sequencing method, where proteins are digested into peptides, chromatographed, and detected by mass spectrometry (MS), has rapidly evolved over the past two decades.[1,2] In this strategy eluting peptide cations have their mass-to-charge (*m/z*) values measured in the MS^1 scan. Then, in order of abundance, precursor *m/z* values are selected for a series of sequential tandem MS events (MS^2). This succession is cycled for the duration of the analysis. The process, called data-dependent acquisition (DDA), is at the very core of shotgun analysis and has not changed for over fifteen years; MS hardware, however, has. Major improvements in MS sensitivity, scan rate, mass accuracy and resolution have been achieved. Orbitrap hybrid systems, for example, routinely achieve low ppm mass accuracy with MS/MS repetition rates of 5-10 Hz.[3,4] Constant operation of such systems generates hundreds of thousands of spectra in days. These MS^2 spectra are then mapped to sequence using database search algorithms.[5-7]

The DDA sampling strategy offers an elegant simplicity and has proven highly useful, across most any MS platform, for discovery-driven proteomics. Of recent years, however, emphasis has shifted from protein identification to peptide-level quantification – often with certain targets in mind. In this context faults in the DDA approach have become increasingly evident. There are two primary limitations of the DDA approach: First, is poor run-to-run reproducibility and, second, is the inability to effectively target peptides of interest.[8] Irreproducibility is caused by the highly stochastic nature of choosing which *m/z* peaks to sample. Dozens and even hundreds of peptides often co-elute so that low-level signals often get selected in one run and not the next. And selecting *m/z* peaks to sequence by abundance certainly does not offer the opportunity to inform the system of pre-selected targets.

Several DDA add-ons and alternatives have been examined. Sampling depth, for example, can be increased by preventing selection of an *m/z* value identified in a prior technical replicate (PANDA).[9] Irreproducibility can be somewhat countered by informing the DDA algorithm of the precursor *m/z* values of desired targets (inclusion list) – if observed this can ensure their selection for MS^2. Frequently, however, low abundance peptides may not have precursor signals above noise so that MS^2 scan, which is requisite for identification, is never triggered. This conundrum is avoided altogether in the data-independent

acquisition approach (DIA).[10] Here no attention is paid to precursor abundance, or even presence, instead consecutive *m/z* isolation windows are dissociated and mass analyzed. A main drawback of DIA is that it requires significantly more instrument analysis time as MS^2 scans from every *m/z* window must be collected.[11] As such, DDA analysis remains the preeminent method for MS data acquisition.

Besides improvements in MS analyzer performance, numerous alternative dissociation methods and scan types have recently advanced. These include collision, electron and/or photon-based fragmentation (i.e., trapHCD, HCD, ETD, IRMPD, etc.), specialized quantification scans (i.e., selected ion monitoring (SIM) or precursor purification methods (QuantMode, QM)), or simply analysis using varied precursor ion targets, *m/z* accuracy, etc.[12-17] Each of these techniques show applicability and superlative performance for some subset of peptide precursors. The result is a dizzying alphabet soup of techniques, scan types, and parameter space that is not easily integrated into the current data acquisition paradigm. Recently we introduced a decision tree (DT) algorithm that used precursor *m*, *z*, and *m/z* to automatically determine, in-real time, whether to employ CAD or ETD during MS^2.[18] The approach significantly improved sequencing success rates and was an important first step in a movement toward development of informed acquisition. Unfortunately only a limited amount of knowledge can be gleaned from the very basic information of *m*, *z*, and abundance.

Here we describe the next advance in DT acquisition technology – instant sequence confirmation (*inSeq*). The *inSeq* algorithm processes MS^2 spectra at the moment of collection using the MS system's on board processing power. With sequence in hand the MS acquisition system can process this knowledge to make autonomous, real-time decisions about what type of scan to trigger next. Here, with the *inSeq* instant identification algorithm, we extend our simple DT method by adding several new decision nodes. These nodes enable novel automated functionalities including: real-time elution prediction, advanced quantification, PTM localization, large-scale targeted proteomics, and increased proteome coverage, among others. This technology provides a direct pathway to transform the current, i.e., DDA, passive data collection paradigm. Specifically, knowing the identity of a peptide that is presently eluting into the MS system permits an ensemble of advanced, automated decision-making logic.

Results

Instant sequence confirmation (*inSeq*)

To develop an advanced DT acquisition schema, which can seamlessly incorporate the myriad of specialized procedures and scans available on modern day MS systems, we must expedite the spectral analysis process – i.e., from off-line to real-time. Knowledge gained from these instant analyses enables automated decision making logic to extract the most information possible from an MS analysis. There are two obvious pathways to incorporate real-time spectral analysis within an MS system. The first approach exports spectra for processing with an external computing system followed by import of the search outcome.[19] A second, more elegant strategy, is to perform all computation within the MS on-board computing system.[20] The former approach circumvents complications in accessing instrument firmware and allows for the use of more sophisticated processing power; however, a serious constraint is the time required for import/export of the information (i.e., ~40 ms). For this reason we have pursued technologies and computational algorithms that integrate real-time spectral analysis into the MS system's on-board processors and firmware. We call this method instant sequence confirmation (*inSeq*). Experimental details (*i.e.,* peptide candidates, scan sequences, etc…) are transferred, on-demand, along with the instrument's method file to the instrument before the experiment commences to enable flexibility in experimental design with minimum configuration.

To establish robustness across platforms we implemented *inSeq* on two distinct MS systems (operating with different code bases) – a dual cell quadrupole linear ion trap-Orbitrap hybrid (LTQ-Velos Orbitrap) and a quadrupole mass filter-Orbitrap hybrid (Q-Exactive). In both cases we modified and extended the instrument firmware to quickly (~ <20 ms in the case of the more modern Q-Exactive system) and accurately (<2% false discovery rate (FDR)) map MS/MS spectra to sequence. The embedded peptide database-matching algorithm processes MS/MS scans immediately (**Figure 1A-B**) by comparing product ions present in the MS/MS scan to those from peptide candidates pre-loaded onto the instrument's firmware. Note the candidate sequences are first filtered so that only sequences whose mass is within 30 ppm of the

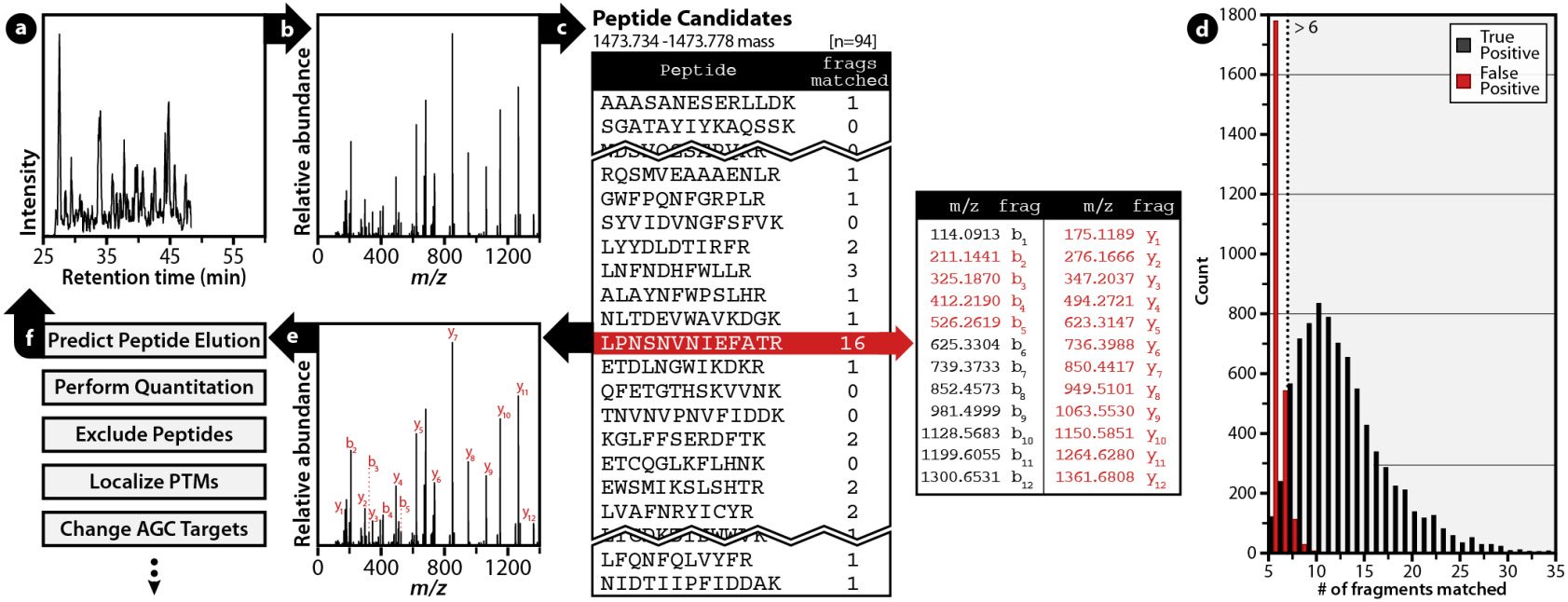


Figure 1. Progression of the *inSeq* logic.
A) A nHPLC-MS/MS chromatogram at 48.35 minutes along with an MS^2 scan B) that was acquired at that time following dissociation of a doubly protonated feature of m/z 737.86. Upon collection of the MS^2 scan, *inSeq* groups all peptide candidates (in silico, n = 94) whose theoretical mass are within 30 ppm of the experimentally determined precursor neutral mass 1473.756 (C). Then *inSeq* performs *in silico* fragmentation to produce a theoretical product ion series for each of the 94 candidates and proceeds to compare each to the experimental spectrum (<10 ppm mass accuracy). D) Plot of *inSeq* identifications compared to conventional post-acquisition searching. *inSeq* agrees (True Positive) >98% of the time when >6 fragment ions are matched.

sampled precursor neutral mass are considered (**Figure 1C**). For each candidate sequence the number of +1 product ions (+2 ions are included for precursors >+2) that matched the spectrum at a mass tolerance < 10 ppm is recorded (**Figure 1C**). The algorithm employs memoization optimization techniques for calculating product ions masses, i.e., calculating the masses only once and storing the results in a fast reference table. Next, it uses straightforward scoring metrics, providing sufficient evidence for the confirmation of a putative sequence without burdening the system with non-essential calculations. On both MS platforms the real-time confirmation algorithm was expediently executed and required no hardware modification, taking an average of 16 ms to perform (Q-Exactive). Note similar processing times were achieved on the older Velos-Orbitrap system; however, a ~ 100 ms overhead was included because complete collection of the Orbitrap transient signal is necessary before the spectrum can be examined.

To characterize the *inSeq* algorithm we performed a nHPLC-MS/MS experiment on tryptic peptides derived from human embryonic stem cells. Prior to analysis, a database consisting of all theoretical tryptic peptides (up to three missed cleavages, 6-50 in length) contained within the human proteome was uploaded to the instrument's (Q-Exactive) on-board computer. A data-dependent top 10 method was employed and analysis proceeded as usual, except following each MS/MS scan the *inSeq* algorithm was executed and the results logged. This manifest of instant identifications was then compared to those made post-acquisition via traditional database searching at a 1% FDR (reverse-decoy method). We assumed the conventional post-acquisition approach to represent the true answer and compared the number of correct instant spectral identifications as a function of matched product ions (**Figure 1D**). From these data we conclude the detection of >6 product ions at high mass accuracy (<10 ppm) by the *inSeq* algorithm produces the correct sequence identification >98% of the time.

inSeq represents a simple, expedient approach to correlate sequence to spectrum and is positioned to become an essential technology in transforming the current passive data collection paradigm. Specifically, learning the identity of a peptide that is presently eluting into the MS system permits an ensemble of advanced, automated decision-making logic. These concepts build upon our previous

development of the data-dependent decision tree (DT) method. There we embedded an on-board algorithm to make unsupervised, real-time decisions of which fragmentation method to engage, based on precursor charge (z) and *m/z*. Here, with the *inSeq* instant identification algorithm, we extend our simple DT method by adding several new decision nodes. These nodes enable automated functionalities including: real-time elution prediction, advanced quantification, PTM localization, large-scale targeted proteomics, and increased proteome coverage, among others (**Figure 1F**).

Predicting peptide elution

Liquid chromatography is the conventional approach to fractionate highly complex peptide mixtures prior to measurement by MS. The highest MS sensitivity is achieved when one tunes the MS system to detect a given target (i.e., execute MS/MS) regardless of its presence in the preceding MS^1 event (i.e., selected reaction monitoring). SRM measurements deliver both sensitivity and reproducibility at the cost of bandwidth. Specifically, if one does not know the elution time of a target, the duration of the nHPLC-MS/MS analysis must be dedicated to conditions for that specific entity. If elution times are known, then multiple SRM scan events can be programmed allowing for detection of multiple targets; however, chromatographic conditions must remain identical or the scheduled SRM elution windows will no longer align. Still, the bandwidth of that approach is low, ~ 100 peptide targets per nHPLC-MS/MS analysis, and compiling such an experiment is highly laborious.[21]

We surmised that *inSeq* could inform the MS system, without human intervention, of which peptide targets are most likely to subsequently elute. Such capability could enable robust, large-scale targeting (> 500 per analysis) in an automated manner. Our approach relies upon relative peptide elution order and, consequently, bypasses the use of absolute retention times, which shift depending on chromatographic conditions and are not directly portable from multiple disparate experiments. Peptide elution order can be obtained in two ways: First, discovery experiments can be employed to determine retention order by normalizing the measured retention time for each detected peptide sequence. Second, the relative hydrophobicity for any sequence can be theoretically determined using existing software (e.g., SSRCalc).[22-

[24] In our experience experimentally determined retention order offers better precision; still, it requires prior knowledge which may not be available. However retention order is determined, the real-time confirmation algorithm maintains a rolling average of the calculated elution order (CEO – a number describing the relative elution order of a target peptide) so that target peptides having nearby CEOs are specifically pursued (**Figure 2B**). **Figure 2** presents an overview of this approach. This example, 60.39 minutes into the chromatograph, highlights the last five *inSeq* identified peptides and their average CEO (26.926 a.u.). The on-board algorithm then computes an asymmetric CEO window (5 a.u., 24.926-29.926) that presents a short list of desired targets having CEOs within that range (**Figure 2C**). With this information the MS system can trigger specialized MS^2 scans specific to this refined target subset. Note that as targets are identified, the CEO window is dynamically adjusted so targets come into and out of the range precisely when they are eluting.

To test this technology, we performed a data-dependent top-10 nHPLC-MS/MS experiment in which tryptic peptides from a human ES cell sample were separated over a 60 minute gradient. Following data collection the resulting MS/MS spectra were mapped to sequence using database searching (1% FDR). The unique peptide identifications (4,237) were sorted by observed retention time – this ordering then served as the CEO. 3,000 of these peptides were randomly selected as "targets" and subsequently loaded onto the instrument firmware (Velos-Orbitrap), along with their respective CEO, as a database for *inSeq*. The sample was then re-analyzed with *inSeq* activated, but with a doubled gradient length (120 min). **Figure 2D** displays the CEO window as calculated in real-time by the MS system (*inSeq*) plotted beside the actual elution time of identified peptides. Greater than 95% of the peptides (2,889) fell within the rolling CEO window and were identified by both *inSeq* and post-acquisition searching. Further, the rolling CEO window width averaged ~ 4 minutes for each of these 3,000 targets. At our present capability we can achieve window widths similar to those used in absolute scheduling type experiments (~ 3-6 minutes) on a scale that is 30X larger (e.g., 3,000 targets vs. 100) with minimal effort.[21,25] Further, we demonstrate that our approach easily adapts to different chromatographic conditions with no negative effects (**Figure 2D**). The key to the high portability and simplicity of our algorithm is the use of *inSeq* for continual, real-time realignment.

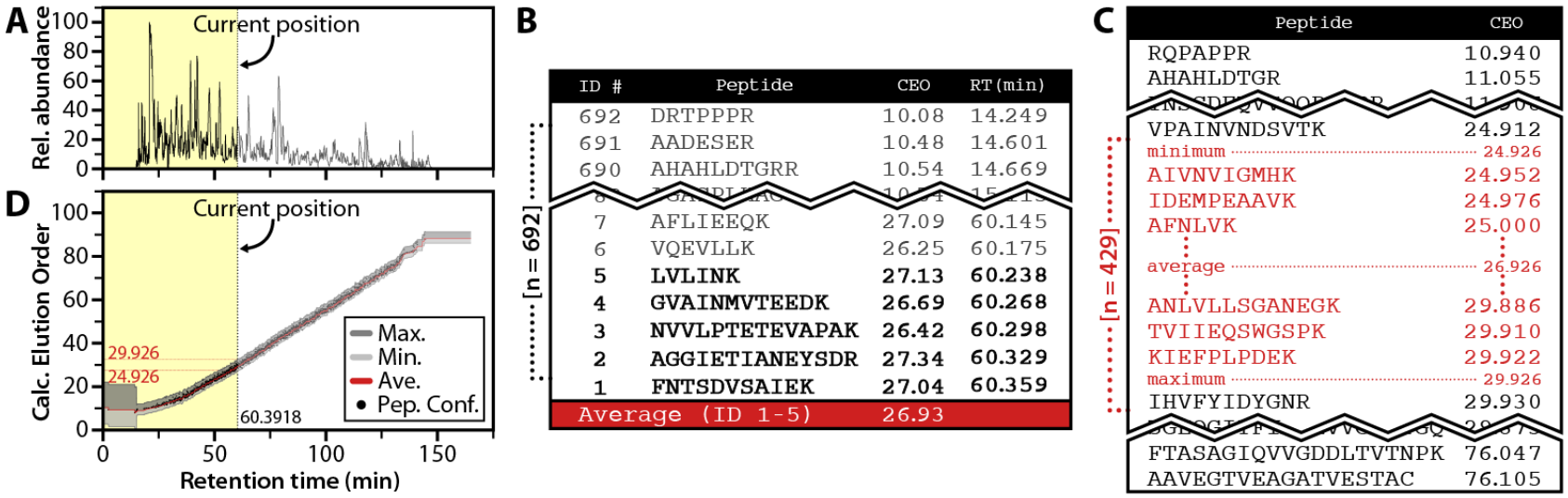


Figure 2. Elution order prediction using *inSeq*.

A) Experimental chromatogram, 60.39 minutes into a 120 minute gradient, obtained while *inSeq* was recording instant identifications. B) The *inSeq* calculated elution order (CEO) obtained by averaging the CEOs of the preceding 5 instantly identified peptides. C) The asymmetric window (5 a.u.) surrounding the instant CEO average (μ = 26.926) and their corresponding sequences. *inSeq* can then refine a target list to only include those having CEOs within the predicted window. (D) This analysis is repeated following each new peptide confirmation to constantly realign the CEO window based on current chromatographic conditions.

Improvement of quantitative accuracy.

The method of stable isotope labeling has greatly propelled large-scale, quantitative analysis.[26-31] While generally robust, these techniques can yield spotty data for certain peptide and protein groups – mainly those present at low abundances. For SILAC, low signal-to-noise (S/N) precursor peaks in the MS^1 scan often result in either omission of that particular feature or quantitative imprecision, if included.[32] For isobaric tagging, low intensity reporter ion signals (MS^2) induces similar shortcomings.[33] We surmised that *inSeq* could be employed to counter these limitations.

First, we developed an *inSeq* module to improve the quality of isobaric label-based measurements. The module analyzes MS^2 spectra, using *inSeq*, and, when a peptide of interest is detected, the quality of quantitative data is assessed. Should the reporter ion signal fall below a specified threshold, *inSeq* triggers follow-up scans to generate increased signal at the very instant the target peptide is eluting. In one implementation, we instructed *inSeq* to automatically trigger three quantitative scans, using the recently developed QuantMode (QM) method, to generate superior quality quantitative data on targets of high value.[17] Triggering three additional MS/MS scans commits ~ 1 second of instrument time to a single precursor. Constitutive operation in this mode would severely hamper duty cycle; however, triggering the scan sequence on only a few hundred pre-selected, high value targets commits a modest time commitment and can deliver exceptional quantitative information on the user-selected peptide targets. Thus, we only trigger these three QM scans whenever a target peptide is identified by *inSeq*. The trio of QM scans are then summed offline.

To assess this decision node we analyzed a sample comprising three biological replicates of human embryonic stem cells (ESCs) pre- and two days-post bone morphogenetic protein 4 (BMP4) treatment (i.e., TMT 6-plex, three pre-treatment and three post BMP4 treatment cell populations). BMP4, a growth factor that induces context-dependent differentiation in pluripotent stem cells, is widely used to study differentiation to biologically relevant cell lineages such as mesoderm and endoderm.[34-36] **Figure 3A** demonstrates the benefit of summing isobaric tag intensities from one, two, or three consecutive quantitation scans for an *inSeq* identified target peptide having the sequence FCADHPFLFFIR from the

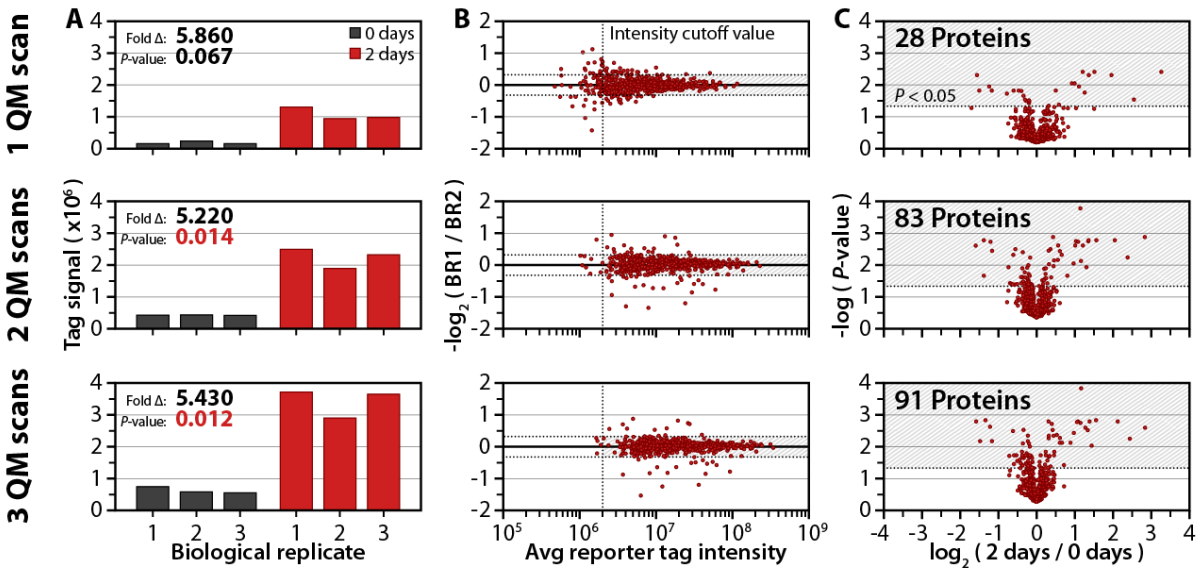


Figure 3. *inSeq* improves quantitative outcomes for isobaric tagging.
An *inSeq* decision node was written to so that a real-time identification of a target sequence prompted automatic acquisition of three consecutive QuantMode (QM) scans. A) Summing the reporter ion tag intensity from one, two, or three QM scans greatly improves the statistical significance of the measurement. B) Summation of QM technical replicates reduces the variation in biological replicate measurement by increasing reporter ion S/N. C) *inSeq*-triggered QM scans increase the number of significantly changing proteins from 28 to 91.

protein SERPINB8. Here the ratio of change between control and treatment cell lines measured in one QM scan is large (5.86) but not significant (P = 0.067, Student's t-test with Storey correction).[37] Note significance testing was accomplished by assessing variation within the three biological replicates of both treatment and control cell lines. The measured ratio remains relatively unchanged (5.22 and 5.43) as reporter tag signal from additional quantitation scans are added; however, the corresponding P-values decrease to 0.014 and 0.012 when two or three quantitation scans are summed. By plotting the $\log_2$ ratio of quantified proteins from the three biological replicates against the average intensity of isobaric labels (**Figure 3B**) we demonstrate this improved significance results from boosted reporter signal-to-noise. Ideally this $\log_2$ ratio would be zero, indicating perfect biological replication; however, when only one quantitation scan is employed this ratio severely deviates from zero with decreasing tag intensity. To improve overall data quality and to omit potentially erroneous measurements we, and others, employ arbitrary reporter signal cutoffs (dashed vertical line in **Figure 3B**). These cutoffs eliminate many low-abundance peptides which often represent proteins of interest. Summation of additional quantitation scans increases the average reporter tag intensity, raising nearly all of the protein measurements above the intensity cutoff value (74, 9, and 4 proteins omitted using one, two, and three quantification scans, respectively). This quantification decision node also increased the number of proteins within 25% of perfect biological replication (horizontal dashed line).

To determine if the method could improve the number of statistically significant differences between the cell populations, we calculated the $\log_2$ ratio of treated vs. control (i.e., 2 days/0 days) for each of the 596 quantified proteins (P<0.05, Student's t-test with Storey correction, **Figure 3C**). To display both fold change and significance the P-value for each protein difference was plotted against its corresponding ratio. Only 28 proteins display significant change when one QM scan is used. By simply adding the reporter tag signal from one additional scan the number of significantly changing proteins increases nearly threefold, from 28 to 83. When all three QM scans are analyzed together, the number of significantly changing proteins increases slightly, to 91.

Many stable isotope incorporation techniques measure heavy and light peptide pairs in MS^1 (e.g., SILAC). This approach, of course, requires the detection of both partners; note low abundance peptides are often identified with low, or no, precursor signal in the MS^1. We supposed that addition of another *inSeq* decision node could circumvent this problem. We cultured human embryonic stem cells (ESCs) in light and heavy media (i.e., Lysine[$^{13}C_6$-$^{15}N_2$]). Protein extract from these cultures was mixed 5:1 (light:heavy), before digestion overnight with LysC. The SILAC node was developed to select precursors from an MS^1 scan only if the monoisotopic mass was within 30 ppm of any target on a list which contained 4,000 heavy and light peptides from a previous discovery run. Targets were selected only if the SILAC ratio deviated from the expected ratio of 5 by 25%, i.e., the subset containing the most error. Following MS/MS, the resulting spectra were analyzed using *inSeq*. When a target of interest was identified, *inSeq* instructed the system to immediately record a SIM scan surrounding the light/heavy pair with a small, charge-dependent isolation window (~8-10 Th). The narrow isolation range provides gas phase enrichment of low abundance target precursors and enables accurate quantification of many peptides whose data would otherwise be discarded.

The average ratio of the light and heavy peptides subtly, but significantly, shifted from 4.47 under normal analysis to 5.34 for the *inSeq* triggered SIM scans (Student's t-test, p-value < $6x10^{-20}$). More importantly, the number of useable measurements, i.e., when both partners of the pair are observed, increased by ~ 20% (2,887 under normal analysis to 3,548 with *inSeq*). The overall distribution of these data is plotted in **Figure 4A**. Panel B of **Figure 4** displays an example of the *inSeq*-triggered SIM scan and the increase in S/N and accuracy it affords. Here the MS/MS scan of the light partner was mapped, in real-time, to the sequence IEELDQENEAALENGIK, a pre-defined target. This event triggered a high resolution SIM scan (8 Th window), which determined the ratio of 4.99:1 (correct ratio 5:1). Here gas phase enrichment was essential to quantify the relative abundance, as the isotopic envelope of the heavy partner was not observed, even with extensive spectral averaging of successive MS^1 scans (~ *30 s*, **Figure 4B**). Whether for MS^1 or MS^2 centric methods, we conclude that *inSeq* technology will significantly improve the quality of quantitative data with only a minimal impact on duty cycle.

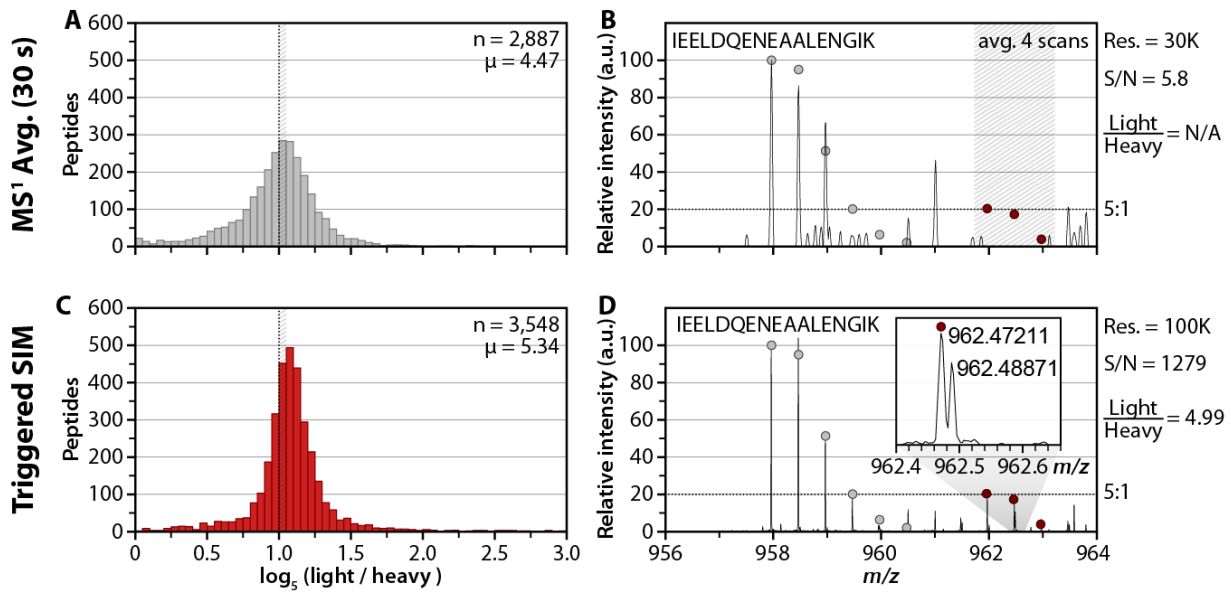


Figure 4. *inSeq* can improve quantitative outcomes for SILAC.
Following an *inSeq* confirmation of a peptide having the sequence, IEELDQENEAALENGIK, a narrow (8 Th), high-resolution (R = 100,000) SIM scan was automatically triggered and increased the S/N from 5.8 to 1,279 (Panels A and B). This SIM scan enabled detection of both partners and yielded the correct ratio of 5:1 (light:heavy). Besides increased dynamic range, the theoretical isotope distribution (shown in open circles) closely matches in the SIM scan (B), while the signal for the heavy partner is not even detectable in the MS^1 (A). Over our entire data set, the *inSeq* triggered SIM scans improve the mean ratio from 4.47 to 5.34, but, more impressively, produced ~ 20% more quantifiable measurements (3,548 vs. 2,887).

Post-Translational Modification Site localization.

The presence, or lack, of post-translational modifications (PTMs) on proteins plays a major role in cellular function and signaling. Unambiguous localization of PTMs to residue demands observation of product ions resulting from cleavage of the residues adjacent to the site of modification, i.e., site-determining fragments (SDFs). Phosphorylation analysis has become increasingly widespread; however, in a typical analysis only about half of the identified phosphorylation sites can be mapped with single amino acid resolution. These ambiguous spectra stymie systems-level data analysis and prevent use of thousands of identified phosphopeptides. We reasoned that *inSeq* could be leveraged to boost PTM localization rates by dynamically modifying MS^2 acquisition conditions when necessary. As such we developed an online PTM localization decision node to determine, within milliseconds, whether a MS/MS spectrum contains SDFs to unambiguously localize the PTM. Should SDFs be lacking, *inSeq* instantly orchestrates further interrogation.

The PTM localization node is engaged when *inSeq* confirms the detection of a PTM-bearing peptide. After the sequence is confirmed, *inSeq* assesses the confidence with which the PTM(s) can be localized to a particular amino acid residue. This procedure is accomplished by computing an online probability score similar to post-acquisition PTM localization software – i.e., AScore.[38] Using the MS system's embedded processors, *inSeq* compares all possible peptide isoforms against the MS/MS spectrum. For each SDF the number of matches at <10 ppm tolerance is counted and an AScore is calculated (*inSeq* uses similar math). If the AScore of the best fitting isoform is above 13 ($p < 0.05$) the PTM is declared localized and the *inSeq* routine ends. When the AScore is lower than 13, however, *inSeq* triggers further characterization of the eluting precursor until either the site has been deemed localized or all decision nodes have been exhausted. Additional characterization can include many procedures such as acquisition of MS/MS spectra using different fragmentation methods (e.g., CAD, HCD, ETD, PD, etc.), varied fragmentation conditions (e.g., collision energy, reaction time, laser fluence, etc.), increased spectral averaging, MS^n, pseudo MS^n, modified dynamic exclusion, and altered AGC target values, among others.[39]

To obtain proof-of-concept results we wrote a simple *inSeq* node that triggered an ETD MS/MS scan of phosphopeptides that were not localized following HCD MS^2. An example of how this logic can function is presented in **Figure 5** where a target phosphopeptide was detected during a shotgun experiment. The sequence, RNSSEASSGDFLDLK, was confirmed to contain a phosphoryl group; however, the *inSeq* algorithm could not confidently localize the PTM to any of the four Ser residues (AScore = 0). *inSeq* determined the most probable sites as either Ser 3 or Ser 4. Next, *inSeq* triggered an ETD MS^2 scan of the same precursor (**Figure 5B**). The resulting spectrum was then analyzed for the presence of the SDFs, $c_3/z^{\bullet}_{12}$. Both of these fragments were present and the phosphorylation site was localized to Ser 3 with an AScore of 31.0129 ($p < 0.00079$). Post-acquisition analysis confirmed the results of our online *inSeq* approach – both spectra (HCD and ETD) were confidently identified and their calculated AScores were 0 and 45.58, respectively. When compared on a global scale, 993 of the 1,134 *inSeq*-identified phosphopeptides had localization judgments that matched post-acquisition A-Score analysis (**Figure 5**). This slight difference is the result of using two slightly different localization algorithms for online and post-acquisition analysis.[30,38] Nonetheless, these data demonstrate that our localization node is highly effective at instantaneously determining whether a PTM site can be localized. Still, we must act upon this information to achieve markedly improved localization outcomes. Unfortunately, only marginal gains were achieved in this basic implementation as most precursors were doubly charged and, therefore, not effectively sequenced by ETD. Next, we modified the *inSeq* decision node to incorporate a dissociation method DT. Here a follow-up ETD or combination ion trap CAD/HCD scan was triggered depending upon precursor charge (*z*) and *m/z*. With the slightly evolved algorithm the *inSeq* method detected 998 phosphopeptides in a single shotgun experiment. It determined that 324 of these identifications lacked the information to localize the PTM site and, in those cases, triggered the new dissociation decision node. 78 of these unlocalizable sites were confidently mapped with this technique - salvaging nearly 25% of the unlocalized sites (**Figure 5D**). These encouraging results demonstrate that *inSeq* has great promise to curtail the problem of PTM localization in a highly automated fashion. We note there are dozens of parameters, e.g., collision energy, reaction time,

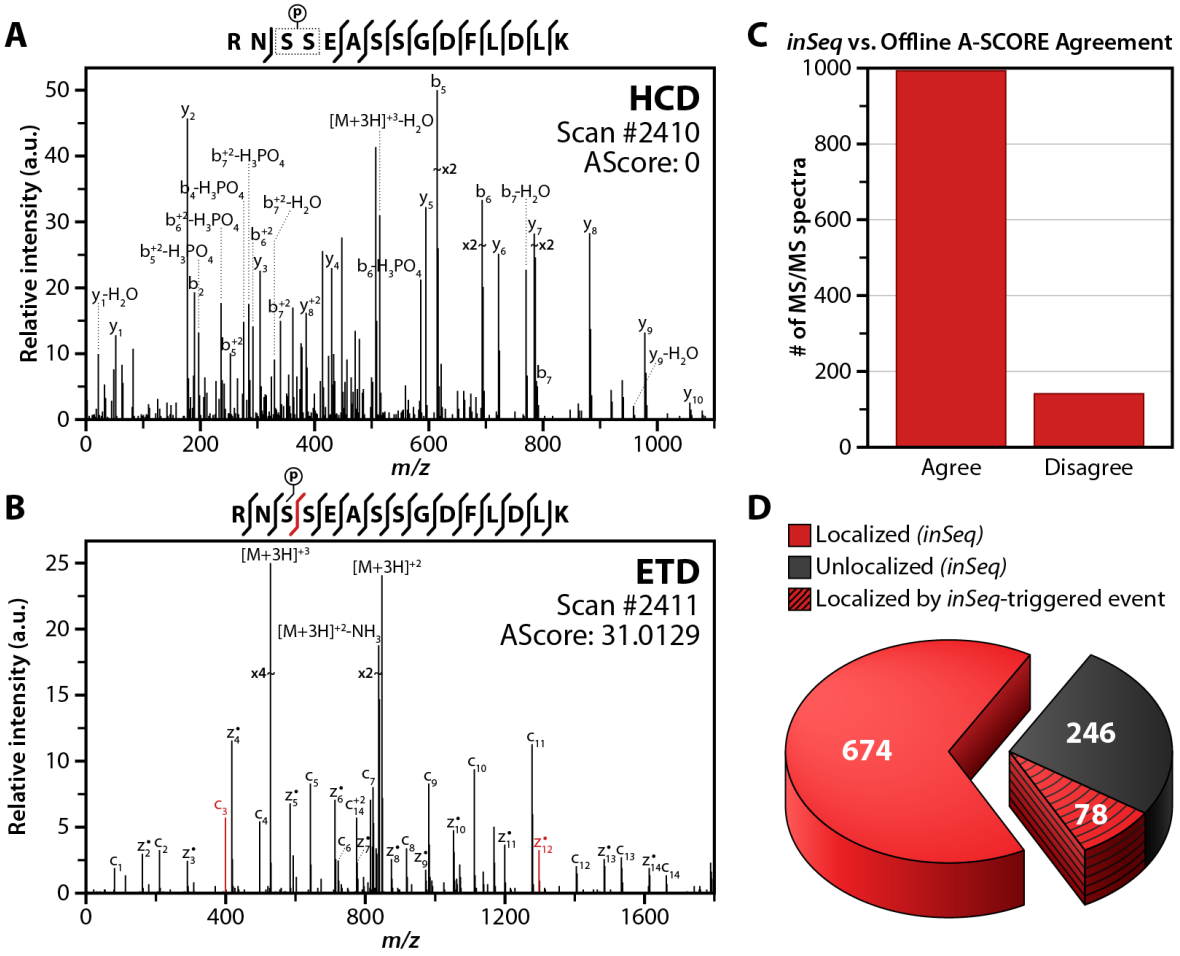


Figure 5. *inSeq* can improve PTM localization rates.
Following MS^2 (HCD) of the singly-phosphorylated precursor RNsSEASSGDFLDLK, *inSeq* could not find sufficient information to confidently localize the modification to either Ser 3 or 4 (A, AScore = 0). *inSeq* immediately triggered an ETD MS^2 scan event on the same precursor (B). This spectrum was assigned an AScore of 31.0129 (phosphorylation on Ser3) and was considered confidently localized – note the SDFs c_3 and $z^{\cdot}_{12}$ ions. C) Globally, the *inSeq* localization calculation agreed with offline analysis using the actual AScore algorithm. D) Using a simple dissociation method DT, *inSeq* produced a confidently localized phosphorylation site for 78 of 324 unlocalizable sites.

laser fluence, spectral averaging, MS^n, adjustment of AGC target values, targeting of a different charge state, etc., to explore in the continued advancement of this PTM localization decision node.

Discussion

Here we describe an instant sequencing algorithm (*inSeq*) that operates using the pre-existing processors of the mass spectrometer. *inSeq* relies on accurate mass MS^2 data for expedient spectral matching with high accuracy. Rapid real-time sequencing affords several novel data acquisition opportunities. To orchestrate these opportunities we constructed an advanced decision tree logic that extends are our earlier use of the method to intelligently select dissociation type. The approach can circumvent longstanding problems with the conventional DDA paradigm. We provide three such examples herein. First, we demonstrate that knowledge of which peptide sequences are eluting can facilitate the prediction of soon to elute targets. This method shows strong promise to revolutionize the way in which targeted proteomics is conducted. Second, we used quantitative decision nodes that fired when *inSeq* detected a peptide sequence of interest. With either SILAC or isobaric tagging, significant gains in quantitative outcomes were documented. Third, we endowed *inSeq* with an instant PTM site-localization algorithm to determine whether or not to initiate more rigorous follow-up at the very instant the peptide of interest was eluting. We show that the *inSeq* site localizer is highly effective (90% agreement with post-acquisition analysis) and that triggering a simple dissociation method DT can improve site localization by ~ 25%. Further development will doubtless deliver additional gains.

Targeted proteomics is an area of increasing importance. Following discovery analysis it is natural to cull the list of several thousand detected proteins to several hundred key players. In an ideal world these key proteins are then monitored in dozens or even hundreds of samples with high sensitivity and reproducibility, without rigorous method development. And this, of course, must be done expediently as hundreds of samples are involved. We envision that advanced DT analysis with *inSeq* could offer such a platform. Using the retention time prediction algorithm we introduced here one can foresee the *inSeq* algorithm quickly and precisely monitoring hundreds of peptides without the extensive labor and pre-

planning required by the selected reaction monitoring (SRM) technique, current state-of-the-art.[40] Other possibilities include automated pathway analysis where user-defined proteins, within a collection of pathways, are simply uploaded to the MS system. Then, *inSeq* automatically determines the best peptides to track, their retention times, and constructs the method. Two key advantages over current SRM technology make this operation possible. First, knowledge of specific fragmentation transitions are not necessary as all products are monitored with high mass accuracy. Second, precise elution time scheduling is not necessary as *inSeq* can use CEO, experimental or theoretical, to dynamically adjust the predicted elution of targets. In this fashion the most tedious components of the SRM workflow can be avoided.

Experimental Procedures

Cell culture

Human embryonic stem cells (line H1) were maintained in feeder independent media as previously described (Chen et al). For SILAC experiments, DMEM/F12 lacking lysine and arginine (Mediatech Inc.) was supplemented with light arginine (Sigma-Alrich) and either heavy labeled lysine (Cambridge Isotopes Laboratories) or light lysine (Sigma-Aldrich). Cells were cultured on Matrigel (BD Biosciences) and split 1:8 at approximately 80% confluency using 0.1 mM EDTA. To harvest cells, TripLE Express (Invitrogen) was applied for five minutes at 37 degrees C. Following cell detachment, an equivalent volume of ice-cold DPBS (Invitrogen) was added before centrifugation. Cell pellets were subsequently washed twice in ice-cold DPBS and stored at –80 degrees C. BMP4-treated cells were grown and harvested as described above, except that 5ng/mL BMP4 (R&D Systems) was added into the media and cells were split using TrypLE (Invitrogen). For BMP4 experiments, single cells were plated at the density of 4 X 104/cm2, for 2 days of treatment. We collected ~108 cells for each analysis.

Cell lysis

For all analysis, human embryonic stem cells were lysed in ice-cold 8M urea, 40 mM NaCl, 50 mM tris (pH 8), 2 mM MgCl2, 50 mM NaF, 50 mM b-glycerol phosphate, 1 mM sodium orthovanadate, 10 mM

sodium pyrophosphate, 1X mini EDTA-free protease inhibitor (Roche Diagnostics), and 1X phosSTOP phosphatase inhibitor (Roche Diagnostics). To solubilize protein and ensure complete lysis, samples were sonicated three times for 15 seconds with 30 second pauses. Total protein was then quantified using a BCA protein assay kit (Thermo Scientific Pierce). Cells were lysed by sonication and protein extracted. For SILAC experiments protein from the two cell cultures was mixed 5:1 light:heavy before digestion.

Isobaric label sample preparation.

For analysis, 250 ug of protein from each sample was reduced by adding DTT to a final concentration of 5 mM, and alkylated with 15 mM iodoacetamide before final capping with 5 mM DTT. Digestion was carried out by adding LysC (Wako Chemicals) at a 1:100 enzyme-to-protein ratio and incubating at 37 degrees C for 2 hours. At this time, the lysate was diluted with 25 mM tris (pH 8) to a final urea concentration of 1.5 M and further digested for 12 hours at 37 degrees C with trypsin (Promega) at a 1:100 enzyme to protein ratio. Peptides were then acidified with TFA to quench the reaction and de-salted using C-18 solid phase extraction (SPE) columns (Waters). TMT labeling was carried out per manufacturer's directions (Thermo Scientific Pierce). Samples were mixed in a 1:1:1:1:1:1 ratio before analysis.

SILAC sample preparation.

Protein from the light and heavy embryonic stem cell cultures was mixed in a 5:1 ratio (light:heavy) by pooling 2.5 mg of light protein and 0.5 mg of heavy protein. The sample was reduced by adding DTT to a final concentration of 5 mM, and alkylated with 15 mM iodoacetamide before final capping with 5 mM DTT. Digestion was carried out by adding LysC (Wako Chemicals) at a 1:100 enzyme-to-protein ratio and incubating at 37 degrees C overnight. Peptides were then acidified with TFA to quench the reaction and de-salted using C-18 solid phase extraction (SPE) columns (Waters).

Phosphopeptide sample preparation.

From an embryonic stem cell culture, 1 mg of protein was reduced by adding DTT to a final concentration of 5 mM, and alkylated with 15 mM iodoacetamide before final capping with 5 mM DTT. Digestion was carried out by adding LysC (Wako Chemicals) at a 1:100 enzyme-to-protein ratio and

incubating at 37 degrees C for 2 hours. At this time, the lysate was diluted with 25 mM tris (pH 8) to a final urea concentration of 1.5 M and further digested for 12 hours at 37 degrees C with trypsin (Promega) at a 1:100 enzyme to protein ratio. Peptides were then acidified with TFA to quench the reaction and de-salted using C-18 solid phase extraction (SPE) columns (Waters). Phosphopeptides were enriched via immobilized metal affinity chromatography (IMAC) using magnetic beads (Qiagen). Following equilibration with water, the beads were treated with 40 mM EDTA (pH 8.0) for 30 minutes with shaking, and washed 3x with water again. The beads were then incubated with 100 mM FeCl3 for 30 minutes with shaking and finally were washed 3 times with 80% acetonitrile/0.1% TFA. Samples were likewise resuspended in 80% acetonitrile/0.15% TFA and incubated with beads for 45 minutes with shaking. The resultant mixture was washed 3 times with 1 mL 80% acetonitrile/0.1% TFA, and eluted using 1:1 acetonitrile:0.7% NH4OH in water. Eluted phosphopeptides were acidified immediately with 4% formic acid and lyophilized to ~5 µL.

nano-High performance liquid chromatography

For all samples online reverse-phase chromatography was performed using a NanoAcquity UPLC system (Waters). Peptides were loaded onto a pre-column (75 µm ID, packed with 7 cm C18 particles, Alltech) for 10 min at a flow rate of 1 µm/min. Samples were then eluted over an analytical column (50 µm ID, packed with 15 cm C18 particles, Alltech) using either a 60 or 120 min linear gradient from 2% to 35% acetonitrile with 0.2% formic acid and a flow rate of 300 nL/min.

Target list construction and inSeq setup

For all experiments, the monoisotopic mass, charge state, and previously determined retention time of target peptides was included for use by the *inSeq* algorithm. In addition, peptides modified on methionines or tyrosines were omitted from all target lists. For peptide elution and isobaric label quantitation *inSeq* experiments, a target list of 4,000 peptides was constructed from a previous nHPLC-MS/MS experiment employing a 90 min nHPLC gradient. For SILAC *inSeq* experiments, peptides identified at 1% FDR in a discovery nHPLC-MS/MS experiment were analyzed to determine the light:heavy partner ratio. A target list of 2,000 peptide pairs (4,000 total peptides) whose ratio deviated

from the expected value of 5 by at least 25% was constructed. This subset of peptides included many measurements in which the signal to noise was low, or a partner was missing. For phosphorylation *inSeq* experiments, phosphopeptides identified at 1% FDR in a discovery nHPLC-MS/MS experiment were analyzed by the Phosphinator to assign phosphosite locations. A target list comprising 2,174 phosphopeptides was constructed and used for both ETD only and decision tree (DT) *inSeq* methods.

Target lists were loaded into the instrument's firmware for instant access during acquisition. Peptide lists were stored in an internal database and sorted based on their precursor mass for fast look ups using a binary search algorithm. A parameter file was preloaded into the firmware prior to each experiment to specific scan sequences and instrument parameters needed for the intended experiment.

Mass spectrometry

All experiments were performed on Thermo LTQ Orbitrap Velos and Q-Exactive mass spectrometers. The LTQ Orbitrap Velos used firmware version 2.6.0.1065 SP3 with additional ion trap control language (ITCL) modifications to enable *inSeq* operation. MS1 scans were performed in the Orbitrap at 30,000 resolution at a max injection time of 500 ms and a target value of 1e6. MS2 scans were also performed in the Orbitrap at a resolution of 7,500 and with HCD normalize collision energy (NCE) of 27%, for a max fill time of 500 ms. The Q-Exactive was operated using version 2.0 Build 142800 with a modified python code base for *inSeq* data acquisition control. Q-Exactive MS1 scans were collected at 70,000 resolution for a max injection time of 120 ms or if the 1e6 AGC target value was reached. MS2 events were measured at 17,500 resolution at a target value of 1e5, 120 ms max injection time and 26% NCE. Instrument methods for both the LTQ Orbitrap Velos and Q-Exactive were overridden during acquisition by the instrument's firmware to provide for dynamic *inSeq* operation.

Database searching and FDR estimation

MS/MS data was analyzed using the Coon OMSSA Proteomics Software Suite (COMPASS).[41] The Open Mass Spectrometry Search Algorithm (OMSSA; version 2.1.8) was used to search spectra against the International Protein Index(IPI) human database version 3.85.[42] Precursor mass tolerance was set to ±4.5 Da and monoisotopic mass tolerance was set to ±0.015 Da for fragments ions. For all experiments,

carbamidomethylation of cysteines was included as a fixed modification, while oxidation of methionines was set as a variable modification. For TMT experiments, TMT on the N-terminus and TMT on lysines were included as fixed modifications and TMT on tryosines was added as a variable modification. For SILAC experiments heavy lysine was added as a variable modification. Results were filtered to a 1% FDR at both the peptide and protein level. For phosphopeptides, the Phosphinator software was used to localize phosphorylation sites[30]

Protein and peptide quantification.

TMT quantification was performed using TagQuant within COMPASS. This program extracts reporter ion intensities multiplies them by injection times to determine counts. Purity correction was performed as previously described. Tag intensities were normalized to ensure that the total signal from each channel was equal. For evaluation of multiple QuantMode (QM) scans, data was analyzed at the peptide level by only quantifying the first, the sum of first and second, or the sum of the first, second, and third QM scans using TagQuant. Peptides were then combined into protein groups (ProteinHerder) and quantified at the protein level (ProteinTagQuant) within COMPASS. Experimental ratios and p-values (Student's t-test assuming equal variance) were determined using Microsoft Excel. To correct for multiple hypothesis testing, we applied Storey correction using the freely available program QVALUE.[37]

SILAC quantification was performed with in-house software that retrieved the peak intensities of both SILAC partners from either a single *inSeq*-triggered SIM scan (monoisotopic peak) or performed an extract ion chromatogram (30 sec window) of identified precursor. A ratio of partner abundance was only calculated if both SILAC partners had an intensity at least twice that of the noise.

Firmware programming methodology.

We employed object-oriented programming (OOP) as the main programming paradigm for *inSeq* due to its ability to encapsulate data efficiently and for its modularity and expandability. Each peptide candidate is stored as object that comprises of an ordered list of amino acids objects, an neutral mass

(double), a calculate elution order (double), along with various other properties (z, modifications, etc...). We supplemented the peptide object with multiple methods to help calculate fragment ions on-the-fly, handle fixed and variable modifications, as well as determining relative hydrophobicity using a modified SSRCalc algorithm.

Peptide lists are stored on the workstation computer in a text-based document that is parsed in prior to the start of the experiment. The peptides are loaded into a list within the firmware which is then sorted on the neutral mass of each peptide to enable quick binary searching capabilities. For the Velos MS, ~4,000 peptide targets could be loaded at any given time due to the limitations of the firmware's memory. The Q-Exactive has a much larger memory capacity (2 GB, expandable to 4 GB) and is capable of loading a complete tryptic digestion of Human proteins accounting for up to 3 missed cleavages (>10 million peptides). In addition to loading peptide targets, a configuration file (text-based) is also downloaded to the firmware to set up the next experiment conditions and properties. The updated instrument control program for the Q-Exactive (Tune 2.0) allows for direct control of *inSeq* operation through the use of the status tree.

At the start of the experiment, the binding CEO window (stored as two doubles) is set to a wide width (*e.g.*, **Figure 2D**, time 0 – 15 min) to ensure the peptide elution prediction will start correctly and not miss the first eluting peptides. Before an MS1 scan is performed, the CEO window is updated (as mentioned in the main text) only if a peptide target(s) was identified in the last round of MS/MS events. Following acquisition of a MS1 scan, peptide candidates that have a CEO value within this CEO window are selected and copied to a new list. This step is performed first to quickly reduce the number of targets that need to be searched for in the MS1 spectrum. Following CEO filtering, the m/z(s) of each selected peptide candidate are searched for in the MS1 spectrum at a specified mass tolerance (*e.g.*, 30 ppm). If a peak was detected within the tolerance, with a signal-to-noise above some threshold, that peptide is placed on top of a stack of future MS/MS scan events to be performed. Each MS/MS event in the stack is then subsequently popped off, performed and analyzed (discussed below). When this stack is empty, the whole process is repeated by triggering a new MS1 scan event until the end of the experiment.

Following acquisition of a MS/MS spectrum, all peptides targets whose neutral mass is within some specified mass window are selected. Each selected peptide is then matched against the MS/MS spectrum using the methods described in the main text. If a peptide is matched, a Boolean is flipped to alert the system to update of the CEO window during the next MS1 stage. Additionally, depending on the experiment details, other flags may be triggered to inform the MS system to collect additional scans (i.e. QuantMode scan, SIM scan, etc...) immediately following the current scan. This is accomplished by placing the new scan event on top of the stack of MS/MS events. Finally, when a peptide is matched, it is effectively removed from future consideration by setting an identified Boolean flag within the peptide object. Peptides that have this flag enabled will be bypassed in all future analyses.

Because OOP is highly flexible and modular, adding additional real-time analyses is straightforward. For example, PTM localization is only triggered when an identified peptide contains a variable modification that can exist in multiple isoforms. Additionally, real-time analyses can be completely skipped if instructed too without interfering with other analyses.

Acknowledgments

We thank Steve Gygi for assistance in analyzing the SILAC datasets and A.J. Bureta for figure illustrations. We also thank Jae Schwartz, John Syka, Mike Senko, and Jens Griep-Raming for helpful discussions. This work was funded by the grants from the NIH (GM081629 and HG004952), the NSF (DBI-0701846), and Thermo Fisher Scientific. J.B was funded by an NSF Graduate Research Fellowship and NIH traineeship (5T32GM08349). C.M.R was also funded by an NSF Graduate Research Fellowship and NIH Traineeship (T32GM008505).

References

1. Washburn MP, Wolters D, & Yates JR. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol*, 2001, 19(3):242-247.

2. Nilsson T, Mann M, Aebersold R, Yates JR, Bairoch A, & Bergeron JJM. Mass spectrometry in high-throughput proteomics: ready for the big time. *Nat Methods*, 2010, 7(9):681-685.

3. Wenger CD, McAlister GC, Xia QW, & Coon JJ. Sub-part-per-million Precursor and Product Mass Accuracy for High-throughput Proteomics on an Electron Transfer Dissociation-enabled Orbitrap Mass Spectrometer. *Mol Cell Proteomics*, 2010, 9(5):754-763.

4. Michalski A, Damoc E, Hauschild JP, Lange O, Wieghaus A, Makarov A, Nagaraj N, Cox J, Mann M, & Horning S. Mass Spectrometry-based Proteomics Using Q Exactive, a High-performance Benchtop Quadrupole Orbitrap Mass Spectrometer. *Mol Cell Proteomics*, 2011, 10(9).

5. Eng JK, Mccormack AL, & Yates JR. An Approach to Correlate Tandem Mass-Spectral Data of Peptides with Amino-Acid-Sequences in a Protein Database. *J Am Soc Mass Spectr*, 1994, 5(11):976-989.

6. Perkins DN, Pappin DJC, Creasy DM, & Cottrell JS. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*, 1999, 20(18):3551-3567.

7. Geer LY, Markey SP, Kowalak JA, Wagner L, Xu M, Maynard DM, Yang XY, Shi WY, & Bryant SH. Open mass spectrometry search algorithm. *J Proteome Res*, 2004, 3(5):958-964.

8. Liu HB, Sadygov RG, & Yates JR. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem*, 2004, 76(14):4193-4201.

9. Hoopmann MR, Merrihew GE, von Haller PD, & MacCoss MJ. Post Analysis Data Acquisition for the Iterative MS/MS Sampling of Proteomics Mixtures. *J Proteome Res*, 2009, 8(4):1870-1875.

10. Venable JD, Dong MQ, Wohlschlegel J, Dillin A, & Yates JR. Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nat Methods*, 2004, 1(1):39-45.

11. Panchaud A, Scherl A, Shaffer SA, von Haller PD, Kulasekara HD, Miller SI, & Goodlett DR. Precursor Acquisition Independent From Ion Count: How to Dive Deeper into the Proteomics Ocean. *Anal Chem*, 2009, 81(15):6481-6488.

12. McAlister GC, Phanstiel DH, Brumbaugh J, Westphall MS, & Coon JJ. Higher-energy Collision-activated Dissociation Without a Dedicated Collision Cell. *Mol Cell Proteomics*, 2011, 10(5).

13. Olsen JV, Macek B, Lange O, Makarov A, Horning S, & Mann M. Higher-energy C-trap dissociation for peptide modification analysis. *Nat Methods*, 2007, 4(9):709-712.

14. Syka JEP, Coon JJ, Schroeder MJ, Shabanowitz J, & Hunt DF. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *P Natl Acad Sci USA*, 2004, 101(26):9528-9533.

15. Little DP, Speir JP, Senko MW, Oconnor PB, & Mclafferty FW. Infrared Multiphoton Dissociation of Large Multiply-Charged Ions for Biomolecule Sequencing. *Anal Chem*, 1994, 66(18):2809-2815.

16. Olsen JV & Mann M. Improved peptide identification in proteomics by two consecutive stages of mass spectrometric fragmentation. *P Natl Acad Sci USA*, 2004, 101(37):13417-13422.

17. Wenger CD, Lee MV, Hebert AS, McAlister GC, Phanstiel DH, Westphall MS, & Coon JJ. Gas-phase purification enables accurate, multiplexed proteome quantification with isobaric tagging. *Nat Methods*, 2011, 8(11):933-935.

18. Swaney DL, McAlister GC, & Coon JJ. Decision tree-driven tandem mass spectrometry for shotgun proteomics. *Nat Methods*, 2008, 5(11):959-964.

19. Graumann J, Scheltema RA, Zhang Y, Cox J, & Mann M. A framework for intelligent data acquisition and real-time database searching for shotgun proteomics. *Molecular & cellular proteomics : MCP*, 2011.

20. Bailey DJ, McAlister GC, Rose CM, Hebert AS, Wenger C, Lee M, Westphall MS, & Coon JJ (2011) How High Mass Accuracy Measurements Will Transform Targeted Proteomics. *Proceedings of the 59th ASMS Conference on Mass Spectrometry and Allied Topics*, (ASMS, Denver, Colorado).

21. Yan W, Luo J, Robinson M, Eng J, Aebersold R, & Ranish J. Index-ion Triggered MS2 Ion Quantification: A Novel Proteomics Approach for Reproducible Detection and Quantification of Targeted Proteins in Complex Mixtures. *Mol Cell Proteomics*, 2011, 10(3).

22. Krokhin OV, Craig R, Spicer V, Ens W, Standing KG, Beavis RC, & Wilkins JA. An improved model for prediction of retention times of tryptic peptides in ion pair reversed-phase HPLC - Its application to protein peptide mapping by off-line HPLC-MALDI MS. *Mol Cell Proteomics*, 2004, 3(9):908-919.

23. Krokhin OV. Sequence-specific retention calculator. Algorithm for peptide retention prediction in ion-pair RP-HPLC: Application to 300-and 100-angstrom pore size C18 sorbents. *Anal Chem*, 2006, 78(22):7785-7795.

24. Kiyonami R, Schoen A, Prakash A, Peterman S, Zabrouskov V, Picotti P, Aebersold R, Huhmer A, & Domon B. Increased Selectivity, Analytical Precision, and Throughput in Targeted Proteomics. *Mol Cell Proteomics*, 2011, 10(2):-.

25. Schmidt A, Gehlenborg N, Bodenmiller B, Mueller LN, Campbell D, Mueller M, Aebersold R, & Domon B. An Integrated, Directed Mass Spectrometric Approach for In-depth Characterization of Complex Peptide Mixtures. *Mol Cell Proteomics*, 2008, 7(11):2138-2150.

26. Gruhler A, Olsen JV, Mohammed S, Mortensen P, Faergeman NJ, Mann M, & Jensen ON. Quantitative phosphoproteomics applied to the yeast pheromone signaling pathway. *Mol Cell Proteomics*, 2005, 4(3):310-327.

27. Choe L, D'Ascenzo M, Relkin NR, Pappin D, Ross P, Williamson B, Guertin S, Pribil P, & Lee KH. 8-Plex quantitation of changes in cerebrospinal fluid protein expression in subjects undergoing intravenous immunoglobulin treatment for Alzheimer's disease. *Proteomics*, 2007, 7(20):3651-3660.

28. de Godoy LMF, Olsen JV, Cox J, Nielsen ML, Hubner NC, Frohlich F, Walther TC, & Mann M. Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature*, 2008, 455(7217):1251-U1260.

29. Xiao KH, Sun JP, Kim J, Rajagopal S, Zhai B, Villen J, Haas W, Kovacs JJ, Shukla AK, Hara MR, Hernandez M, Lachmann A, Zhao S, Lin YA, Cheng YS, Mizuno K, Ma'ayan A, Gygi SP, & Lefkowitz RJ. Global phosphorylation analysis of beta-arrestin-mediated signaling downstream of a seven transmembrane receptor (7TMR). *P Natl Acad Sci USA*, 2010, 107(34):15299-15304.

30. Phanstiel DH, Brumbaugh J, Wenger CD, Tian SL, Probasco MD, Bailey DJ, Swaney DL, Tervo MA, Bolin JM, Ruotti V, Stewart R, Thomson JA, & Coon JJ. Proteomic and phosphoproteomic comparison of human ES and iPS cells. *Nat Methods*, 2011, 8(10):821-U884.

31. Lee MV, Topper SE, Hubler SL, Hose J, Wenger CD, Coon JJ, & Gasch AP. A dynamic model of proteome changes reveals new roles for transcript alteration in yeast. *Mol Syst Biol*, 2011, 7.

32. Bakalarski CE, Elias JE, Villen J, Haas W, Gerber SA, Everley PA, & Gygi SP. The Impact of Peptide Abundance and Dynamic Range on Stable-Isotope-Based Quantitative Proteomic Analyses. *J Proteome Res*, 2008, 7(11):4756-4765.

33. Zhang Y, Askenazi M, Jiang JR, Luckey CJ, Griffin JD, & Marto JA. A Robust Error Model for iTRAQ Quantification Reveals Divergent Signaling between Oncogenic FLT3 Mutants in Acute Myeloid Leukemia. *Mol Cell Proteomics*, 2010, 9(5):780-790.

34. Lengerke C, Schmitt S, Bowman TV, Jang IH, Maouche-Chretien L, McKinney-Freeman S, Davidson AJ, Hammerschmidt M, Rentzsch F, Green JBA, Zon LI, & Daley GQ. BMP and Wnt specify hematopoietic fate by activation of the Cdx-Hox pathway. *Cell Stem Cell*, 2008, 2(1):72-82.

35. Yang L, Soonpaa MH, Adler ED, Roepke TK, Kattman SJ, Kennedy M, Henckaerts E, Bonham K, Abbott GW, Linden RM, Field LJ, & Keller GM. Human cardiovascular progenitor cells develop from a KDR plus embryonic-stem-cell-derived population. *Nature*, 2008, 453(7194):524-U526.

36. Yu PZ, Pan GJ, Yu JY, & Thomson JA. FGF2 Sustains NANOG and Switches the Outcome of BMP4-Induced Human Embryonic Stem Cell Differentiation. *Cell Stem Cell*, 2011, 8(3):326-334.

37. Storey JD & Tibshirani R. Statistical significance for genomewide studies. *P Natl Acad Sci USA*, 2003, 100(16):9440-9445.

38. Beausoleil SA, Villen J, Gerber SA, Rush J, & Gygi SP. A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nat Biotechnol*, 2006, 24(10):1285-1292.

39. Schroeder MJ, Shabanowitz J, Schwartz JC, Hunt DF, & Coon JJ. A neutral loss activation method for improved phosphopeptide sequence analysis by quadrupole ion trap mass spectrometry. *Anal Chem*, 2004, 76(13):3590-3598.

40. Picotti P, Rinner O, Stallmach R, Dautel F, Farrah T, Domon B, Wenschuh H, & Aebersold R. High-throughput generation of selected reaction-monitoring assays for proteins and proteomes. *Nat Methods*, 2010, 7(1):43-U45.

41. Wenger CD, Phanstiel DH, Lee MV, Bailey DJ, & Coon JJ. COMPASS: A suite of pre- and post-search proteomics software tools for OMSSA. *Proteomics*, 2011, 11(6):1064-1074.

42. Kersey PJ, Duarte J, Williams A, Karavidopoulou Y, Birney E, & Apweiler R. The International Protein Index: An integrated database for proteomics experiments. *Proteomics*, 2004, 4(7):1985-1988.

Chapter 4

Multipurpose dissociation cell for enhanced ETD of intact proteoforms

CMR designed research, performed experiments, analyzed data, and wrote the paper.

This chapter has been published:

Rose C. M.*, Russell J. D.*, Ledvina A. R., McAlister G. C., Syka J. E. P., Westphall M. S., Griep-Raming J., Schwartz J. C., Coon J. J. "Multipurpose Dissociation Cell for Enhanced ETD of Intact Protein Species." *Journal of the American Society for Mass Spectrometry*, **2013**, 24(6):816-827.

*co-first author

Abstract

We describe and characterize an improved implementation of ETD on a modified hybrid linear ion trap-Orbitrap instrument. Instead of performing ETD in the mass-analyzing quadrupole linear ion trap (A-QLT), the instrument collision cell was modified to enable ETD. We partitioned the collision cell into a multi-section RF ion storage and transfer device to enable injection and simultaneous separate storage of precursor and reagent ions. Application of a secondary (axial) confinement voltage to the cell end lens electrodes enables charge-sign independent trapping for ion-ion reactions. The approximately two-fold higher quadrupole field frequency of this cell relative to that of the A-QLT, enables higher reagent ion densities and correspondingly faster ETD reactions, and, with the collision cell's longer axial dimensions, larger populations of precursor ions may be reacted. The higher ion capacity of the collision cell permits the accumulation and reaction of multiple full loads of precursor ions from the A-QLT followed by FT Orbitrap *m/z* analysis of the ETD product ions. This extends the intra-scan dynamic range by increasing the maximum number of product ions in a single MS/MS event. For analyses of large peptide/small protein precursor cations, this reduces or eliminates the need for spectral averaging to achieve acceptable ETD product ion signal-to-noise levels. Using larger ion populations, we demonstrate improvements in protein sequence coverage and aggregate protein identifications in LC-MS/MS analysis of intact protein species as compared to the standard ETD implementation.

Introduction

Electron capture and electron transfer dissociation (ECD and ETD, respectively), are unique in their proclivity to cleave the inter-residue N-C_α bonds of peptides with relative indifference to amino acid composition.[1-3] In addition, both reactions become faster and more efficient as precursor charge increases. These unique features have enabled researchers to study large, highly charged protein molecules that were previously difficult or impossible to characterize using dissociation methods involving vibrational activation.[4-7]

Electron transfer dissociation has been demonstrated on a variety of single and multiple mass analyzer instruments including quadrupole ion trap[1,2], triple quadrupole[8,9], quadrupole-time-of-flight (Q-TOF)[10-13], linear ion trap-Orbitrap[14,15], and quadrupole-FT-ICR instruments[16]. With this variety of instrument types, there has been a corresponding diversity in the ion confinement devices wherein the ETD reaction has been performed. ETD reactions have been performed in Paul Traps (3D RF quadrupole ion traps) [17-19], linear RF multipole ion traps (linear quadrupole and hexapole traps) [8,9,16], and RF ion pipes (RF stacked ring ion guides).[10] The more successful implementations of ETD, including all of the commercial implementations, involve trapping of reagent ions.

The types of traps used for ETD have primarily been determined by what pre-existing instrument ion path component could be conveniently and effectively modified and repurposed to perform ETD without compromising its usual function. This means ETD is mostly performed within ion traps whose design was optimized for mass analysis, or within ion guides (usually originally designed and optimized for use as collision cells) which are converted into linear ion traps by altering the applied voltages. The first implementation of ETD involving FT Orbitrap analysis of product ions utilized a modified commercial hybrid quadrupole linear ion trap–Orbitrap instrument, LTQ Orbitrap XL.[14] In that and subsequent commercial implementations of ETD on hybrid QLT-Orbitrap instruments, the ETD reaction was conducted in the quadrupole linear ion trap *m/z* analyzer (A-QLT) in a substantially identical manner to how it was performed with the corresponding commercial standalone quadrupole linear ion trap instruments. The overall approach has been little altered from that used for the original demonstration of

ETD.(2) Features of an ideal implementation of ETD would include efficient accumulation of both precursor cations and reagent anions, minimization of ETD reaction times, complete control of the initiation and subsequent quenching of ETD reactions, ability to expose precursor ions to IR activation before and during the ETD reaction, high trapping capacity for both precursor and reagent ions, retention of widest possible *m/z* range of product ions and efficient transfer of product ions to the *m/z* analyzer.

The introduction of a quadrupole collision cell into the designs of recent generations of commercial hybrid QLT-Orbitrap instruments [20] offers the opportunity to perform ETD in a device better suited for ion-ion reactions than a quadrupole ion trap of design specifically optimized for *m/z* analysis. Here, we describe a modified hybrid dual cell quadrupole linear ion trap-Orbitrap mass spectrometer, LTQ Orbitrap Velos ETD, (Thermo Fisher Scientific, Bremen Germany). As presented previously at various scientific conferences,[21,22] we modified the standard RF quadrupole collision cell to enable ETD and activated-ion ETD (AI-ETD) [23,24] as well as collision cell-type CID (HCD). This multi-dissociation cell (MDC) was created by partitioning the collision cell quadrupole ion guide into four segments with independently controllable DC biases. This partitioning of the cell enables injection and simultaneous separate storage of precursor cations and reagent anions. Provision was also made for the application of secondary RF (axial) confinement voltage to the cell end plate lens electrodes to enable charge-sign independent trapping (CSIT) during the ETD reaction. The approximately two fold higher operating quadrupole field frequency of this cell, relative to that of the A-QLT, enables higher reagent ion densities and correspondingly higher ETD reaction rates. The higher trapping field frequency, in combination with the collision cell's longer axial dimensions, permits reaction of larger populations of precursor ions. These modifications have not compromised the function of the MDC as a conventional collision cell. Conversion from operation of as a collision cell to an ETD reaction cell is accomplished solely by altering the voltages applied to the device's electrode.

In addition to facilitating faster ion-ion reactions, the MDC enables advanced scan functions; for example, multiple loads of precursor ions can be injected into the MDC, allowing for very large ion populations (~5×10^6 charges) to be accumulated prior to ETD activation. Similar scan functions have

recently been described, during which multiple loads of precursor cations were injected into an FT-ICR prior to ECD activation.[25-27] These scan functions are particularly well suited for top-down analysis, as analysis of even modest-sized whole proteins typically necessitates spectral averaging, requiring long acquisition times. Using the MDC, it is possible to instead use multiple cycles of ion injection, isolation, and transfer to the MDC, to accumulate a large precursor ion population. Following the multiple cycles of precursor accumulation, single ETD activation and *m/z* analysis events are performed. We demonstrate this procedure decreases the overall time required to obtain high quality ETD spectra of large peptides or proteins. We directly compare analytical figures of merit for ETD performed in the MDC against ETD performed in the A-QLT pertaining to product ion generation, signal-to-noise, ion statistics, MS/MS duty cycle, and top-down LC-MS/MS data quality and quantity. Our results illustrate substantial improvements in protein sequence coverage and aggregate protein identifications facilitated by interrogating large precursor populations using multi-dissociation cell ETD.

Results

Characterization of ETD in the Multi-dissociation Cell

The differing physical and operational parameters of the A-QLT and MDC which have the potential to cause difference in ETD functionality would be the use of different damping/collision gases and pressures (A-QLT: Helium at ~3×10^{-3} Torr, MDC: Nitrogen at > 1×10^{-3} Torr) [28], the axial dimensions of the two devices (A-QLT: ~62 mm, MDC: 133 mm), and quadrupole trapping field frequencies (A-QLT:1144 kHz , MDC: 2365 kHz).

The choice of nitrogen as the collision/damping gas rather than helium was driven by the desire to preserve the functionality of the MDC for collision cell type CID (HCD). Further the function of the C-Trap relies on being pressurized by nitrogen flowing from the MDC cell (or HCD cell). The damping/collision gas type and pressure potentially could influence the equilibrium reagent cloud radius and density, and therefore the reaction kinetics. Gas type and pressure could also influence the rate of collisional cooling of the intact charge reduced precursor ions after electron transfer and therefore influence

the partitioning in the generation of ETD and ETnoD products. We made no effort characterize the extent of such effects for various MDC nitrogen pressures. The fixed helium pressure in the high pressure cell (HPC) of the A-QLT (~3×10^{-3} Torr) was determined by the instrument's designers though co-optimization of ion injection and *m/z* isolation efficiency rather than any attribute specific to ETD. Similarly the setting of the MDC nitrogen pressures for our work was determined by co-optimization of transfer efficiency from the A-QLT and the efficiency of transfer to the C-Trap and Orbitrap.

The greater length of the MDC cell (~ 133 mm) relative to the length of the HPC of dual cell A-QLT (~62 mm) has both advantageous and disadvantageous aspects. For comparable trapping field conditions (ion Mathieu q-value/LMCO, field frequency, and damping gas type and pressure, etc.) increased axial dimension means that proportionally more reagent ions would be required to achieve a given reagent ion density, so injection times would be correspondingly be longer. However, longer axial dimension increases the path length for collisional stabilization of injected ions and can lead to higher injection efficiency. In this specific instance, the ion optics path between the MDC cell and the reagent source contained far fewer elements than the path between the reagent source and the A-QLT (**Figure 1E**). The efficiency of transfer of reagent ions was expected to be 2-4 fold higher to the MDC than to the HPC of the A-QLT. In our work the injection times/Reagent Target settings (as determined by and set for the A-QLT) required to reach maximal ETD reaction rates were comparable or slightly lower for the MDC than for A-QLT (see below).

The different number and length of the segments between the two devices has an influence on the total number of precursor ions available for ETD. For ETD in the A-QLT the precursor ions are sequestered (trapped) in the front section of the three section HPC quadropole while the reagent ions were injected into the center section of the HPC. The front section of the 3 section HPC quadrupole was ~ 12 mm long with an r_0 of ~4.75 mm. The relatively short length the front section (~12 mm) in combination with the device's r_0 (~4.75 mm) created a very restricted precursor trapping region (perhaps ~ 3 - 5 mm length) which limited the storage (sequestration) space charge capacity to 3×10^5 - 1×10^6 charges. During ETD reagent ion injection and accumulation in the rear section of the MDC, the precursor ions were sequestered in the front

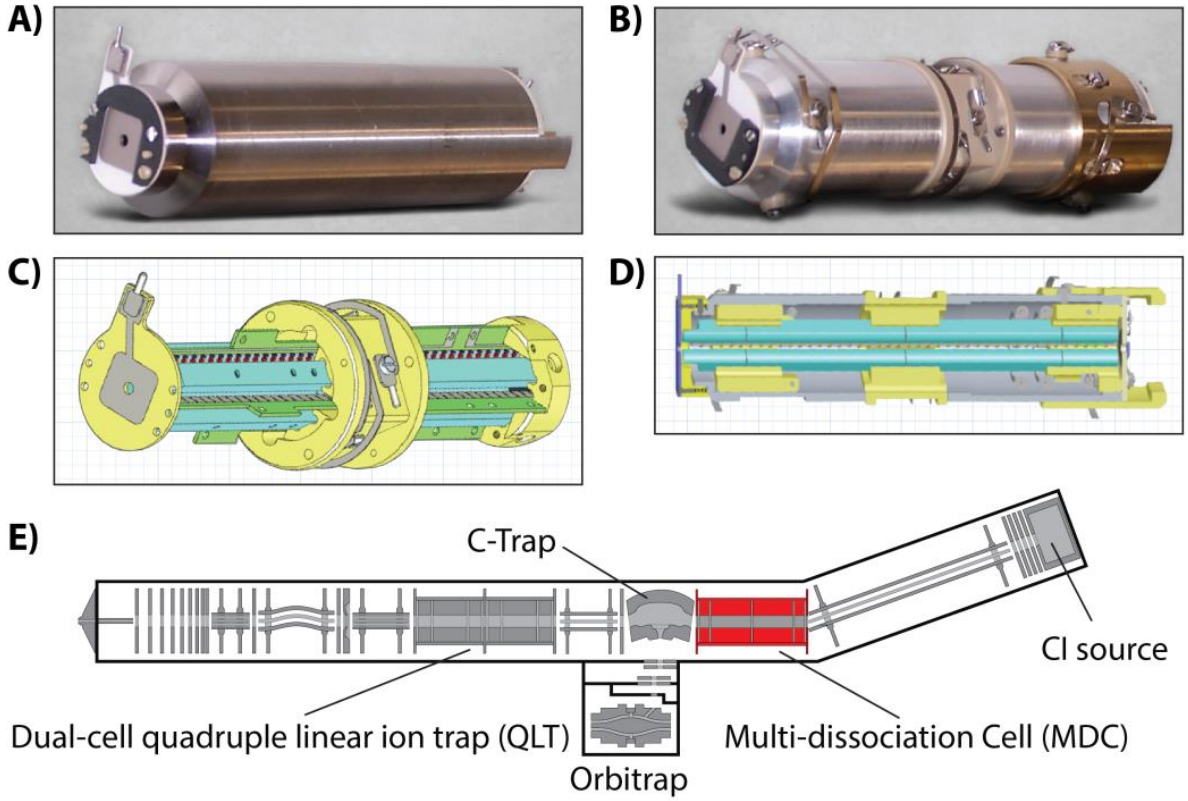


Figure 1. Multi-dissociation cell installed onto a hybrid QLT-Orbitrap mass spectrometer. In place of the pre-existing HCD cell (A), we installed a custom ion-ion reaction vessel (multi-dissociation cell, B-D). The multi-dissociation cell (MDC) comprises four separate sections, allowing for axial manipulation of ion populations. The MDC occupies the same space within the mass spectrometer as a normal collision cell (E).

two sections of the MDC (combined length ~68 mm). The storage space charge capacity of these combined sections of the MDC quadrupole probably exceeded 6×10^6 charges.

The different quadrupole trapping field frequencies would account for the large differences in the ETD functionality between the A-QLT and the MDC. When operated with identical low *m/z* cutoffs, the MDC, with its higher quadrupole field frequency, should provide stronger radial confinement of ions and, thus, enable higher reagent ion cloud density for the same number of trapped reagent ions per unit length. For ETD reaction, in either device, the ETD reagent ion population was always in sufficient excess of precursor ion population ions such that kinetics of the ETD reaction were pseudo first order. Under these conditions population of precursor $N_p(t)$ ions would be expected to decay with reaction time, t, as

$$N_p(t) = N_p(0)e^{-k[R]t} \qquad \textbf{(Equation 1)}$$

where [R] is the average number density of the portion of the reagent ion cloud that overlaps with the precursor ion cloud, and k is ETD ion-ion reaction rate constant. Based on the treatment of ion-ion kinetics of McLuckey and Stephenson [29,30], we expect that the rate constant, k, could be expressed as

$$k = c(|v|)\, Z_p^2\, Z_r^2\, ((m_p+m_{\mathbf{r}})/\, m_p\, m_{\mathbf{r}})^2 \qquad \textbf{(Equation 2)}$$

where Z_p, m_p ,Z_r and $m_{\mathbf{r}}$ are the charge states and masses of the precursor and reagent ion species respectively and the quantity $c(|v|)$ is a function of the distribution the magnitude of differential velocities, $|v|$, of precursor and reagent ions in the overlapping ion clouds. In both the A-QLT and MDC trapped ions are subject to large numbers of collisions with a collision/damping gas and are expected to assume near thermal velocity distributions [31], so we assumed that $c(|v|)$ was about the same for reactions in both devices. This is perhaps a dubious assumption as the factor $c(|v|)$ varies as $|v|^{-3}$ for overlapping ion clouds where $|v|$ is uniform (crossed mono-energic beams). Thermal velocities vary with mass ($m^{-½}$) so the rate constant, k, depends on m_p and $m_{\mathbf{r}}$ both explicitly and implicitly though $c(|v|)$. Since the mass of the reagent anion (202

Da for the Fluoranthene anion) was usually be considerable smaller than the precursor peptide or protein cation masses (such as the 3+ Angiotenisn I cations ($m/z_p \approx 433$ Th, $m_p \approx 1299$ Da), k is relatively weakly dependent upon the precursor ion mass.

As all known ETD reagent anions are singly charged, ETD reaction rates will be primarily influenced by the precursor charge state (quadratic dependence), reagent ion mass and the average number density of the portion of the reagent ion cloud that overlaps with the precursor ion cloud . For relatively low numbers of stored ions, the ion cloud diameter is determined by the kinetic energy of the trapped ions, the *m/z* of the ions and the trapping field intensity (for RF-only 2D quadrupole fields, this is compactly defined by ion's Mathieu q value) and the quadrupole trapping field frequency, f. For low numbers of trapped ions, implying the absence of significant effects due to mutual repulsion of ions (i.e., space charge effects), the average ion number density in the cloud scales proportionally with the number of trapped ions. Using the adiabatic theory of ion motion (pseudo-potential approximation) the maximum radius of ion motion, r_{max}, for a given ion kinetic energy, KE_{avg}, (averaged over an RF cycle) can be expressed as

$$r_{max} = (2\, KE_{avg}/m_{ion})^{1/2} /(\pi\, f\, q_{ion}) \qquad \textbf{Equation 3}$$

where q_{ion} and m_{ion} are the Mathieu q and mass of the ion respectively. This expression indicates that for a "low density" reagent ion cloud of fixed ion number with a near thermal distribution of kinetic energies and maintained at a fixed q in the field (i.e., operating with fixed LMCO), the cloud radius will vary as 1/f so the average number density of the ions in the cloud will vary as f^2. As the number of trapped ions are further increased, ion-ion repulsion will cause the ion cloud to expand and the average number density in the ion cloud will grow less than linearly with trapped ion number. Eventually, further increases in trapped ion numbers will be accommodated completely by growth in the ion cloud radius without significant increase in ion number density in bulk of the ion cloud.[31] The estimated saturation number density, $[N_{ion}]_{max}$, can be given as

$$[N_{ion}]_{max} = \varepsilon_0 \, m_{ion} \, (\pi \, f \, Z_{ion} \, e \, q_{ion})^2 \qquad \textbf{Equation 4}$$

where ε_0 is the permittivity of free space, *e* is the elemental charge, and Z_{ion} is the ion charge state. Therefore, ETD reaction rates, k[R], in both the A-QLT and MDC were expected to increase linearly with reagent ion number (as controlled by the Reagent Target setting) for relatively low reagent ion numbers and then saturate at higher reagent ion numbers. We systematically measured ETD reaction rates of the 3+ cation of Angiotensin I in both devices over wide range of Reagent Target settings. The low mass cutoff for all experiments was maintained at ~*m/z* 89 so the reagent ion Mathieu q was always ~ 0.4 during the reactions. The reaction rates were determined by a semi-logarithmic least squares fit of the recorded abundance of the surviving precursor *m/z* peak as a function of reaction time. Both the A-QLT data and MDC kinetics data were in good qualitative match to the predictions based on adiabatic theory – linear increase with Reagent Target eventually saturating at a maximal reaction rate. The saturation reaction rate for the A-QLT was ~50/sec. For the MDC the maximal reaction rate was ~ 130 /sec (**Figure 2**). According to adiabatic theory, the maximal ion densities should vary quadratically with trapping field frequency. We saw a ~2.6 fold increase in reaction rates, rather than the predicted 4.3 fold increase expected based on the MDC and A-QLT frequencies. A Mathieu q-value of 0.4 is at the far upper limit for applicability of the adiabatic model, so high populations of reagent ions may been subject to a fair amount of RF heating in the MDC resulting in increased reagent velocities and a corresponding reduction in the magnitude of the c(|*v*|) term. The extent of this effect would likely be amplified by the use of the higher molecular weight collision/damping gas (nitrogen rather than helium). This could account for the lower than predicted increase in saturation reaction rate.

When analyzing whole protein MDC ETD data it was evident that a larger amount of charge reduced precursor exists as compared to A-QLT ETD. We suspect this is from the presence of the nitrogen adduct of Fluoranthene (*m/z* 216) which preferentially performs PTR. This adduct is usually removed during the reagent anion selection step when performing A-QLT ETD, but because the MDC is situated after the C-trap there is no means for anion selection.

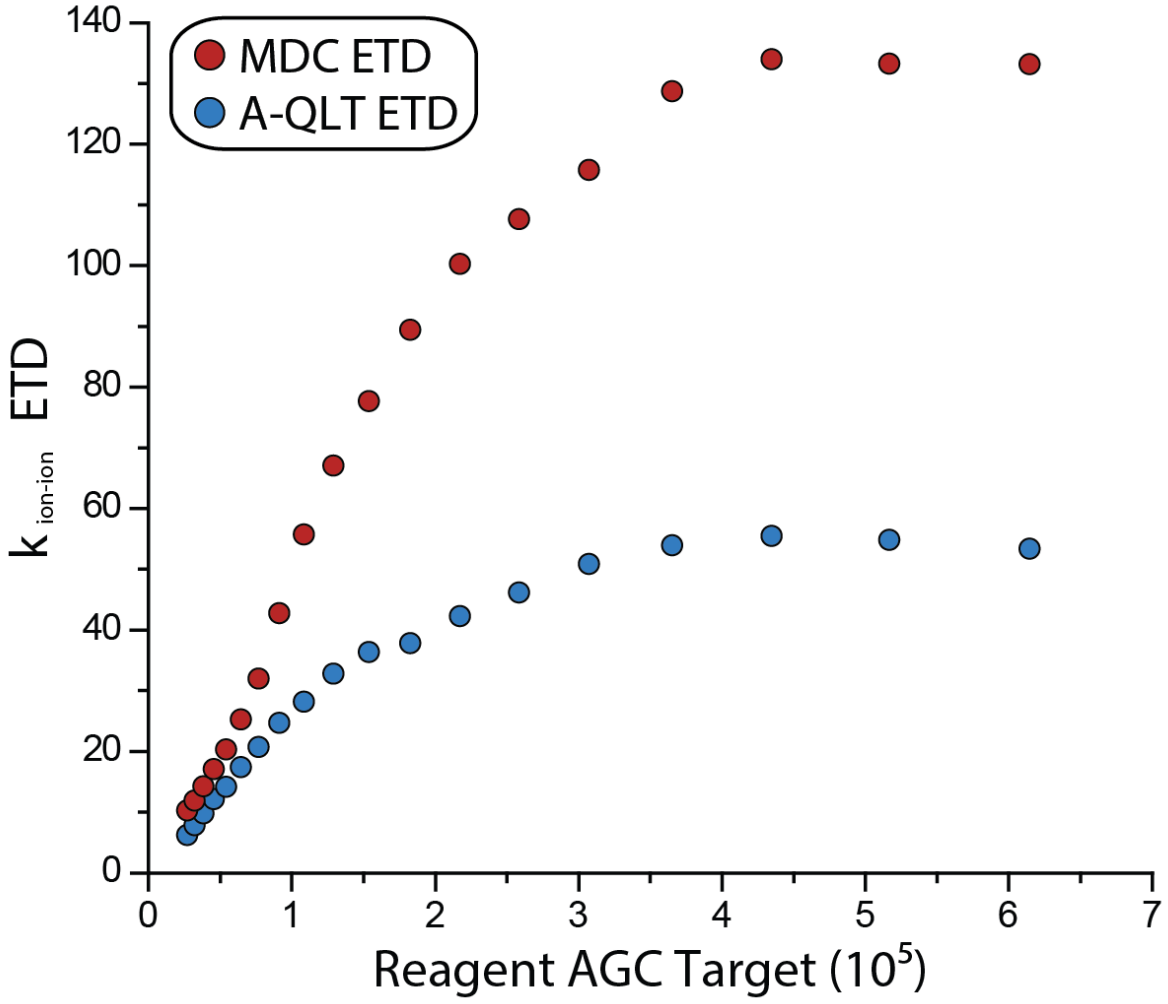


Figure 2. Reaction kinetics for ETD performed in the A-QLT or MDC.
MDC ETD demonstrates a higher reaction rate. ETD reactions were conducted with a precursor AGC target of $2X10^5$.

Characterization of Multiple Precursor Fills Scan Function

The isolation space charge limit in the A-QLT is ~ 5 x 10^5 charges when isolating with a 3 Th window.[32] While both the activation and storage space charge capacities are many-fold larger [33], the total charge of the precursor ion population is restricted to this value when ions are isolated and reacted in the in the same device. For Orbitrap FT analyses, 5x10^5 precursor charges are more than sufficient to produce peptide MS/MS spectra with high signal-to-noise; however, it is rarely adequate to produce quality MS/MS spectra of intact proteins without substantial signal (transient) averaging. A principle reason for this is the much larger number of potential backbone cleavages and the larger number of charge states for each the possible product ion structure. Further, product ions are of higher average molecular weight which means that the signal for each product ion structure may be distributed across several isotopic *m/z* peaks as well as between different charge states. For ETD, which is relatively non-selective in backbone cleavage points, this means that that the number of different dissociation product ions (*m/z* peaks) grows faster than linearly with precursor molecular weight. The number of expected product ion *m/z* peaks of from a 15000 Da precursor ion, is well more than 10 fold greater than what expected to be produced from a 1500 Da precursor ion. The MDC has an estimated storage space charge limit of 2 x 10^7 charges, accommodating a precursor population many-fold greater than what can be isolated and then reacted in the A-QLT. To build up a large precursor ion population in the MDC for MS/MS, multiple loads of protein precursors are iteratively isolated in the A-QLT and transmitted to the MDC; a process we refer to as multiple fills **Figure 3**). The ions are then activated in one ETD reaction with subsequent analysis in the Orbitrap.

To investigate the utility of multiple fills, recombinant human histone H3.3 was infused and the +15 charge state (*m/z* 1015.2) was subjected to ETD conducted in the A-QLT (A-QLT ETD) or MDC (MDC ETD) (**Figure 4**). As a starting point, H3.3 was analyzed using A-QLT ETD with a 5x10^5 MS/MS automatic gain control (AGC) target value with product ion analysis in the Orbitrap (60k resolving power, 6 ms activation) using 10 microscans (transient averaging). We then performed ETD in the MDC using 10 precursor fills (1 fill ≈ 5x10^5 charges) with product ion analysis in the Orbitrap using 1 microscan, or 10 fills with 4 Orbitrap microscans, corresponding to an equivalent per specturm number of precursor charges

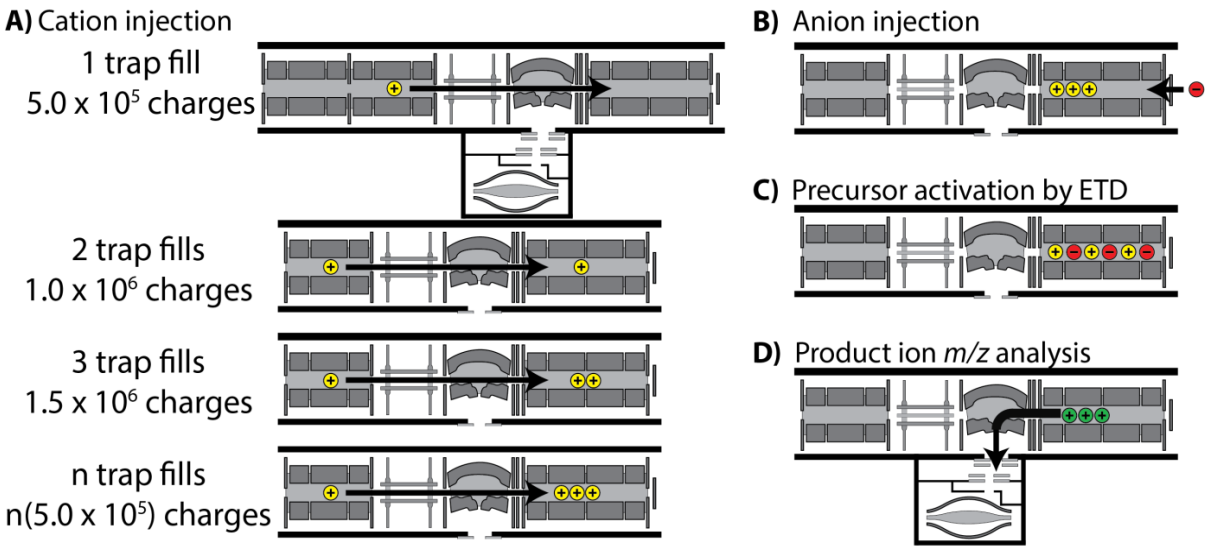


Figure 3. Multiple fills injection scheme used to build large precursor populations for ETD conducted in the MDC.

A precursor cation population equal to the MS/MS AGC target (e.g., $5x10^5$) is isolated in the A-QLT and then transmitted and stored in the MDC. This process is repeated until the desired population is reached in the MDC. The cations are sequestered in the front portion of cell followed by injection of the reagent anions into the back portion of the cell. ETD is carried out via charge-sign-independent trapping and product ions are analyzed in the Orbitrap.

($5x10^6$) or similar acquisition time (~9 s), respectively. The activation duration for MDC ETD (60 ms) was scaled to generate approximately the same relative amount of intact charge-reduced precursor to unreacted precursor as was observed in the reference A-QLT ETD spectra (**Figure 4B**). Note that MDC ETD activation could be reduced to 20-30 ms without appreciable decreases in product ion S/N (as computed by the data system). However, we determined that the ratio of charge-reduced to unreacted precursor was a fair strategy to normalize for reaction completeness. The scan cycle depicting precursor injection, activation, and product ion analysis for each ETD MS/MS experiment is given in **Figure 4A**. The major contributor to the MS/MS cycle is the time required to record the ion image current signal transient for product ion *m/z* analysis in the Orbitrap (60k res. = 768 ms).[34] High MS and MS/MS resolving power is required to facilitate precursor mass determination and product ion assignment from the abundant and highly charged species generated from ETD. To increase product ion S/N for A-QLT ETD, multiple transients could be averaged, but at the expense of a corresponding increase in time required to obtain the product ion spectrum. By comparison, only a small time penalty is incurred by using multiple fills in increase product ion S/N for MDC ETD as the time required to accumulate, transmit, and store a load of precursor ions is relatively short (~ 50 ms for a fill of $5x10^5$ charges) for these infusion experiments.

Product ions from the 600-900 *m/z* range in the histone spectra were examined in detail to directly compare ETD MS/MS approaches detailed above (**Figure 4B**). A-QLT ETD with 10 microscans produced 47 product ions ($S/N \geq 3$) in this region with an average S/N of 12. MDC ETD with 10 fills and 1 microscan generated 49 product ions ($S/N \geq 3$) with an average S/N of 14, while MDC ETD with 10 fills and 4 microscans produced 53 product ions ($S/N \geq 3$) with an average S/N of 27. For MDC ETD data quality was improved in terms of both product ion generation and S/N relative to A-QLT ETD. Superior S/N was achieved for MDC ETD with 10 fills and 4 microscans where more than half of the product ions had S/N > 16 (**Figure 4C**). MDC ETD with 10 fills and 1 microscan also produced a higher quality MS/MS spectrum than A-QLT ETD with 10 microscans in ~25% of the scan cycle time.

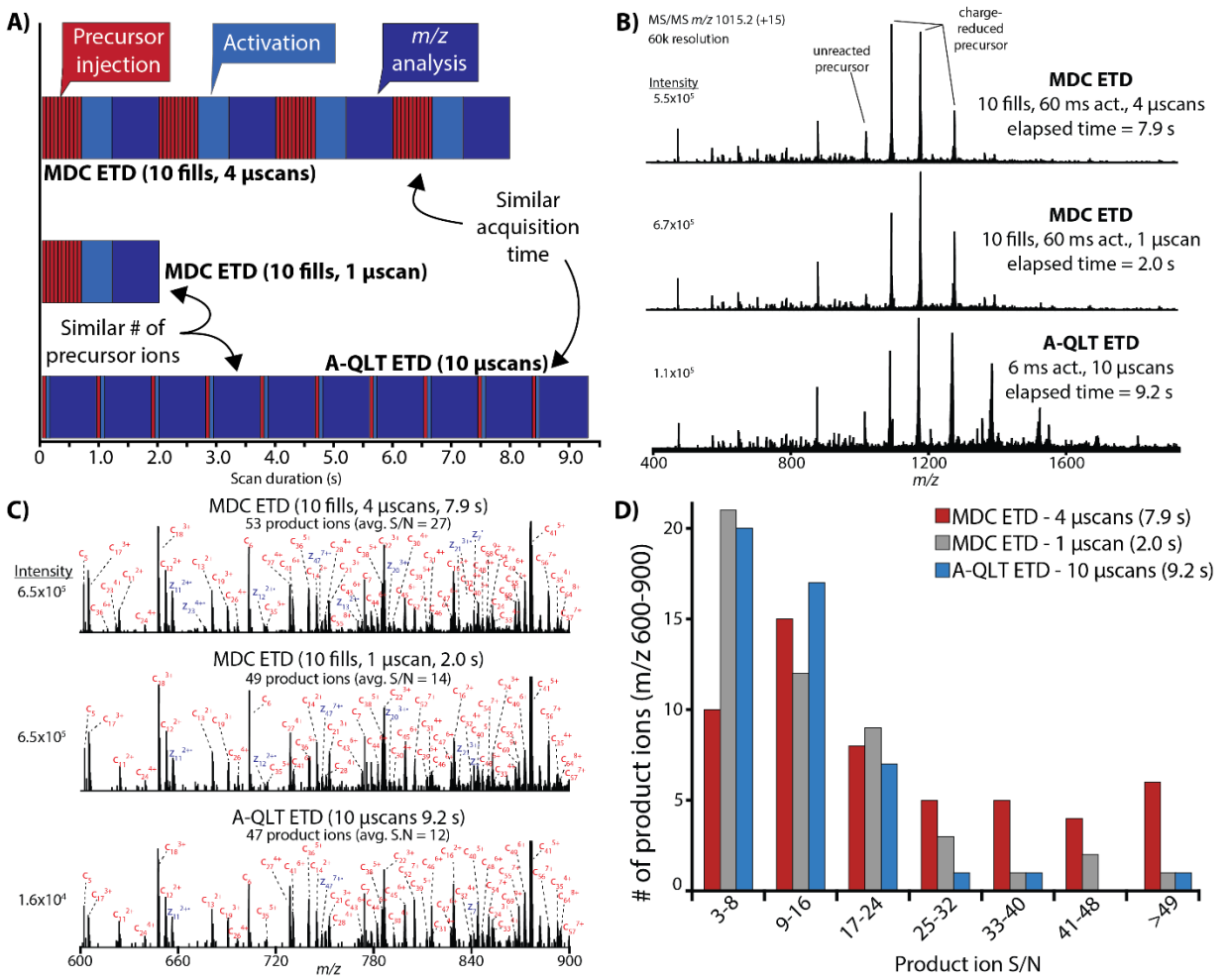


Figure 4. Comparison of A-QLT and MDC ETD for Histone H3.3.

(A) The major scan events contributing to the MS/MS duty cycle for conducting ETD in the MDC using multiple fills and ion trap ETD for histone infusion. The major scan events are divided into precursor injection ($5x10^5$ AGC target), precursor activation by ETD, and Orbitrap product ion analysis (60k resolution, 768 ms transient). **(B)** MS/MS spectra of histone H3.3 (+15, *m/z* 1015.2, 60k resolving power) analyzed using MDC ETD with 10 fills and 4 microscans (top), MDC ETD with 10 fills and 1 microscan (middle), and A-QLT ETD using 10 microscans. Reaction completeness for each ETD reaction was determined from the relative ratios of charge-reduced precursors to the unreacted precursor. A comparison of the top and bottom spectra represent similar acquisition times (7.9 s and 9.2 s), while a comparison of the middle and bottom spectra represent the same total number of precursors analyzed during the course of acquisition (~ $5x10^6$ charges). **(C)** Annotated portion of the histone H3.3 spectra. A product ion was considered a match if the S/N was > 3 and with a mass error less than 10 ppm. **(D)** Distribution of product ion S/N from (A).

Comparison of Multiple Precursor Fills and Transient Averaging

To characterize both the effect of the number of fills for MDC ETD and the effect of microscans for A-QLT ETD on product ion S/N, 10 representative product ions from the histone spectra were measured as the number of fills (with 1 microscan) or microscans were increased from 2 up to 10 (**Figure 5A**). The average S/N of A-QLT ETD product ions increased from 8.4 to 20.2. This observation is in agreement with what is predicted from signal averaging (S/N = 18.8) where S/N increases with the square root of the number of average measurements. The average S/N of MDC ETD product ions increased from 18.3 to 41.6. If the only source of noise was electrical in nature (thermal noise) then S/N should proportionally increase with each additional fill.[35-37] The expected average S/N should be 91.5 after 10 fills. However, chemical noise from electrospray ionization is a major contributor to noise as measured by the system and increases with each fill decreasing the expected S/N by more than half.[38] Yet, the increase in product ion S/N per each additional fill for MDC ETD was more than double than that from each additional microscan for A-QLT ETD. After 10 microscans and more than 9 s of acquisition, the average product ion S/N was 20.2. Comparable S/N (18.3) was realized in less than 1.4 s using MDC ETD with 2 fills. An additional fill boosted the S/N to 24.7 in less than 1.5 s of total acquisition.

There is limit to the number of fills that can be utilized for MDC ETD. The magnitude of the increase in product ion S/N diminished after 9 fills in this experiment. In general, the S/N increase generally levels off at ~ 10 fills for all MDC ETD experiments regardless of precursor attributes (charge state, *m/z*, length).We believe that this effect is due to the extraction space charge limit of the C-Trap because the we expect storage space charge capacities of both the MDC and the C-Trap to exceed the ~$1x10^7$ charges that would correspond to 10 fills.

Collecting microscans significantly lengthens the MS/MS duty cycle due to the time required to make multiple measurements of the time-domain signal transient. By comparison, the amount time required to accumulate additional fills is relatively small (**Figure 5B**). For the infusion experiments, each consecutive microscan added an additional 880 ms to the time required to produce a MS/MS spectrum (effective scan time), whereas each consecutive fill added 50 ms. The highest quality data, however, was

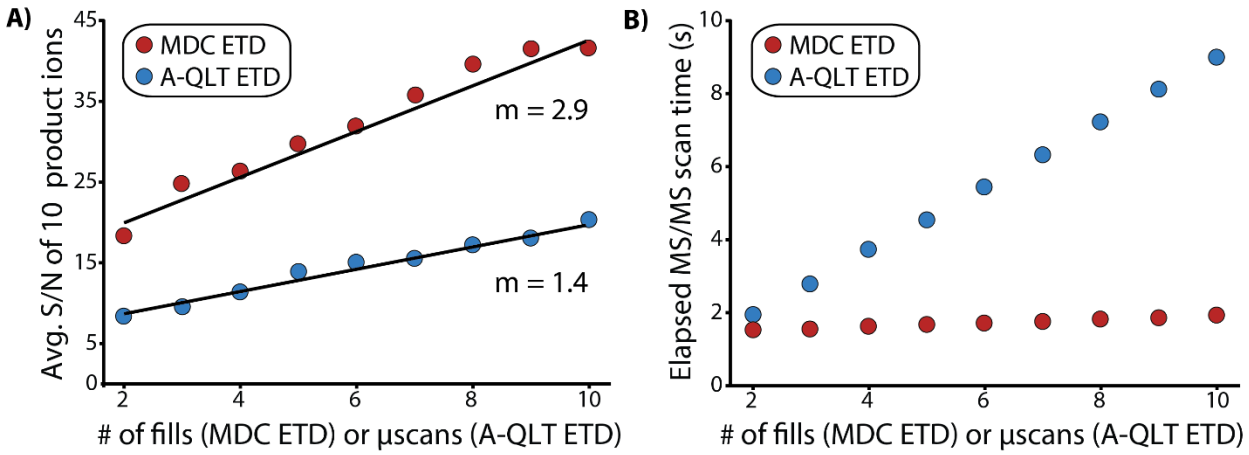


Figure 5. Comparison of transient averaging and multiple fills for selected product ions of Histone H3.3.

A) The average S/N of 10 representative product ions from **Figure 4** as a function of the number of microscans (A-QLT ETD) or fill (MDC ETD). **B**) The average elapsed MS/MS scan time to perform A-QLT ETD using microscans or MDC ETD using multiple fills. These values are from histone infusion experiments where the average time required to accumulate $5x10^5$ precursor charges for MS/MS was ~ 4.5 ms. The Orbitrap resolving power was set to 60k (768 ms transient).

achieved through the use of both multiple fills and transient averaging. Minimization of effective scan time is not a primary concern for protein infusion experiments. Effective scan times of more than 8 s are unacceptably long for LC analyses of complex mixtures. This is where the MDC ETD/multiple fills strategy has the greatest benefit.

Comparison of A-QLT and MDC ETD for Top-Down LC-MS/MS

Proteins extracted from yeast lysate were filtered to remove proteins greater than 30 kDa, the current upper mass limit for routine analysis on this instrument type. To provide a direct comparison of the analysis of similar numbers of precursor charges the sample was analyzed in duplicate over an 80-minute gradient using A-QLT ETD with 6 microscans or MDC ETD with 6 fills and 1 microscan. Performing the ETD reaction in the MDC resulted in a 132% increase in the number of protein-spectrum matches (PrSMs) identified, a 36% increase in the number proteoforms identified, and a 36% increase in the number of unique proteins (accession numbers) that could be confidently identified using ProSightPC (**Figure 6A**). Different charge states of the same proteoforms were identified multiple times leading to the large difference in the number of PrSMs and proteoforms that were identified. The average elapsed MS/MS scan times were 2.5 s and 7.6 s for MDC ETD and A-QLT ETD, respectively. Clearly, performing MDC ETD for intact protein analysis is advantageous for LC analyses due to for its shorter effective scan time.

In addition to shorter effective scan time, MDC ETD systematically produced spectra of higher signal-to-noise ratio. Spectral database correlation using ProSightPC relies upon the conversion of product ions to neutral monoisotopic masses through charge-state determination/deconvolution and deisotoping.[39-41] Critical to the correct assignment of monoisotopic masses is the ability to generate product ions with isotope distributions that match theoretical isotope distributions.[42-47] Determination of monoisotopic masses is confounded by low S/N and poor ion statistics. Low-level ions buried in the noise may never rise to a detectable level despite extensive transient averaging; however, these low-level ions may be boosted to a detectable level if the initial precursor ion population is sufficiently large as demonstrated in the histone infusion experiments. A portion of the ribosome spectrum generated by MDC ETD and A-QLT ETD is

A)

	PrSMs	Proteo-forms	Unique accession #
MDC ETD, 6 fills, 1 µscan	1,073	191	169
A-QLT ETD, 6 µscans	463	140	124

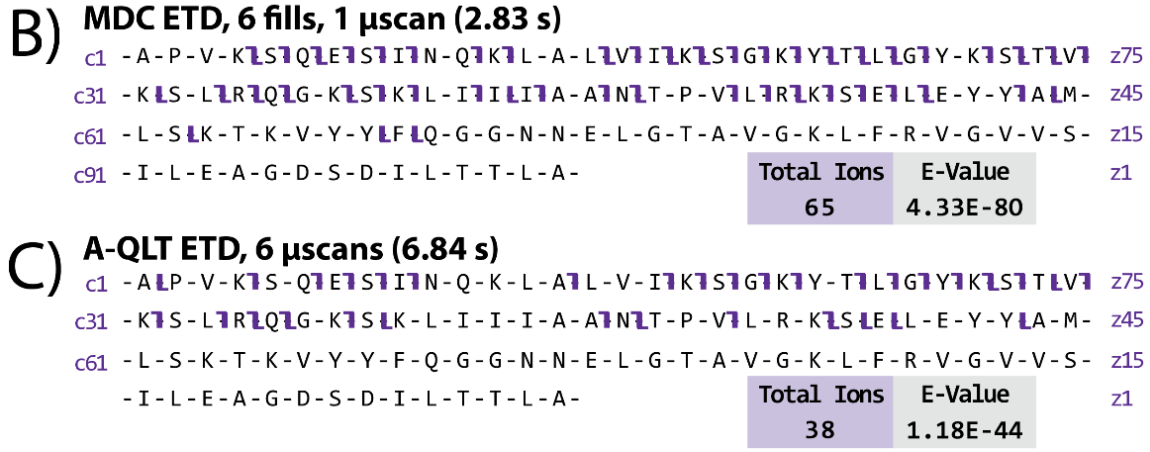


Figure 6. Results of LC-MS/MS analyses of yeast lysate enriched in proteins < 30 kDa interrogated with either MDC ETD or A-QLT ETD (reported values are the average of duplicates)

(**A**). MDC ETD facilitated the identification of more protein-spectrum matches (PrSMs), proteoforms, and unique protein accession numbers compared to A-QLT ETD. The 60S ribosomal protein L30 was identified with both MDC ETD (**B**) and A-QLT ETD (**C**). MDC ETD produced 71% more fragment ions leading to a better E-value and better sequence coverage compared to A-QLT ETD using ProSightPC 2.0 SP1 for raw file processing and database searching.

shown in **Figure 7A**. The isotope distribution from two product ions, c_{24}^{3+} and c_{46}^{5+}, were compared to the theoretical isotope distribution of each ion based upon the elemental composition of each (**Figure 7B**). In both cases, product ions from MDC ETD have a greater number of isotopic peaks appearing above the instrument's applied S/N threshold and their relative heights more closely matched those of the theoretical distributions. The analysis of a larger product ion population, rather than performing transient averaging, led to better ion statistics and increased the intrascan dynamic range. Improved ion statistics and intrascan dynamic range also improve the automated detection of isotopic clusters and aid in the ability to perform peak subtraction of overlapping product ion isotopic clusters to make use of all of analytical information available in the MS/MS spectrum.[45]

For example the 60S ribosomal protein L30 (11.3 kDa, accession # A6ZUW4) was identified with both MDC ETD and A-QLT ETD at 1.41 and 1.44 ppm mass error, respectively. Of the 206 possible protein backbone cleavages resulting in the production of c- and z˙-type product ions, 65 product ions were identified from the MDC ETD MS/MS spectrum compared to 38 for A-QLT ETD (**Figure 6B-C**). The expectation-value (E-value) score for the MDC ETD spectrum is substantially improved leading to a more confident match.

Discussion

By performing the ETD reactions in a multi-dissociation cell (MDC) – combination ETD and HCD cell - rather than in the analyzing quadruple linear ion trap (A-QLT) we have enabled ETD of 6-10 fold larger initial populations of precursor ions. Further the ~2 fold higher operating frequency of the MDC enables 2.4 fold higher reaction rates for reactions performed at comparable low mass cutoffs (LMCO). We have demonstrated an advanced scan function utilizing multiple cycles of precursor injection and isolation in the A-QLT followed by transfer to and accumulation in the MDC, There this aggregated ion population is subjected to ETD and then extracted to the Orbitrap for high resolution *m/z* analysis. Using this multi-fill approach for LC-ETD-MS/MS analysis of intact proteins yields more protein identifications with higher confidence due to the increased number and higher quality of MS/MS spectra. Further studies will continue

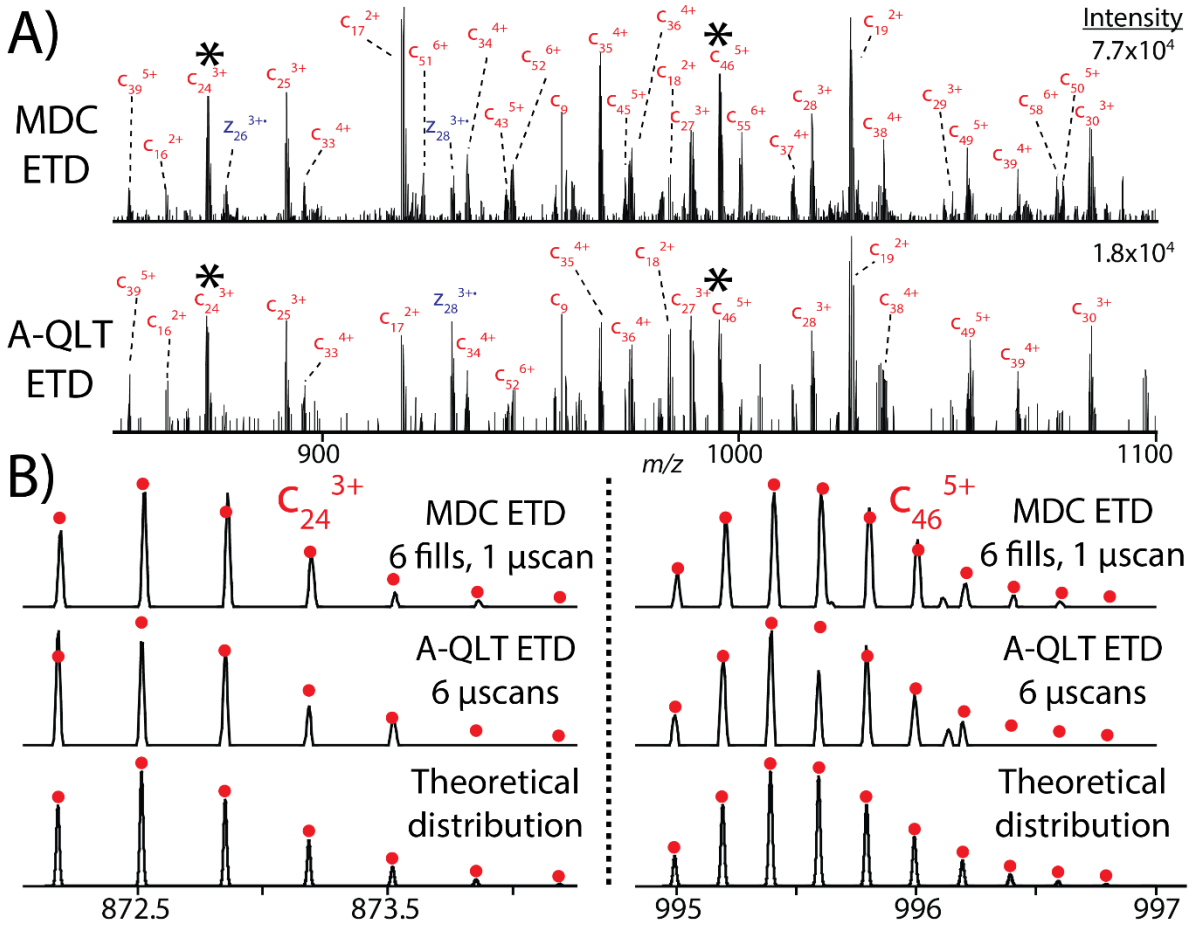


Figure 7. Comparison of yeast 60S ribosomal protein fragmentation.
A) Annotated portion of the yeast 60S ribosomal protein L30 MS/MS spectrum generated by MDC ETD (6fills, 1 microscan, top) and A-QLT ETD (6 microscans, bottom). **B**) The analysis of the isotope distributions of the two product ions demonstrate that the improved ion statistics engendered by A-QLT ETD lead to a larger intrascan dynamic range with isotope distributions more closely matching theoretical distributions modeled in Xcalibur.

a comparison of microscans (A-QLT ETD) and multiple fills (MDC ETD) on newer generation Orbitrap instruments that provide higher resolution for equivalent transient collection times.[34,48]

Experimental Procedures

Mass Spectrometry Instrumentation

All experiments were performed on an ETD-enabled, hybrid, dual cell-quadrupole ion trap-Orbitrap mass spectrometer (Velos-Orbitrap, Thermo Fisher Scientific, San Jose, CA). The primary modification was the installation of a modified multi-dissociation cell, MDC, in place of the pre-existing collision (HCD) cell (**Figure 1**). The MDC retained the basic layout of the HCD cell it replaced, with its quadrupole ion guide having the same length and r_0 (2.75 mm). The rod electrodes of the MDC quadrupole were divided into four sections with lengths, starting from end of the closest to the C-Trap, of 19 mm, 48.45 mm, 47.65 mm and 16.85 mm respectively. New quadrupole electronics were installed which could deliver sufficiently high RF voltages at 2,365 kHz to provide for a low *m/z* cutoff of ~90 Th (Mathieu q= 0.4 for *m/z* 202) during ETD reactions. Each section of the MDC quadrupole was driven by a separate filer of the of the new RF system's tuned circuit transformer coil permitting the sections to have independently controlled DC biases whilst having common applied RF voltages. An additional RF electronics system was used to provide RF up to ~ 500 $V_{0\text{-}P}$ at ~1222 kHz to the MDC end lens plate electrodes to provide the axial confinement for charge sign independent trapping (CSIT) of ions during ETD reactions. To support the control of the MDC device's additional voltages and to enable newly possible scan functions, we extensively modified the instrument control code.

As there was no provision on the instrument for a direct pressure measurement of the MDC cell pressure, MDC cell pressure was set and monitored according its contribution to the Orbitrap chamber pressure as measured by a Penning ionization gauge (delta pressure). Typical MDC Nitrogen pressure/flow settings for ETD experiments corresponded to a delta pressure of ~0.3×10^{-10} Torr for kinetics studies and ~0.1×10^{-10} Torr for protein analyses.

For MS/MS events using FT analysis, the A-QLT performed most of its usual functions, namely accumulation, trapping and isolation of precursor cations. Subsequent to isolation, cations were transmitted to the MDC. This sequence could be repeated several times, such that multiple populations of selected precursor cations would be delivered to the MDC. This multiple fill scheme increased the overall number of precursor ions subjected to ETD and thus product ions analyzed in each Orbitrap FT *m/z* analysis event. Following accumulation, cation precursors were sequestered in the front sections of the MDC, while the DC voltages of the back sections were set to positive values suitable for accumulation of ETD reagent anions. Anions generated in the reagent ion source were transmitted to and accumulated in the rear sections of the MDC. ETD was initiated by applying axial confining RF voltages to the end lenses of the MDC, and by setting all the DC bias voltage of the MDC to 0 V. The ETD reaction was quenched by setting the center two sections to negative DC offsets. Using the identical procedure for analysis of CID product ions produced in the collision cell, the ETD product ions were then transferred to the C-Trap and injected to the Orbitrap for FT *m/z* analysis.

In all experiments the radical anion of Fluoranthene (*m/z* 202) was the primary ETD reagent ion. There was no provision for *m/z* selection of the reagent ions for ETD in the MDC so the Fluoranthene radical anion comprised only ~95 % of the reagent ions delivered to the MDC for ion-ion reactions. The other 5% of the reagent ions were the Fluoranthene Nitrogen ion molecule reaction product anion, *m/z* 216, and various lower level background anions. The lack of purity of the reagent ion population would led to somewhat more prominent intact charge reduced product *m/z* peaks produced by proton transfer reactions with the reagent impurity anions. **Figure 8** provides evidence of this occurring as reaction of the +15 Histone H3.3 precursor in the MDC (10 Multiple Fills, 60 ms reaction time) produces a higher abundance of +13 and +14 charge-reduced precursors relative to the unreacted precursor than the same precursor is subjected to A-QLT ETD (10 microscans, 6 ms reaction time).

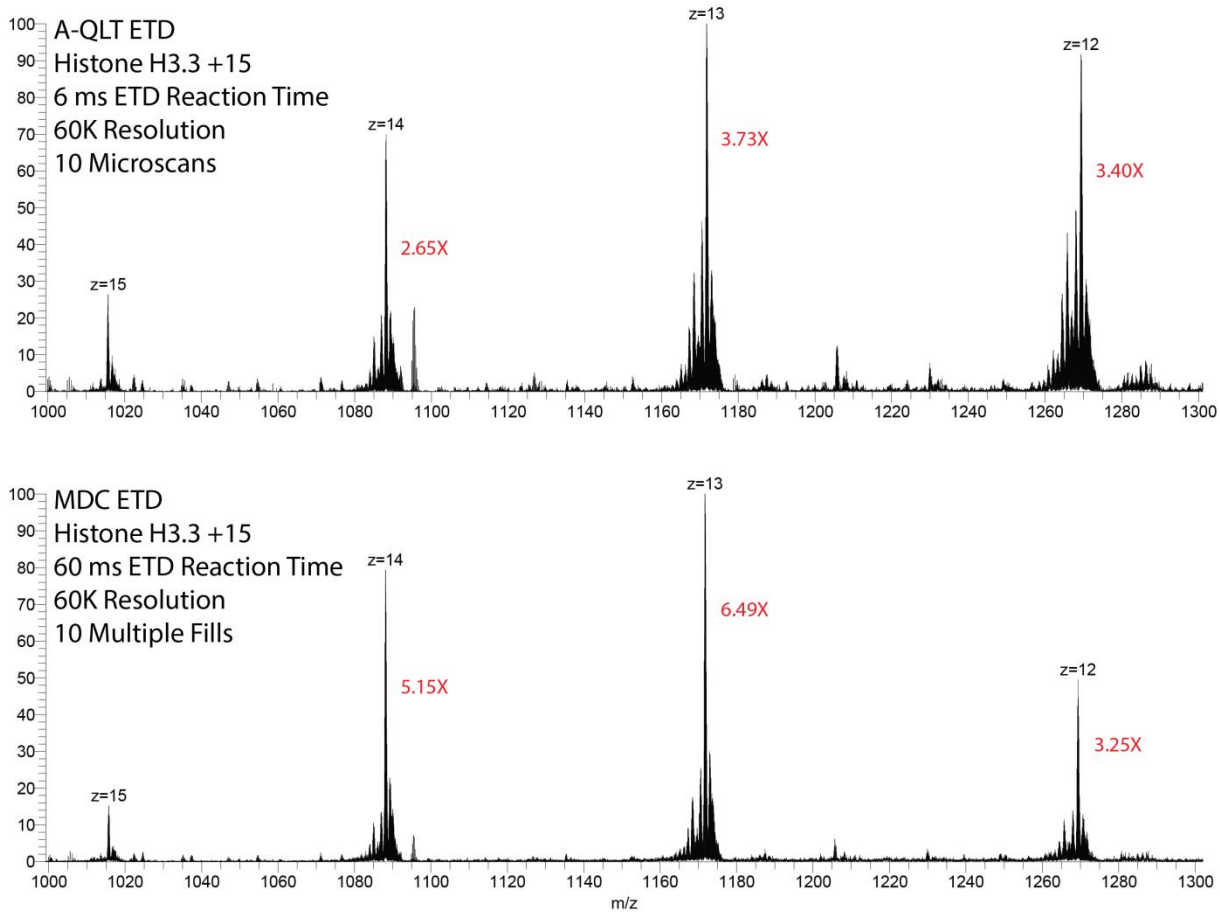


Figure 8. Demonstration of higher proton transfer in MDC ETD.
The +15 precursor ion of Histone H3.3 was reacted in the A-QLT or MDC and the ratio of charge reduced precursor to unreacted precursor is displayed in red.

Sample preparation

Recombinant, isotopically labeled ($^{13}C_2$-leucine) human histone H3.3 was acquired through collaboration with New England BioLabs Inc. (Ipswich, MA). Histones were desalted using 50 mg C_8 Sep-Pak cartridge (Waters Corp., Milford, MA) and reconstituted in 50% ACN with 0.2% formic acid to a concentration of 5 pmol/µL for direct infusion experiments. Wild-type *Saccharomyces cerevisiae* was cultured, lysed, and protein was extracted as previously described.[49] Protein estimation on the supernatant was performed by BCA and diluted to 5 ug/uL with lysis buffer and stored at -80°C. Protein lysate was reduced with dithiothreitol (DTT) at 37 °C for 30 min and alkylated with iodoacetamide for 45 min in the dark. Protein aliquots (30 µL diluted to 300 µL with lysis buffer) were passed through a 30 kDa nominal molecular weight centrifugal cut-off filter (Microcon YM-30, Millipore Corp., Billerica, MA) to deplete high-molecular weight proteins. The eluate (~250 µL) was buffer-exchanged against 3 volumes of 50 mM phosphate buffer (pH 7.2) in 5% acetonitrile using a 5 kDa nominal molecular weight centrifugal cut-off filter (Spin-X® UF, Corning Life Science, Corning, NY). The retentate (~ 50 µL), enriched in proteins between approximately 5-30 kDa, was diluted to an estimated 1 µg/µL in exchange buffer.

MS/MS parameters for LC analyses

Orbitrap MS^1 analyses were performed with the 60k resolution setting and 3 microscans (transient averaging) with an AGC target value of $1x10^6$ charges. Orbitrap MS/MS analyses with ETD conducted in the ion trap were performed at the 60k resolution setting and 6 microscans (transient averaging) with an AGC target value of $5x10^5$ charges. The ETD reagent AGC target value was $5x10^5$ charges. Orbitrap MS/MS analyses with ETD conducted in the MDC were performed at the 60k resolution setting and 1 microscans (no transient averaging) and 6 fills were used with each fill representing ~ $5x10^5$ charges (effective AGC target value of $3x10^6$ charges). The ETD reagent AGC target value was $5x10^5$ charges and the injection time calculated to reach this value was used for both A-QLT and MDC ETD; however, the actual reagent population in the MDC was estimated to be ~ $1.0x10^6$ – $1.5x10^6$ charges. The increase was related to improved transmission efficiency to the MDC relative to the A-QLT due to considerably fewer

ion optical elements between the reagent ion source and the MDC than between the reagent ion source and the A-QLT.

Charge-state dependent reaction time scaling was custom coded into the instrument firmware with reaction times determined from prior LC scouting runs and optimized reaction times from protein infusion experiments. For example, for A-QLT ETD, a +5, +10, and +15 precursor populations were reacted for 17.5 ms, 5.5 ms, and 4.0 ms, respectively. The same charge states were reacted for 50 ms, 28 ms, and 16 ms, respectively, using MDC ETD. Unassigned charge states were allowed to be sampled for MS/MS, but +1, +2, and +3 charge states were excluded. Unassigned precursor charge states were assumed to be highly charged and were reacted for 3.5 ms (QLT ETD) or 6 ms (MDC ETD). MDC ETD reaction times are ~ 3 – 4X longer than A-QLT ETD reaction times because a larger number of ions (6-10X) are reacted during MDC ETD. However, this additional reaction time is negligible when comparing multiple fills to microscans as each additional microscan consists of an ETD reaction followed by the collection of a transient (768 ms at 60k resolving power).

The 2 most intense precursors were selected by data-dependent analysis with dynamic exclusion enabled (exclusion time: 30 s, repeat count: 1, exclusion width: +/- 1.5 Th). The isolation width for both ETD experiments was 5 Th to capture the entire precursor isotopic cluster and to limit isolation space charge effect. The first mass was fixed at *m/z* 225. The maximum MS and MS/MS ion injection times were set to 300 ms and 750 ms, respectively. The MS/MS max injection time was relatively long, as compared to peptide analysis, but only ~ 8% of scans reached this value. Raw data files are available for download at http://www.chem.wisc.edu/~coon/Downloads/MDC%20Data/.

Liquid chromatography

Micro-capillary columns with an integrated ESI emitter were manufactured as previously described. [50] Analytical columns were prepared from fused silica tubing (360 μm o.d. x 75 μm i.d.) slurry-packed to 15 cm in length with Magic C18AQ, 5 μm, 300 Å particles (Michrom Bioresources Inc., Auburn, CA). Precolumns, with cast chemical frits, were made from 360 μm o.d. x 75 μm i.d. fused silica and slurry-

packed with 5 cm of the same material used for analytical columns.[51] Gradient elution was achieved using a nanoACQUQLTY Ultra Performance LC® system (Waters Corporation, Milford, MA) with analytical flow rates of 300 nL min^{-1}. Mobile phase A consisted of 0.2% formic acid and mobile phase B contained 99.8% ACN/0.2% formic acid. Sample concentration/desalting onto the precolumn was carried out at 1.5 µL/min for 10 min with 5% mobile phase B. The gradient began at 10% B and increased to 25% B at 30 min. The gradient then linearly increased from 25%B to 50%B at 80 min.

Automated database searching

Automated database searching was performed using ProSightPC 2.0 SP1 (Thermo Fisher).[41,52] Spectra were converted to neutral monoisotopic masses using the Xtract algorithm only considering product ions with S/N greater than 2.5 with a maximum charge of 20. Top-down spectra were searched against the "top_down_simple" *Saccharomyces cerevisiae* database downloaded from the ProSightPC database website (ftp://prosightpc.northwestern.edu/). The database contained 15,303 basic protein sequences with 29,050 protein forms. Searches were performed by first using the Absolute Mass search mode with a 3.0 Da precursor and 15 ppm fragment ion tolerances using monoisotopic masses. The minimum number of matching fragments to be considered for identification was set to 10. Spectra that failed to produce a protein-spectrum match or matches with E-value scores less than 10^{-4} were searched in Biomarker mode with 10 ppm precursor and fragment ion tolerances with a minimum of 7 fragment ions for a protein-spectrum match. Include Modified Forms was enabled along with the Δm option in Biomarker mode.

Acknowledgments

The authors gratefully acknowledge support from Thermo Fisher Scientific, NSF CAREER grant 0747990, and NIH grant R01 GM080148. C.M.R was funded by an NSF Graduate Research Fellowship and NIH Traineeship (T32GM008505).

References

1. Syka JEP, Coon JJ, Schroeder MJ, Shabanowitz J, & Hunt DF. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proceedings of the National Academy of Sciences of the United States of America*, 2004, 101(26):9528-9533.

2. Coon JJ, Syka JEP, Schwartz JC, Shabanowitz J, & Hunt DF. Anion dependence in the partitioning between proton and electron transfer in ion/ion reactions. *International Journal of Mass Spectrometry*, 2004, 236(1-3):33-42.

3. Zubarev RA, Kelleher NL, & McLafferty FW. Electron capture dissociation of multiply charged protein cations. A nonergodic process. *Journal of the American Chemical Society*, 1998, 120(13):3265-3266.

4. Kelleher NL. Top-down proteomics. *Anal. Chem.*, 2004, 76(11):196A-203A.

5. Chi A, Bai DL, Geer LY, Shabanowitz J, & Hunt DF. Analysis of intact proteins on a chromatographic time scale by electron transfer dissociation tandem mass spectrometry. *Int. J. Mass Spectrom.*, 2007, 259(1-3):197-203.

6. Tsybin YO, Fornelli L, Stoermer C, Luebeck M, Parra J, Nallet S, Wurm FM, & Hartmer R. Structural Analysis of Intact Monoclonal Antibodies by Electron Transfer Dissociation Mass Spectrometry. *Analytical Chemistry*, 2011, 83(23):8919-8927.

7. Fornelli L, Damoc E, Thomas PM, Kelleher NL, Aizikov K, Denisov E, Makarov A, & Tsybin YO. Analysis of Intact Monoclonal Antibody IgG1 by Electron Transfer Dissociation Orbitrap FTMS. *Molecular & cellular proteomics : MCP*, 2012, 11(12):1758-1767.

8. Wu J, Hager JW, Xia Y, Londry FA, & McLuckey SA. Positive Ion Transmission Mode Ion/Ion Reactions in a Hybrid Linear Ion Trap. *Analytical Chemistry*, 2004, 76(17):5006-5015.

9. Liang X, Hager JW, & McLuckey SA. Transmission Mode Ion/Ion Electron-Transfer Dissociation in a Linear Ion Trap. *Analytical Chemistry*, 2007, 79(9):3363-3370.

10. Rand K, Pringle S, Morris M, Engen J, & Brown J. ETD in a Traveling Wave Ion Guide at Tuned Z-Spray Ion Source Conditions Allows for Site-Specific Hydrogen/Deuterium Exchange Measurements. *Journal of the American Society for Mass Spectrometry*, 2011, 22(10):1784-1793.

11. Xia Y, Chrisman PA, Erickson DE, Liu J, Liang X, Londry FA, Yang MJ, & McLuckey SA. Implementation of Ion/Ion Reactions in a Quadrupole/Time-of-Flight Tandem Mass Spectrometer. *Analytical Chemistry*, 2006, 78(12):4146-4154.

12. Xia Y, Thomson BA, & McLuckey SA. Bidirectional Ion Transfer between Quadrupole Arrays: MSn Ion/Ion Reaction Experiments on a Quadrupole/Time-of-Flight Tandem Mass Spectrometer. *Analytical Chemistry*, 2007, 79(21):8199-8206.

13. Huang T-y & McLuckey SA. Top-down protein characterization facilitated by ion/ion reactions on a quadrupole/time of flight platform. *PROTEOMICS*, 2010, 10(20):3577-3588.

14. McAlister GC, Berggren WT, Griep-Raming J, Horning S, Makarov A, Phanstiel D, Stafford G, Swaney DL, Syka JEP, Zabrouskov V, & Coon JJ. A Proteomics Grade Electron Transfer Dissociation-Enabled Hybrid Linear Ion Trap-Orbitrap Mass Spectrometer. *Journal of Proteome Research*, 2008, 7(8):3127-3136.

15. McAlister GC, Phanstiel D, Good DM, Berggren WT, & Coon JJ. Implementation of Electron-Transfer Dissociation on a Hybrid Linear Ion Trap−Orbitrap Mass Spectrometer. *Analytical Chemistry*, 2007, 79(10):3525-3534.

16. Kaplan DA, Hartmer R, Speir JP, Stoermer C, Gumerov D, Easterling ML, Brekenfeld A, Kim T, Laukien F, & Park MA. Electron transfer dissociation in the hexapole collision cell of a hybrid quadrupole-hexapole Fourier transform ion cyclotron resonance mass spectrometer. *Rapid Communications in Mass Spectrometry*, 2008, 22(3):271-278.

17. Hartmer R & Lubeck M. New Approach for Characterization of Post Translational Modified Peptides Using Ion Trap MS with Combined ETD/CID Fragmentation. *LC-GC Europe*, 2005, 18:11-13.

18. Sobott F, Watt SJ, Smith J, Edelmann MJ, Kramer HB, & Kessler BM. Comparison of CID Versus ETD Based MS/MS Fragmentation for the Analysis of Protein Ubiquitination. *Journal of the American Society for Mass Spectrometry*, 2009, 20(9):1652-1659.

19. Pitteri SJ, Chrisman PA, Hogan JM, & McLuckey SA. Electron Transfer Ion/Ion Reactions in a Three-Dimensional Quadrupole Ion Trap: Reactions of Doubly and Triply Protonated Peptides with SO2•. *Analytical Chemistry*, 2005, 77(6):1831-1839.

20. Olsen JV, Schwartz JC, Griep-Raming J, Nielsen ML, Damoc E, Denisov E, Lange O, Remes P, Taylor D, Splendore M, Wouters ER, Senko M, Makarov A, Mann M, & Horning S. A Dual Pressure Linear Ion Trap Orbitrap Instrument with Very High Sequencing Speed. *Molecular & Cellular Proteomics*, 2009, 8(12):2759-2769.

21. Russell JD, Ledvina AR, McAlister GC, Westphall MS, Syka JEP, Griep-Raming J, & Coon JJ. Single scan, top-down intact protien analysis of a Velos Orbitrap modified with a dedicated high-capacity ion/ion reaction cell. . *Procceedings of the 59th ASMS Conference on Mass Spectrometry and Allied Topics, Denver, CO*, 2011.

22. Yang C, Russell JD, Ledvina AR, Westphall MS, Brumbaugh J, & Coon JJ. Profiling histone H3 isoforms in Human embryonic stem cells using a Velos Orbitrap modified with a segmented reaction cell. *Procceedings of the 59th ASMS Conference on Mass Spectrometry and Allied Topics, Denver, CO*, 2011.

23. Ledvina AR, Beauchene NA, McAlister GC, Syka JEP, Schwartz JC, Griep-Raming J, Westphall MS, & Coon JJ. Activated-Ion Electron Transfer Dissociation Improves the Ability of Electron Transfer Dissociation to Identify Peptides in a Complex Mixture. *Anal. Chem.*, 2010, 82(24):10068-10074.

24. Ledvina AR, McAlister GC, Gardner MW, Smith SI, Madsen JA, Schwartz JC, Stafford GC, Syka JEP, Brodbelt JS, & Coon JJ. Infrared Photoactivation Reduces Peptide Folding and Hydrogen-Atom Migration following ETD Tandem Mass Spectrometry. *Angew. Chem.-Int. Edit.*, 2009, 48(45):8526-8528.

25. Schaub TM, Hendrickson CL, Horning S, Quinn JP, Senko MW, & Marshall AG. High-performance mass spectrometry: Fourier transform ion cyclotron resonance at 14.5 tesla. *Anal. Chem.*, 2008, 80(11):3985-3990.

26. Tsybin YO, Witt M, Baykut G, & Hakansson P. Electron capture dissociation Fourier transform ion cyclotron resonance mass spectrometry in the electron energy range 0-50 eV. *Rapid Commun. Mass Spectrom.*, 2004, 18(14):1607-1613.

27. Hakansson K, Chalmers MJ, Quinn JP, McFarland MA, Hendrickson CL, & Marshall AG. Combined electron capture and infrared multiphoton dissociation for multistage MS/MS in a Fourier transform ion cyclotron resonance mass spectrometer. *Anal. Chem.*, 2003, 75(13):3256-3262.

28. McAlister GC, Phanstiel DH, Brumbaugh J, Westphall MS, & Coon JJ. Higher-energy Collision-activated Dissociation Without a Dedicated Collision Cell. *Molecular & Cellular Proteomics*, 2011, 10(5).

29. Compton PD, Strukl JV, Bai DL, Shabanowitz J, & Hunt DF. Optimization of Electron Transfer Dissociation via Informed Selection of Reagents and Operating Parameters. *Anal. Chem.*, 2012, 84(3):1781-1785.

30. McLuckey SA, Stephenson JL, & Asano KG. Ion/Ion Proton-Transfer Kinetics: Implications for Analysis of Ions Derived from Electrospray of Protein Mixtures. *Analytical Chemistry*, 1998, 70(6):1198-1202.

31. Tolmachev AV, Udseth HR, & Smith RD. Modeling the ion density distribution in collisional cooling RF multipole ion guides. *International Journal of Mass Spectrometry*, 2003, 222(1):155-174.

32. McAlister GC, Phanstiel DP, & Coon JJ (2008) A dual reaction cell, ETD-enabled orbitrap mass spectrometer for top-down proteomics. *56th Proceedings of the American Society for Mass Spectrometry and Allied Topics Conference*.

33. Schwartz J, Senko M, & Syka J. A two-dimensional quadrupole ion trap mass spectrometer. *J. Am. Soc. Mass Spectrom.*, 2002, 13(6):659-669.

34. Michalski A, Damoc E, Lange O, Denisov E, Nolting D, Müller M, Viner R, Schwartz J, Remes P, Belford M, Dunyach J-J, Cox J, Horning S, Mann M, & Makarov A. Ultra high resolution linear ion trap Orbitrap mass spectrometer (Orbitrap Elite) facilitates top down LC MS/MS and versatile peptide fragmentation modes. *Mol. Cell. Proteomics*, 2012, 11(3):pages numbers not available.

35. Makarov A, Denisov E, Kholomeev A, Balschun W, Lange O, Strupat K, & Horning S. Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. *Anal. Chem.*, 2006, 78(7):2113-2120.

36. Perry RH, Cooks RG, & Noll RJ. Orbitrap mass spectrometry: Instrumentation, ion motion and applications. *Mass Spectrom. Rev.*, 2008, 27(6):661-699.

37. Alexander M (2010) Theory and practice of the Orbitrap mass analyzer. *Practical Aspects of Trapped Ion Mass Spectrometry, Volume IV,* (CRC Press), pp 251-272.

38. Cech NB & Enke CG. Practical implications of some recent studies in electrospray ionization fundamentals. *Mass Spectrom. Rev.*, 2001, 20(6):362-387.

39. Taylor GK, Kim Y-B, Forbes AJ, Meng F, McCarthy R, & Kelleher NL. Web and database software for identification of intact proteins using "top down" mass spectrometry. *Anal. Chem.*, 2003, 75(16):4081-4086.

40. LeDuc RD, Taylor GK, Kim Y-B, Januszyk TE, Bynum LH, Sola JV, Garavelli JS, & Kelleher NL. ProSight PTM: an integrated environment for protein identification and characterization by top-down mass spectrometry. *Nucleic Acids Res.*, 2004, 32(suppl_2):W340-345.

41. Zamdborg L, LeDuc RD, Glowacz KJ, Kim Y-B, Viswanathan V, Spaulding IT, Early BP, Bluhm EJ, Babai S, & Kelleher NL. ProSight PTM 2.0: improved protein identification and characterization for top down mass spectrometry. *Nucleic Acids Res.*, 2007, 35(suppl_2):W701-706.

42. Yergey JA. A general approach to calculating isotopic distributions for mass spectrometry. *Int. J. Mass Spec. Ion Physics*, 1983, 52(2–3):337-349.

43. Rockwood AL, Van Orden SL, & Smith RD. Rapid calculation of isotope distributions. *Anal. Chem.*, 1995, 67(15):2699-2704.

44. Senko MW, Beu SC, & McLafferty FW. Determination of monoisotopic masses and ion populations for large biomolecules from resolved isotopic distributions. *J. Am. Soc. Mass Spectrom.*, 1995, 6(4):229-233.

45. Horn DM, Zubarev RA, & McLafferty FW. Automated reduction and interpretation of high resolution electrospray mass spectra of large molecules. *J. Am. Soc. Mass Spectrom.*, 2000, 11(4):320-332.

46. Renard B, Kirchner M, Steen H, Steen J, & Hamprecht F. NITPICK: peak identification for mass spectrometry data. *BMC Bioinformatics*, 2008, 9(1):355.

47. Jaitly N, Mayampurath A, Littlefield K, Adkins J, Anderson G, & Smith R. Decon2LS: An open-source software package for automated processing and visualization of high resolution mass spectrometry data. *BMC Bioinformatics*, 2009, 10:87.

48. Tran JC, Zamdborg L, Ahlf DR, Lee JE, Catherman AD, Durbin KR, Tipton JD, Vellaichamy A, Kellie JF, Li M, Wu C, Sweet SMM, Early BP, Siuti N, LeDuc RD, Compton PD, Thomas PM, & Kelleher NL. Mapping intact protein isoforms in discovery mode using top-down proteomics. *Nature*, 2011, 480(7376):254-258.

49. Swaney DL, McAlister GC, & Coon JJ. Decision tree-driven tandem mass spectrometry for shotgun proteomics. *Nat Meth*, 2008, 5(11):959-964.

50. McAlister GC, Russell JD, Rumachik NG, Hebert AS, Syka JEP, Geer LY, Westphall MS, Pagliarini DJ, & Coon JJ. Analysis of the acidic proteome with negative electron-transfer dissociation mass spectrometry. *Anal. Chem.*, 2012, 84(6):2875-2882.

51. Ficarro SB, Zhang Y, Lu Y, Moghimi AR, Askenazi M, Hyatt E, Smith ED, Boyer L, Schlaeger TM, Luckey CJ, & Marto JA. Improved electrospray ionization efficiency compensates for diminished chromatographic resolution and enables proteomics analysis of tyrosine signaling in embryonic stem cells. *Anal. Chem.*, 2009, 81(9):3440-3447.

52. LeDuc RD & Kelleher NL (2007) Using ProSight PTM and related tools for targeted protein identification and characterization with high mass accuracy tandem MS data. *Curr. Protoc. Bioinform.*, (John Wiley & Sons, Inc.), pp 19:13.16.11-13.16.28.

Chapter 5

Activated ion ETD performed in a modified collision cell on a hybrid QLT-Orbitrap mass spectrometer

CMR designed research, performed experiments, analyzed data, and wrote the paper.

This chapter has been published:

Ledvina A. R.*, **Rose C. M.***, McAlister G. C., Syka J. E. P., Westphall M. S., Griep-Raming J., Schwartz J. C., Coon J. J. "Activated Ion ETD Performed in a Modified Collision Cell on a Hybrid QLT-Orbitrap Mass Spectrometer." *Journal of the American Society for Mass Spectrometry*, **2013**, 24(11):1623-1633.

*co-first author

Abstract

We describe the implementation and characterization of activated ion electron transfer dissociation (AI-ETD) on a hybrid QLT-Orbitrap mass spectrometer. AI-ETD was performed using a collision cell that was modified by partitioning the cell's quadrupole ion guide into a multi-section RF ion storage and transfer device enabling injection and simultaneous separate storage of precursor and reagent ions. Application of a secondary (axial) confinement voltage to the cell end lens electrodes enables charge-sign independent trapping and ETD. The instrument manifold was modified to enable irradiation of ions along the axis of this modified cell with IR photons from a CO_2 laser. Laser power settings were optimized for both charge (z) and mass to charge (*m/z*) and the instrument control firmware was updated to allow for automated adjustments to the level of irradiation. This implementation of AI-ETD yielded 1.6 fold more unique identifications than ETD in an LC-MS/MS analysis of tryptic yeast peptides. Furthermore, we investigated the application of AI-ETD on large scale analysis of phosphopeptides, where laser power aids ETD, but can produce b- and y-type ions due to the phosphoryl moiety's high IR adsorption. LC-MS/MS analysis of phosphopeptides derived from human embryonic stem cells using AI-ETD yielded 2.4 fold more unique identifications than ETD alone, demonstrating a promising advance in ETD sequencing of PTM containing peptides.

Introduction

Electron capture dissociation, ECD,[1] and electron transfer dissociation, ETD,[2,3] are advantageous over collisional methods of peptide fragmentation due to their tendency to randomly fragment peptides while preserving labile post-translational modifications (PTMs). This feature has enabled the interrogation of peptides, proteins, and PTMs that were previously difficult to characterize: including localization and characterization of sites of glycosylation[4-8], ADP-ribosylation[9,10], and phosphorylation[11-13].

As noted by ourselves and others, ETD exhibits low dissociation efficiency for high *m/z* (low charge density) peptide precursor cations.[14-18] Ideally, the ETD process is initiated by the transfer of an electron to the precursor cation from a reagent anion. This is followed by a radical rearrangement that, ultimately, results in the production of *c*- and *z*•-type ions. Sometimes, however, precursor peptides receive an electron but fail to separate into *c*- and *z*•-type ions. The probability of precursor peptides undergoing such non-dissociative electron transfer (ETnoD)[18] is elevated for high precursor *m/z* peptides. McLafferty and co-workers, having observed non-dissociative electron capture products in ECD experiments, posited that non-covalent interactions bind the dissociated *c*- and *z*•-type ion pairs together.[19] Practitioners of ECD have devised a number of techniques to mitigate the detrimental effects of non-dissociative electron capture: activated-ion ECD (AI-ECD). These techniques share the common strategy of disruption of non-covalent cation peptide intramolecular interactions via vibrational activation prior to electron capture.[20-25] This disruption leads to more efficient generation of *c*- and *z*•-type ions, and consequently higher dissociation efficiency.

Motivated by the activated-ion ECD methods, similar strategies have been pursued in order to enhance the dissociation product ion yield in ETD experiments. Increasing, the bath gas temperature of a quadrupole ion trap during ETD results in some disruption of peptide secondary structure, allowing for improved ETD efficiency.[26] Alternatively, ETnoD products can be dissociated post reaction into the respective *c*- and *z*•-type ions, using either ion trap-type, or beam-type collisional activation (termed ETcaD).[14,18] Though effective for increasing peptide sequence coverage, ETcaD techniques produce odd electron *c*-type ions (c•-) and even electron *z*-type ions. These product ions, both of which are shifted ~1

Da in mass of the expected *c* and *z*• product ion masses, result from the abstraction of a hydrogen atom from the *c*-type ion by the *z*•-type ion while the two ions are bound together by non-covalent interactions.[27] ETcaD spectra thus contain product ion *m/z* peak "clusters" which are the superposition of the isotopic envelopes of both odd, and even electron species that differ in mono-isotopic mass by 1 Da. This complicates both manual and automated spectral interpretation.

Recently, we have demonstrated a technique involving the vibrational activation of precursor ions with IR photons during the ETD reaction, activated-ion ETD (AI-ETD).[15,28] By continually disrupting peptide secondary structure, AI-ETD increases dissociation product ion yield (efficiency), while minimizing the opportunities for hydrogen transfer; thus resulting in the near-exclusive production of even electron *c*- and odd electron *z*•-type ions. Our initial AI-ETD work was with a standalone ETD equipped linear trap instrument. On that instrument the IR photon beam was transmitted through a window on the back of the instrument's vacuum manifold, through the reagent source, and directly down the axis of the analyzing quadrapole linear ion trap (A-QLT). Given the demonstrated merits of AI-ETD, it was only natural that we would seek to adapt AI-ETD our ETD capable hybrid QLT-Orbitrap instrument with its high resolution capabilities. Unfortunately, the mechanical design of the LTQ Orbitrap Velos ETD instrument was such that there was no comparably convenient way to illuminate the A-QLT's axis with an IR photon beam. The Orbitrap associated portion of the instrument was connected to the back of the ion trap analyzer vacuum chamber. We concluded that relocating the ETD reaction to collision cell modified to perform ETD would be the most mechanically expedient approach to adapting AI-ETD to this hybrid instrument.

To enable AI-ETD on a hybrid QLT-Orbitrap mass spectrometer the vacuum manifold and reagent transfer multipole were modified to allow irradiation of ions with IR photons from a CO_2 laser. The AI-ETD reaction is performed in a newly altered collision cell that contains four segments with individually controllable DC biases. A secondary RF voltage was also applied to the end lenses of the cell. These changes allow charge sign independent trapping (CSIT) of both precursor cations and reagent anions within the multi-dissociation cell (MDC), enabling ETD and AI-ETD to be performed without compromising the

function of the MDC as a conventional collision cell. Using this implementation of AI-ETD we demonstrate an increase in fragmentation efficiency, as well as a reduction of hydrogen abstraction as previously described on standalone ion trap systems.[15,28] In addition, implementation of AI-ETD on an Orbitrap system provides charge state (z) information for precursors, enabling automated control of ion irradiation to dissociate precursors, based on z and *m/z*, to achieve optimal product ion yield. Using real-time precursor ion data to determine the optimum laser power, we demonstrate that AI-ETD performed in the MDC identifies 1.6 fold more unique yeast tryptic peptides and 2.4 fold more unique human phosphopeptides than A-QLT ETD when used for LC-MS/MS analysis.

Results

AI-ETD Performed on a hybrid QLT-Orbitrap MS

Traditionally, ETD has underperformed when interrogating precursors with high *m/z* values (i.e., low charge densities). Peptide secondary structure, more prevalent with increasing precursor *m/z*, can bind newly formed *c*- and $z^{\bullet}$-type product ions following electron transfer, preventing the formation of sequence informative fragment ions (ETnoD). Early work in our group sought to address this shortcoming through the use of resonant-excitation of the ETnoD species to induce separation of the bound *c*- and $z^{\bullet}$-type product ions (ETcaD). The major deficiency associated with the ETcaD approach is that, while they are bound by peptide secondary structure, the $z^{\bullet}$-type ion may abstract a hydrogen atom from the *c*- type ion, shifting the *m/z* value of each product ion by ~ 1 Da, and confounding spectral interpretation. An alternative approach is to immerse precursor peptides in IR photons during the ETD reaction (Activated-Ion ETD, AI-ETD).[15,28] In so doing, we continually disrupt secondary structure, and improve ETD fragmentation efficiency. Moreover, AI-ETD spectra show little evidence of hydrogen abstraction, leading to improved responsiveness to automated searching algorithms.[28]

Prior to the present study, AI-ETD has been limited to implementation on stand-alone ion trap systems. For many applications, particularly whole protein characterization and large-scale peptide analysis, high resolution *m/z* analysis has had a transformative effect.[29] To enable this combination, we

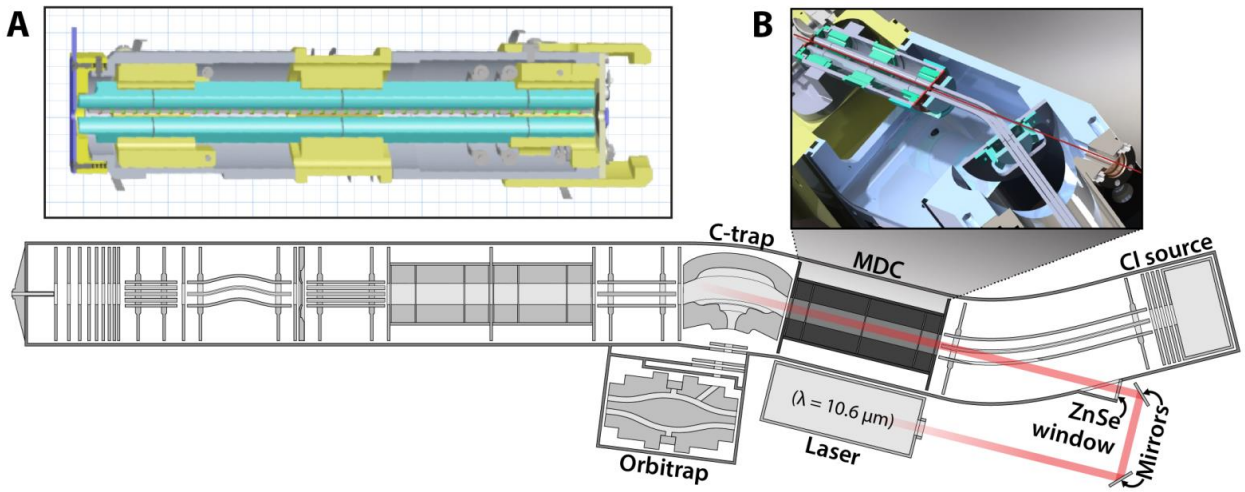


Figure 1. Modified LTQ-Orbitrap Velos mass spectrometer

A) Schematic of the modified collision cell (multi dissociation cell, MDC) capable of performing ETD which replaced the usual collision cell. B) Adaptations to instrument to allow AI-ETD. In addition to the installation of the MDC, we have excavated a photon passage through the transfer multipole, which conducts anions from the chemical ionization (CI) source to forward sections of the mass spectrometer. Using external mirrors and a ZnSe window, we enable the irradiation of the trapping volume of the MDC with IR photons generated using an external laser.

utilized a modified collision cell comprising four sections with independently controllable DC biases (**Figure 1A**). This change allows for simultaneous and separate storage of precursor and reagent ions. The end lenses of the collision cell were modified for the application of secondary RF (axial) confinement voltage, enabling charge-sign independent trapping (CSIT) and ETD. This modified collision cell has been termed the multi-dissociation cell (MDC) as it enables ETD and AI-ETD in addition to remaining a collision cell for beam-type CAD. The mass spectrometer manifold was modified such that photons were introduced to the MDC through a ZnSe window (**Figure 1B**). A hole was excavated in the transfer multipole, which conducts anions from the CI source to either the A-QLT, or the MDC (**Figure 1B**). This modification enables the immersion of ion-ion participants in IR photons, allowing for AI-ETD. The instrument firmware was then modified to trigger the external IR laser upon commencement of ETD within the MDC.

To demonstrate the application of AI-ETD on a hybrid LTQ-Orbitrap instrument the doubly protonated peptide RPKPQQFFGLM was infused and interrogated with ETD or AI-ETD performed in the MDC (**Figure 2**). Upon introduction of IR photons nearly all fragment ions increased in intensity when this peptide is activated for the same time, 60 ms (**Figure 2**). Specifically, c_6 dramatically increased when fragmented by AI-ETD, while c_4, c_5, and $z^{\bullet}_9$ became visible (**Figure 2B**). In addition to increased fragmentation, AI-ETD reduced hydrogen abstraction caused by non-covalent interactions. The fragment ion c_8 displays a peak at 1045.58 *m/z* representing the $c^{\bullet}_8$ ion resulting from abstraction of a hydrogen (**Figure 2A**). Activation by AI-ETD relieved this hydrogen abstraction as the amount of $c^{\bullet}_8$ was negligible as compared to the proper c_8 ion at 1046.58 *m/z* (**Figure 2B**). These data demonstrate the updated implementation of AI-ETD still improves fragmentation and reduces hydrogen abstraction, with the added benefit of high mass accuracy *m/z* analysis of fragment ions.

Optimization of Laser Power Decision Tree Values

AI-ETD was first implemented on a standalone linear trap instrument, which cannot determine charge states in normal scan modes. This limited all AI-ETD reactions to the same reaction time and laser power setting for all peptides regardless of z and *m/z*. High resolution MS^1 analysis yields charge state

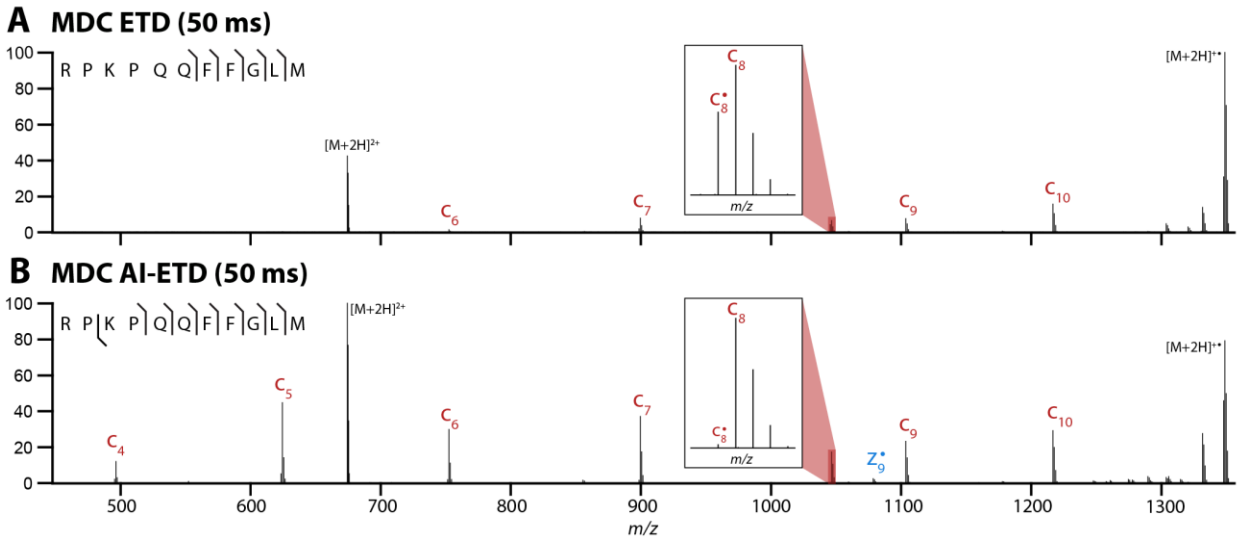


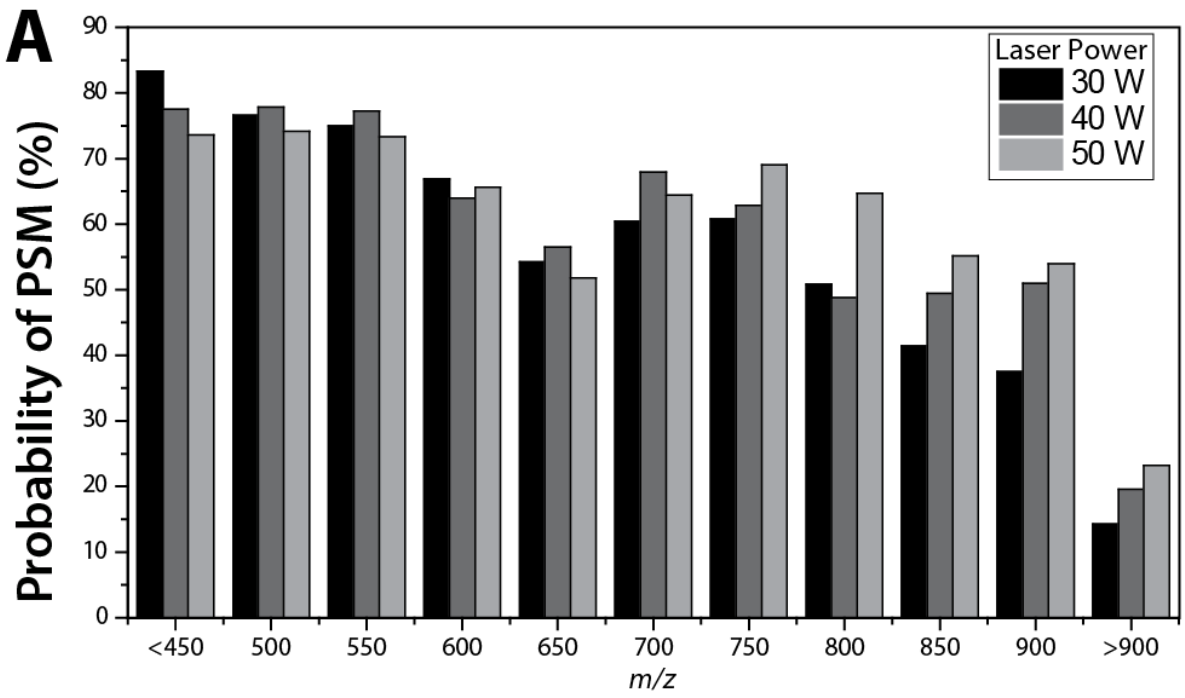
Figure 2. AI-ETD of the doubly charged peptide RPKPQQFFGLM.
AI-ETD performed in the MDC (B) improves fragmentation of the peptide RPKPQQFFGLM as compared to MDC ETD alone (A). AI-ETD also reduces hydrogen abstraction which can inhibit manual and automated spectral interpretation (insets).

information, enabling construction of decision tree logic to apply laser irradiation at differing levels depending on z and *m/z*.

To determine the laser power setting for decision tree analysis, yeast tryptic peptides were analyzed in nine separate LS-MS/MS experiments analyzing precursors of charge state 2, 3/4, or 5+ at laser powers of 30, 40 or 50 W. Identifications resulting from these experiments were grouped into 50 *m/z* bins from 450 to 900 and the probability of identification for each group was determined by dividing the number of identifications by the number of precursors selected in the given bin. **Figure 3A** depicts the result of this analysis for all triply charged precursors. Here, low *m/z* precursors are fragmented best with laser powers of 30 or 40 W, however, for peptides greater than 700 *m/z* 50 W becomes the most successful laser power setting. Using the probability of identification for each charge state and *m/z* bin we constructed a decision tree logic that mapped laser power to the peptide z and *m/z* (**Figure 3B**). The instrument control firmware was updated to enable control of laser power in real-time corresponding to these values. This decision tree logic was used for large scale comparison of AI-ETD to ETD using yeast tryptic peptides.

High Resolution AI- ETD Analysis of Yeast Tryptic Peptides

High resolution AI-ETD utilizing the MDC is more amenable to large scale analysis of peptides as compared to A-QLT ETD as MDC AI-ETD is faster and provides better fragmentation for high *m/z* precursors. To demonstrate improved identification of peptides for large scale analysis two LC-MS/MS experiments were performed using a ~120 minute LC gradient to separate a limited tryptic digest of yeast protein extract. In one experiment, the A-QLT was used to conduct ETD; in the other experiment, we used the MDC to conduct AI-ETD. Using A-QLT ETD, 4643 peptide spectral matches (PSMs), at a 1% false-discovery rate (FDR) were produced, corresponding to 2368 unique peptides (**Figure 4A**). Using the MDC and conducting AI-ETD, we obtained 8256 PSMs, correlating to 3789 unique peptides (**Figure 4A**). MDC AI-ETD, identified nearly all unique peptides obtained using conventional ETD, while also identifying a large number of peptides that were not identified using conventional ETD (**Figure 4B**). The ~80% increase in the number of PSMs is due to the use of the MDC increasing the number of MS/MS scans, and IR



B

z \ *m/z*	< 450	450 - 500	500 - 550	550 - 600	600 - 650	650 - 700	700 - 750	750 - 800	800 - 850	850 - 900	>900
	Laser Power Settings (W)										
2	30	40	40	40	40	40	50	50	50	50	50
3	30	40	40	40	40	40	50	50	50	50	50
4	50	30	30	30	30	30	30	30	30	50	50
5	50	30	40	40	30	40	50	30	40	30	50
> 5	30	40	30	40	40	50	50	40	30	40	50

Figure 3. Construction of decision tree laser power settings.
Precursors of varying charge state were interrogated by AI-ETD with laser powers at 30, 40, and 50 W. The probability of identification for 50 *m/z* bins for +3 precursors demonstrates the need for higher laser powers at higher *m/z* (A). Similar plots for charge states 2, 3, 4, 5, and >5 were used to construct decision tree logic for AI-ETD laser power settings depending on precursor *z* and *m/z* (B).

activation increasing the probability of peptide identification for each individual MS/MS event. Use of the MDC enables faster ETD reactions, and requires less time to accumulate reagent anions relative to the A-QLT; the result is that the average scan time for MDC AI-ETD events is ~ 40 ms faster than A-QLT ETD events. This, in turn, enables the collection of ~1500 more MS/MS scans.

To examine the difference in identification rate, individual PSMs were binned by precursor *m/z*, and divided by the total number of spectral features sampled over the course of the analysis. The resulting data provide a measure of peptide identification rate, as a function of precursor *m/z*, for both conventional ETD, and MDC AI-ETD (**Fig 4C**). ETD performance drops quickly as precursor *m/z* increases, a phenomenon previously noted in the literature.[17] This diminishing performance is attributed to decreased peptide charge density, and higher magnitudes of gas-phase secondary structure, which impair separation of *c*- and *z*- type ions. By disrupting the secondary structure using AI-ETD, performance still diminishes somewhat with increasing precursor *m/z*; in general, however, performance is far more uniform across a wide range of precursor *m/z* values.

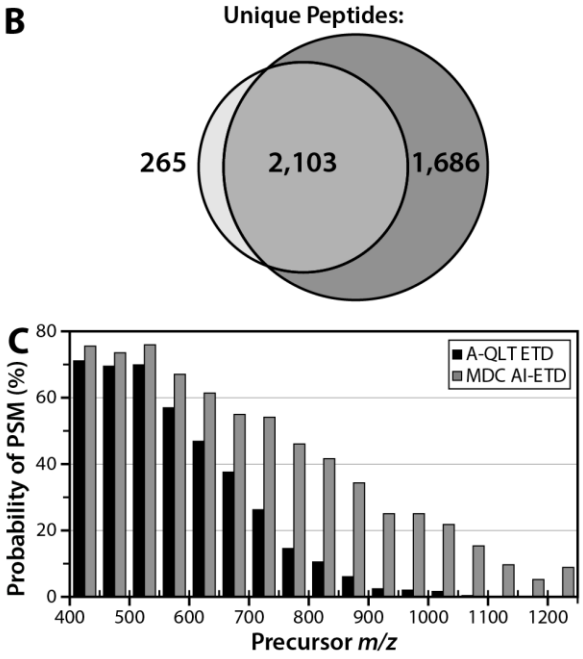
An example of the advantages offered by AI-ETD, as compared with ETD, for precursor peptides with high (> 800) *m/z* values, is shown in **Figure 4D**. ETD of triply protonated peptide SVEMHHEQLEQGVPGDNVGFNVK produces only two sequence informative ions, translating to 9% peptide sequence coverage. AI-ETD, in addition to requiring only 40% of the ion-ion reaction time relative to the A-QLT, results in the generation of 27 product ions, with peptide sequence coverage of 82%. Use of MDC and IR photons ultimately results in superior fragmentation in less time than conventional ETD performed in the A-QLT.

High Resolution AI- ETD Analysis of Phosphopeptides

There has been tremendous interest in the application of ETD to the analysis of phosphopeptides; the initial description of ETD included multiple MS/MS spectra, demonstrating the ability to localize sites of phosphorylation,[2] and a number of high-profile phosphorylation studies, utilizing ETD, have verified its importance to the field.[30-32] It is, therefore, important to evaluate the MDC and AI-ETD in the context of

A

	A-QLT ETD	SC AI-ETD
Unique peptides:	2,368	3,789
Avg. MS/MS time:	652 ms	616 ms
# of MS/MS scans:	16,230	17,661



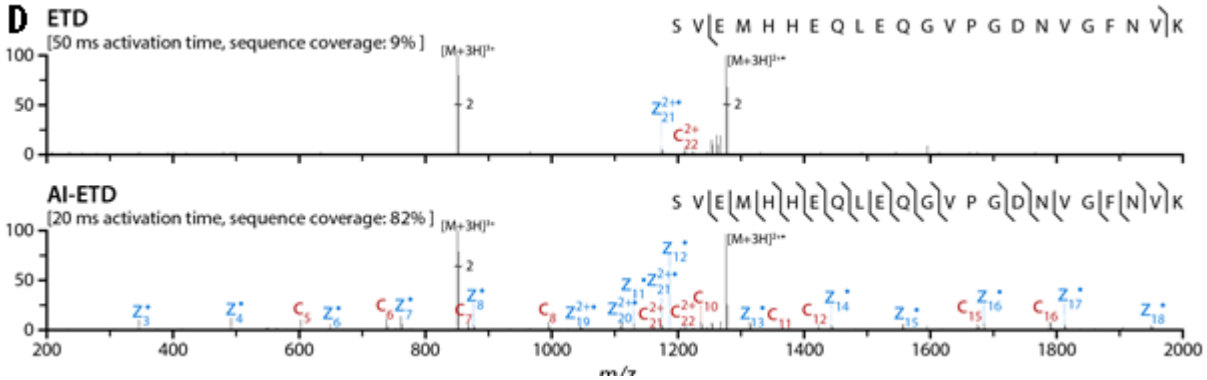


Figure 4. Comparison of A-QLT ETD and MDC AI-ETD.
Use of the MDC augments our implementation of AI-ETD because the reduced reaction time and reagent accumulation time requirements result in a shorter MS/MS time and more MS/MS scans over the course of the LC-MS/MS analysis (A). Analysis of the overlap of peptide identifications demonstrates AI-ETD identifies a large number of peptides not sequenced by A-QLT ETD. Using laser activation, the probability of producing a peptide spectral match (PSM) is greatly increased, particularly for peptide precursors having a high *m/z* value (C). D) ETD and AI-ETD of triply protonated SVEMHHEQLEQGVPGDNVGFNVK.

phosphopeptide analysis. An apparent caveat associated with AI-ETD of phosphopeptides, is the unintended production of *b*- and *y*- type ions. Such ions result from vibronic excitation and are not typically observed in the AI-ETD spectra of unmodified peptides. We rationalize that this is because phosphopeptides have much higher IR photon absorption efficiencies than unmodified peptides.[30,31]

To investigate this phenomenon, AI-ETD was performed at a number of different laser powers (20, 30, 40, and 50 W) for phosphopeptides containing one (LPISASHsEKTR) or two (LPIAASHssKTR) phosphorylated residues (**Figure 5**). As with unmodified peptides, irradiation with IR photons during ETD increased the fragmentation efficiency of peptides as compared to ETD alone (**Figure 5**). Analysis of the doubly charged forms of these standard peptides demonstrated increased adsorption of IR radiation, evidenced by increased vibronic-type and neutral loss ions, as the number of phosphorylated residues increased. For the singly phosphorylated peptide LPISASHsEKTR, AI-ETD appears to be most efficient at a laser power of 40 W (**Figure 5A**), resulting in 10 *c*- or $z^{\bullet}$-type fragment ions while maintaining a low percentage of *b*- and *y*-type ions. Conversely, a doubly phosphorylated peptide, LPIAASHssKTR, required lower level of irradiation to disrupt secondary structure (20 or 30 W) and exhibited a substantial amount of unwanted fragmentation at a laser powers of 40 and 50 W (**Figure 5B**). Increasing power to 50 W created additional fragment types (*a*- and *x*-type) as well as substantial neutral loss (NH_3, H_2O, and H_3PO_4) from all ion types as demonstrated by un-annotated peaks in the AI-ETD spectrum (**Figure 5B**). Using these data, a laser power of 35 W was used in large scale analysis of Human phosphopeptides, as it represented an effective compromise between detrimental formation of vibronic-type products and effective peptide secondary structure disruption.

High Resolution AI- ETD Analysis of Human ES Cell Phosphopeptides

To investigate the efficacy of MDC AI-ETD for large-scale phosphopeptide experiments, two LC-MS/MS experiments were conducted. In one experiment, the A-QLT was used to conduct ETD; in the other experiment, the MDC was used to perform AI-ETD. Analysis using the A-QLT produced 1164 PSMs, corresponding to 980 phosphopeptides, 647 of them unique (**Figure 6A**). In contrast, use of MDC AI-ETD

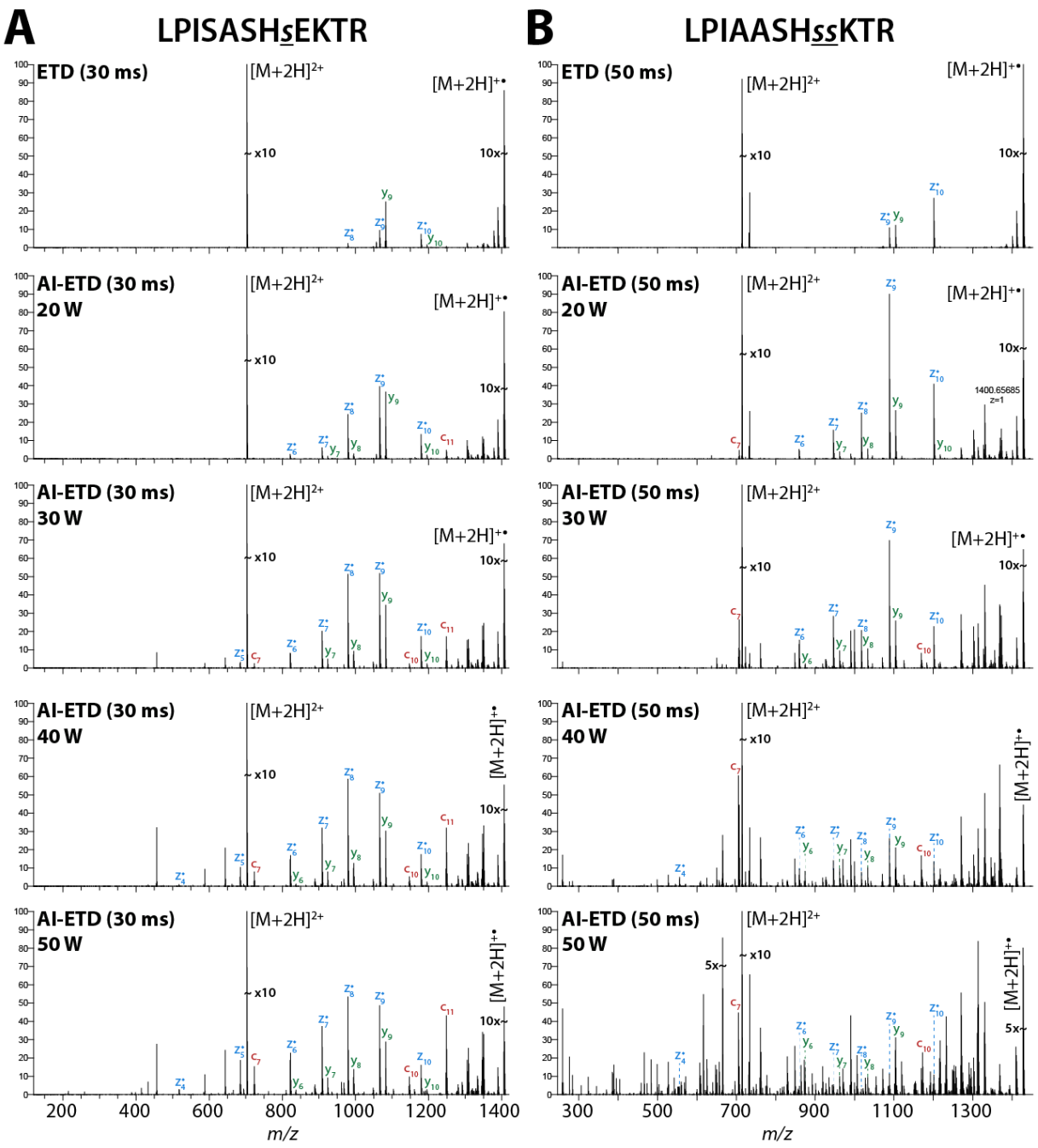


Figure 5. Comparison of ETD and AI-ETD for standard phosphopeptides.
A singly phosphorylated peptide, LPISASHsEKTR (A), and a double phosphorylated peptide, LPIAASHssKTR (B), were interrogated using ETD and AI-ETD at laser powers of 20, 30, 40, or 50 W. The increased IR adsorption of additional phosphoryl groups causes increased unwanted fragmentation (neutral loss, b-, y-, a-, and x-type ions) at relatively low laser power settings (40 W).

results in 3271 PSMs, corresponding to 2736 phosphopeptides and 1575 unique phosphopeptides (**Figure 6A**). As with unmodified peptides, MDC AI-ETD identified most phosphopeptides sequences by A-QLT ETD, while also identifying 1169 phosphopeptides not sequenced by A-QLT ETD (**Figure 6B**). These resulting data were treated similarly to the large scale experiment using unmodified peptides (*vide supra*). Data were again searched using OMSSA, and PSMs within a 1% FDR were binned by precursor *m/z*, and divided by the number of spectral features sampled, providing a measure of success for both ETD and AI-ETD, as a function of precursor *m/z* (**Figure 6C**). AI-ETD produces more phosphopeptide PSMs than ETD, particularly for high *m/z* precursors. Increasing precursor *m/z* has a detrimental effect on the ability of ETD to generate PSMs. While AI-ETD is, likewise, less effective for high *m/z* precursors compared to low *m/z* precursors, the difference in overall performance is far less dramatic.

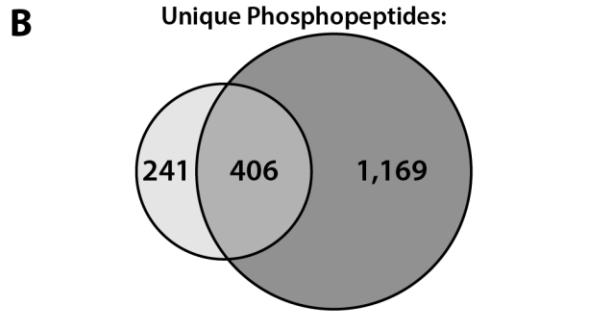
An example of the improvement MDC AI-ETD offers, over conventional ETD conducted in the A-QLT, is shown in **Figure 6D**. Here a triply charged phosphopeptide VVDYSQFQESDDADEDYGR was fragmented by A-QLT ETD or AI-ETD performed in the MDC. AI-ETD results in much higher ETD fragmentation efficiency and a much greater diversity of product ions as compared to A-QLT ETD. AI-ETD of this peptide produces a small amount of collisional activation as b- and y-type ions are produces upon IR irradiation, however these ions do not inhibit peptide identification through database searching. The improvements in identification rates demonstrate that AI-ETD performed in the MDC will improve large scale analysis of PTM containing peptides as compared to A-QLT ETD.

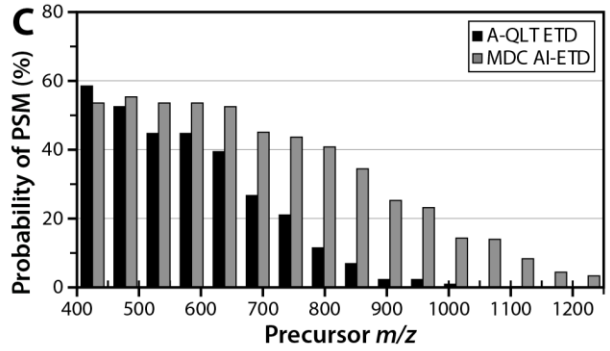
Discussion

This implementation of AI-ETD on a hybrid QLT-Orbitrap mass spectrometer represents an advance in ETD fragmentation that will aid in large scale identifications of both unmodified and PTM containing peptides. We have demonstrated that simple instrument modifications allowed the introduction of IR photons to the multi dissociation cell (MDC), enabling the use of IR photon activation concomitant with ETD (AI-ETD). Utilizing high mass accuracy MS^1 scans yielded charge state information for selected precursors, which enabled the construction of decision tree logic for the applied level of irradiation during

A

	A-QLT ETD	SC AI-ETD
Pep. spectral matches:	1,164	3,271
Phosphopeptides:	980	2,736
U. phosphopeptides:	647	1,575





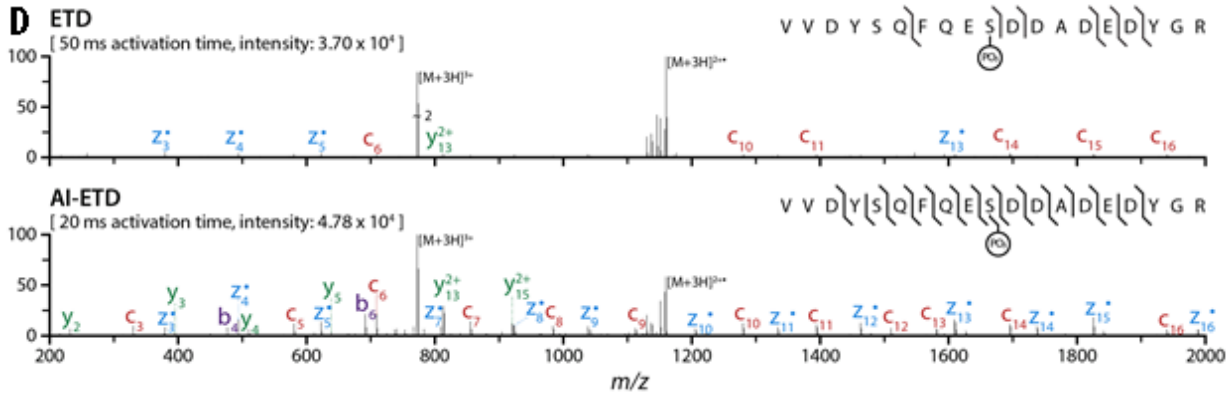


Figure 6. Comparison of A-QLT-based ETD and MDC-based AI-ETD for phosphopeptide analysis.

Use of the MDC AI-ETD results in a greater number of PSMs, phosphopeptide PSMs, and unique phosphopeptides identified (A), while retaining a high proportion of the unique phosphopeptides obtainable using A-QLT-based ETD (B). This is largely due to greatly enhanced probability of PSM, particularly for high *m/z* precursors (C) afforded by the IR activation of precursors during ETD. D) ETD and AI-ETD of the triply protonated phosphopeptide VVDYSQFQEsDDADEDYGR.

AI-ETD. Use of AI-ETD substantially enhanced ETD fragmentation efficiency for unmodified peptides and, importantly, phosphopeptides. When used in large scale analysis AI-ETD resulted in nearly a two-fold increase in the number of unique unmodified peptides identified (2,368 vs. 3,789), and almost three times as many unique phosphopeptides (647 vs. 1,575) over an LC-MS/MS experiment, as compared to unassisted ETD performed in the A-QLT.

The improved performance of AI-ETD will be valuable for further studies of PTM containing peptides and proteins that are difficult to sequence with collisional methods. LC-MS/MS analysis of larger peptides or proteins should benefit from the greater fragmentation efficiency of AI-ETD. Likewise, using AI-ETD to study glycopeptides could present an avenue to collecting spectra containing both peptide and glycan fragmentation in a single spectrum. In addition, the added specificity of high mass accuracy through Orbitrap analysis removes limits of low resolution analysis on ion trap instruments when analyzing bigger species, such as proteins. Lastly, the recent description of TMT reagents amenable to ETD fragmentation will allow researchers to quantify peptides using only ETD fragmentation.[32] In these types of analyses, AI-ETD would represent a significant advance in improving data quality for quantitative analysis of biological systems.

Experimental procedures

Sample preparation

All peptides and chemicals were purchased from Sigma-Aldrich (St. Louis, MO). Standard peptides were prepared by diluting stock solutions to ~5 pmol/uL in 50/50 ACN with 0.2% FA.

Wild-type yeast (*Saccharomyces cerevisiae*) was grown in YPD medium at 30 °C to an optical density (OD_{600}) of ~0.6. Cells were collected and centrifuged at 8000 rpm for 10 min at 4 °C. The resulting cell pellet was washed twice with sterile water and centrifuged at 5000 rpm for 5 min at 4 °C. Lysis buffer of approximately three times the cell pellet volume was added. The lysis buffer contained 8 M urea, 75 mM NaCl, 50 mM Tris (pH 8), 1 mM sodium orthovanadate, 100 mM sodium butyrate, complete mini EDTA-free protease inhibitor (Roche Diagnostics), and phosSTOP phosphatase inhibitor (Roche Diagnostics).

Drops of yeast lysate were flash frozen in liquid nitrogen to form lysis popcorn. Equal volumes of lysis popcorn and acid-washed glass beads (Sigma) were added to the grinding jar. Cells were ruptured by bead-beating on a Restek MM4000 Mixer Mill using 3 x 4 minute cycles at 30 Hz. Lysates were transferred to fresh tubes and centrifuged at 5000 rpm for 15 min at 4 °C. Cysteine residues were reduced and alkylated by incubating lysate with 5 mM DTT (final concentration) for 45 min at 37 °C followed by incubation in 15 mM IAA for 1 hour at room temperature in the dark. The alkylation reaction was capped by incubating the reaction with DTT for 15 minutes at room temperature. Proteins were digested for ~ 60 min at ~5 °C after the addition of 1 mM $CaCl_2$, 50 mM Tris (to decrease urea to 1 M), and adjusting to pH 8 at an enzyme-to-substrate ratio of 1:200 of trypsin (Promega, Madison, WI). The digest was quenched by the addition of TFA to a final concentration of 0.5% (pH 2), and desalted via solid phase extraction on a 50-mg tC_{18} SepPak cartridge (Waters, Milford, MA).

To prepare phosphopeptides, human embryonic stem cells were lysed in ice-cold 8M urea, 40 mM NaCl, 50 mM tris (pH 8), 2 mM $MgCl_2$, 50 mM NaF, 50 mM beta-glycero phosphate, 1 mM sodium orthovanadate, 10 mM sodium pyrophosphate, 1X mini EDTA-free protease inhibitor (Roche Diagnostics), and 1X phosSTOP phosphatase inhibitor (Roche Diagnostics). To solubilize protein, and ensure complete lysis, samples were sonicated three times, for 15 seconds, with 30 second pauses. Total protein was then quantified using a BCA protein assay kit (Thermo Scientific Pierce), reduced by adding DTT to a final concentration of 5 mM, and alkylated with 10 mM iodoacetamide. Digestion was carried out by adding trypsin (Wako Chemicals) at a 1:100 emzyme-to-protein ratio, and incubating at 37 degrees C for 2 hours. At this time, the lysate was diluted with 25 mM tris (pH 8) to a final urea concentration of 1.5 M, and further digested for 12 hours at 37 degrees C with trypsin (Promega), at a 1:100 enzyme-to-protein ratio. Peptides were then acidified with TFA, to quench the reaction, and de-salted using C-18 solid phase extraction (SPE) columns (Waters).

Phosphopeptides were enriched from 1 mg of protein, via immobilized metal affinity chromatography, (IMAC) using magnetic beads (Qiagen). Following equilibration with water, the beads were treated with 40 mM EDTA (pH 8.0) for 30 minutes with shaking, and washed 3x with water, again.

The beads were then incubated with 100 mM $FeCl_3$ for 30 minutes with shaking, and finally were washed 3 times with 80% acetonitrile/0.1% TFA. Samples were likewise re-suspended in 80% acetonitrile/0.15% TFA, and incubated with beads for 45 minutes with shaking. The resultant mixture was washed 3 times with 1 mL 80% acetonitrile/0.1% TFA, and eluted using 1:1 acetonitrile-to-0.7% NH_4OH in water. Eluted phosphopeptides were acidified immediately with 4% formic acid and lyophilized to ~5 µL.

Mass spectrometry and chromatography

All experiments were performed on an ETD-enabled, hybrid, dual-cell, quadrupole ion trap - Orbitrap mass spectrometer (Velos-Orbitrap, Thermo Fisher Scientific, San Jose, CA). A modified collision cell comprising four sections with independently controllable DC biases and supplemental RF voltage applied to the cell end lenses replaced the usual collision cell. This multi dissociation cell (MDC) is capable of performing charge sign independent trapping, enabling ETD, in addition to remaining a collision cell for beam-type CAD fragmentation. To support the new devices and the associated scan functions, we extensively modified the instrument control code.

The instrument was also modified to allow for concurrent excitation of the ETD precursor population by IR photon irradiation. IR photons were generated with an external Firestar T-100 Synrad 100-W CO_2 continuous wave laser (Mukilteo, WA), which was triggered on and off through the mass spectrometer firmware. To create a line-of-sight between our IR photon source and the MDC, a ZnSe window was installed, concentric with the trapping volume of the MDC (**Fig. 2**). A photon passage was excavated through the ETD reagent anion transfer multipole, which conducts ions to the MDC. This enabled the introduction of IR photons to the trapping volume of the MDC.

During AI-ETD MS/MS analysis, the A-QLT was used to trap and isolate precursor cations. Cations were then transferred to the MDC and sequestered in the front sections of the cell. The DC voltages of the back sections were set to positive values suitable for accumulation of ETD reagent anions. After anion injection ETD was initiated by setting all DC offsets of the MDC to 0 V and applying an axial RF voltage to the end lenses of the MDC. At this time IR photons were introduced at a level associated with

the charge and mass to charge of the selected precursor. The AI-ETD reaction was quenched by setting the center two sections to negative DC offsets, followed by *m/z* analysis using the Orbitrap.

LC separations of unmodified and phosphopeptides were carried out using a NanoAcquity UPLC system (Waters, Milford, MA) as previously described: using a 90 minute gradient of 2% to 10% B (0.2% formic acid in ACN) over 30 seconds, followed by a linear gradient increasing buffer B to 28% over 60 minutes, followed by a ramp up to 70% B over 2 minutes, and held for 5 minutes.[29] The gradient was dropped back to 98% A (0.2% formic acid in H2O) over a period of 2 minutes, and allowed to re-equilibrate for 20 minutes. During the LC-MS/MS analysis of unmodified and phosphopeptide complex mixtures, mass spectrometric methods consisted of an MS^1 scan, followed by consecutive ETD, and AI-ETD data-dependent MS/MS scans of the five most intense precursors. Precursors were dynamically excluded for 45 s using an isolation window of ± 1.5 Th. Unless otherwise specified, AGC target values were 1×10^6 for MS^1, 1×10^5 for MS/MS analysis, and an ETD reagent AGC target of 2×10^5.

Database Searching and data analysis

For phosphopeptide and unmodified peptide LC-MS/MS analyses, data reduction was performed with COMPASS,[33] a program that converts output files to searchable text files, as described previously. OMSSA (version 2.1.8, www.yeastgenome.org) was used to search spectra against the concatenated target-decoy SGD yeast database (downloaded 01-05-2010). Average mass tolerances of +/- 5 Th and +/- 0.01 Th were used for precursor and product *m/z*, respectively, with carbaminomethylation of cysteine set as a fixed modification, and oxidation of methionine set as a variable modification. All analyses were independently filtered to 1% false discovery rate, at the unique peptide level, using the concatenated forward-reverse database method, as previously described.[34-36] *C*- and *z*- type fragment ions were searched for both ETD and AI-ETD spectra. For LC-MS/MS analysis of the unmodified, limited tryptic digest, up to eight missed cleavages were considered; for LC-MS/MS analysis of the complex mixture of phosphopeptides, up to three missed cleavages were considered. Phosphopeptide site localization, and assignment of peptides (both unmodified and phosphorylated) to corresponding proteins, was carried out as previously described.[12]

Acknowledgments

The authors gratefully acknowledge support from Thermo Fisher Scientific, NSF CAREER grant 0747990, and NIH grant R01 GM080148. C.M.R was funded by an NSF Graduate Research Fellowship and NIH Traineeship (T32GM008505).

References

1. Zubarev RA, Kelleher NL, & McLafferty FW. Electron capture dissociation of multiply charged protein cations. A nonergodic process. *J. Am. Chem. Soc.*, 1998, 120(13):3265-3266.

2. Syka JEP, Coon JJ, Schroeder MJ, Shabanowitz J, & Hunt DF. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.*, 2004, 101(26):9528-9533.

3. Coon JJ, Syka JEP, Schwartz JC, Shabanowitz J, & Hunt DF. Anion dependence in the partitioning between proton and electron transfer in ion/ion reactions. *Int. J. Mass Spectrom.*, 2004, 236(1-3):33-42.

4. Chalkley RJ, Thalhammer A, Schoepfer R, & Burlingame AL. Identification of protein O-GlcNAcylation sites using electron transfer dissociation mass spectrometry on native peptides. *Proc. Natl. Acad. Sci. U. S. A.*, 2009, 106(22):8894-8899.

5. Khidekel N, Ficarro SB, Clark PM, Bryan MC, Swaney DL, Rexach JE, Sun YE, Coon JJ, Peters EC, & Hsieh-Wilson LC. Probing the dynamics of O-GlcNAc glycosylation in the brain using quantitative proteomics. *Nat. Chem. Biol.*, 2007, 3(6):339-348.

6. Hogan JM, Pitteri SJ, Chrisman PA, & McLuckey SA. Complementary structural information from a tryptic N-linked glycopeptide via electron transfer ion/ion reactions and collision-induced dissociation. *J. Proteome Res.*, 2005, 4(2):628-632.

7. Darula Z, Chalkley RJ, Baker P, Burlingame AL, & Medzihradszky KF. Mass spectrometric analysis, automated identification and complete annotation of O-linked glycopeptides. *Eur. J. Mass Spectrom.*, 2010, 16(3):421-428.

8. Scott NE, Parker BL, Connolly AM, Paulech J, Edwards AVG, Crossett B, Falconer L, Kolarich D, Djordjevic SP, Hojrup P, Packer NH, Larsen MR, & Cordwell SJ. Simultaneous Glycan-Peptide Characterization Using Hydrophilic Interaction Chromatography and Parallel Fragmentation by CID, Higher Energy Collisional Dissociation, and Electron Transfer Dissociation MS Applied to the N-Linked Glycoproteome of Campylobacter jejuni. *Mol. Cell. Proteomics*, 2011, 10(2).

9. Zee BM & Garcia BA. Electron Transfer Dissociation Facilitates Sequencing of Adenosine Diphosphate-Ribosylated Peptides. *Analytical Chemistry*, 2010, 82(1):28-31.

10. Messner S, Altmeyer M, Zhao HT, Pozivil A, Roschitzki B, Gehrig P, Rutishauser D, Huang DZ, Caflisch A, & Hottiger MO. PARP1 ADP-ribosylates lysine residues of the core histone tails. *Nucleic Acids Research*, 2010, 38(19):6350-6362.

11. Brumbaugh J, Hou Z, Russell JD, Howden SE, Yu P, Ledvina AR, Coon JJ, & Thomson JA. Phosphorylation regulates human OCT4. *Proceedings of the National Academy of Sciences*, 2012, 109(19):7162-7168.

12. Phanstiel DH, Brumbaugh J, Wenger CD, Tian SL, Probasco MD, Bailey DJ, Swaney DL, Tervo MA, Bolin JM, Ruotti V, Stewart R, Thomson JA, & Coon JJ. Proteomic and phosphoproteomic comparison of human ES and iPS cells. *Nat. Methods*, 2011, 8(10):821-U884.

13. Rose CM, Venkateshwaran M, Volkening JD, Grimsrud PA, Maeda J, Bailey DJ, Park K, Howes-Podoll M, den Os D, Yeun LH, Westphall MS, Sussman MR, Ané J-M, & Coon JJ. Rapid

Phosphoproteomic and Transcriptomic Changes in the Rhizobia-legume Symbiosis. *Molecular & Cellular Proteomics*, 2012, 11(9):724-744.

14. Swaney DL, McAlister GC, Wirtala M, Schwartz JC, Syka JEP, & Coon JJ. Supplemental activation method for high-efficiency electron-transfer dissociation of doubly protonated peptide precursors. *Anal. Chem.*, 2007, 79(2):477-485.

15. Ledvina AR, McAlister GC, Gardner MW, Smith SI, Madsen JA, Schwartz JC, Stafford GC, Syka JEP, Brodbelt JS, & Coon JJ. Infrared Photoactivation Reduces Peptide Folding and Hydrogen-Atom Migration following ETD Tandem Mass Spectrometry. *Angew. Chem.-Int. Edit.*, 2009, 48(45):8526-8528.

16. Ben Hamidane H, Chiappe D, Hartmer R, Vorobyev A, Moniatte M, & Tsybin YO. Electron Capture and Transfer Dissociation: Peptide Structure Analysis at Different Ion Internal Energy Levels. *J. Am. Soc. Mass Spectrom.*, 2009, 20(4):567-575.

17. Good DM, Wirtala M, McAlister GC, & Coon JJ. Performance characteristics of electron transfer dissociation mass spectrometry. *Mol. Cell. Proteomics*, 2007, 6(11):1942-1951.

18. Han HL, Xia Y, & McLuckey SA. Beam-type collisional activation of polypeptide cations that survive ion/ion electron transfer. *Rapid Commun. Mass Spectrom.*, 2007, 21(10):1567-1573.

19. Horn DM, Ge Y, & McLafferty FW. Activated ion electron capture dissociation for mass spectral sequencing of larger (42 kDa) proteins. *Anal. Chem.*, 2000, 72(20):4778-4784.

20. Cooper HJ, Hakansson K, & Marshall AG. The role of electron capture dissociation in biomolecular analysis. *Mass Spectrom. Rev.*, 2005, 24(2):201-222.

21. Oh H, Breuker K, Sze SK, Ge Y, Carpenter BK, & McLafferty FW. Secondary and tertiary structures of gaseous protein ions characterized by electron capture dissociation mass spectrometry and photofragment spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.*, 2002, 99(25):15863-15868.

22. Oh HB & McLafferty FW. A variety of activation methods employed in "activated-ion" electron capture dissociation mass spectrometry: A test against bovine ubiquitin 7+ions. *Bull. Korean Chem. Soc.*, 2006, 27(3):389-394.

23. Senko MW, Speir JP, & McLafferty FW. COLLISIONAL ACTIVATION OF LARGE MULTIPLY-CHARGED IONS USING FOURIER-TRANSFORM MASS-SPECTROMETRY. *Anal. Chem.*, 1994, 66(18):2801-2808.

24. Breuker K, Oh HB, Horn DM, Cerda BA, & McLafferty FW. Detailed unfolding and folding of gaseous ubiquitin ions characterized by electron capture dissociation. *J. Am. Chem. Soc.*, 2002, 124(22):6407-6420.

25. Horn DM, Breuker K, Frank AJ, & McLafferty FW. Kinetic intermediates in the folding of gaseous protein ions characterized by electron capture dissociation mass spectrometry. *J. Am. Chem. Soc.*, 2001, 123(40):9792-9799.

26. Pitteri SJ, Chrisman PA, & McLuckey SA. Electron-transfer ion/ion reactions of doubly protonated peptides: Effect of elevated bath gas temperature. *Anal. Chem.*, 2005, 77(17):5662-5669.

27. O'Connor PB, Lin C, Cournoyer JJ, Pittman JL, Belyayev M, & Budnik BA. Long-lived electron capture dissociation product ions experience radical migration via hydrogen abstraction. *J. Am. Soc. Mass Spectrom.*, 2006, 17(4):576-585.

28. Ledvina AR, Beauchene NA, McAlister GC, Syka JEP, Schwartz JC, Griep-Raming J, Westphall MS, & Coon JJ. Activated-Ion Electron Transfer Dissociation Improves the Ability of Electron Transfer Dissociation to Identify Peptides in a Complex Mixture. *Anal. Chem.*, 2010, 82(24):10068-10074.

29. McAlister GC, Phanstiel D, Wenger CD, Lee MV, & Coon JJ. Analysis of Tandem Mass Spectra by FTMS for Improved Large-Scale Proteomics with Superior Protein Quantification. *Anal. Chem.*, 2010, 82(1):316-322.

30. Flora JW & Muddiman DC. Selective, sensitive, and rapid phosphopeptide identification in enzymatic digests using ESI-FTICR-MS with infrared multiphoton dissociation. *Anal. Chem.*, 2001, 73(14):3305-3311.

31. Crowe MC & Brodbelt JS. Infrared multiphoton dissociation (IRMPD) and collisionally activated dissociation of peptides in a quadrupole ion trap with selective IRMPD of phosphopeptides. *J. Am. Soc. Mass Spectrom.*, 2004, 15(11):1581-1592.

32. Werner T, Becher I, Sweetman G, Doce C, Savitski MM, & Bantscheff M. High-Resolution Enabled TMT 8-plexing. *Analytical Chemistry*, 2012, 84(16):7188-7194.

33. Wenger CD, Phanstiel DH, Lee MV, Bailey DJ, & Coon JJ. COMPASS: A suite of pre- and post-search proteomics software tools for OMSSA. *Proteomics*, 2011, 11(6):1064-1074.

34. Moore RE, Young MK, & Lee TD. Qscore: An algorithm for evaluating SEQUEST database search results. *J. Am. Soc. Mass Spectrom.*, 2002, 13(4):378-386.

35. Elias JE & Gygi SP. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods*, 2007, 4(3):207-214.

36. Wenger CD, Lee MV, Hebert AS, McAlister GC, Phanstiel DH, Westphall MS, & Coon JJ. Gas-phase purification enables accurate, multiplexed proteome quantification with isobaric tagging. *Nat. Methods*, 2011, 8(11):933-935.

Chapter 6

Normalization of ETD reaction times for optimized product ion abundance and peptide identifications

CMR designed research, performed experiments, analyzed data, and wrote the paper.

This chapter has been submitted for review:

Rose CM, Mullen C, Rush MJP, Westphall MS, Syka JEP, Coon JJ. "Normalization of ETD Reaction Times for Optimized Product Ion Abundance and Peptide Identifications." *Analytical Chemistry*, **2014**, *submitted for review*.

Abstract

Electron transfer dissociation (ETD) has traditionally been conducted using static reaction times that are determined without calibration or normalization. Here we introduce a methodology to standardize ETD reaction times across experiments and instruments. We first introduce a calibration routine that determines the correct number of reagent anions necessary to reach a desired ETD reaction rate. Using the calibrated number of reagent anions, we determine the optimal reaction time based on the theoretical time constant for each charge state. Analysis of ETD product ion intensity in both the ion trap and Orbitrap determined that the optimal reaction time was ~2x the time constant for each charge state. Utilizing a scalable normalized ETD reaction time, similar to a normalized collision energy for CID fragmentation methods, we demonstrate peptide identifications in a complex sample are optimized for both the ion trap and Orbitrap analysis at a reaction time corresponding to ~1-2x the time constant. Lastly, we evaluated this calibration routine implemented on a Q-OT-qIT mass spectrometer and found that using calibrated ETD times increased unique peptide identifications more than 50% as compared to a control utilizing static reaction times.

Introduction

Enabled by new generations of faster mass spectrometers, analysis of entire proteomes within a single run is becoming possible.[1] Improvements in ion injection and ion trap mass analysis have lowered the time required to inject a precursor, fragment *via* CAD, and analyze the resulting fragments to as low as 50 ms.[1,2] Coupling fast MS/MS scans with parallelization, new mass spectrometers have achieved ~20 MS/MS events / sec, enabling the collection of more than 80,000 MS/MS scans within a single one hour experiment.[1] Despite improvements in many aspects of MS data collection, the implementation and utilization of electron transfer dissociation (ETD)[3,4] is not standardized.

ETD has become an important alternative to collisional-based peptide dissociation methods, enabling the identification of unique portions of the proteome as well as peptides modified with labile PTMs.[5-8] Despite the widespread utilization of ETD, operational parameters are neither standardized across instruments nor experiments. Collisional activation methods, CAD and beam type CAD (HCD),[9,10] are implemented with normalized collision energies ensuring default settings produce spectra with high probability of being identified.[11,12] No such calibration exists for ETD and many important parameters are often empirically determined.[13-20] Compton *et al.* reported that ETD reaction times are often much higher than necessary and reducing reaction times leads to improvements in scan speed in the analysis of BSA digested with LysC.[21] Aside from this study, work aimed at optimizing ETD reactions to maximize peptide identifications are absent.

Successful ETD reactions rely on many operational parameters including the number of precursor and reagent ions, charge of the precursor, characteristics of the ion-ion reaction vessel, and reaction time.[21-25] To date, most ETD experiments have utilized a range of static reaction times typically varying from 50 to 150 ms.[8,21,26-28] As we explain below, ion-ion reaction models demonstrate the rate of ETD reactions scale with the square of the precursor charge.[22,23] A scaled ETD reaction time algorithm is available on some mass spectrometers,[29] but this implementation represents only a rough estimate of the relationship between precursor charge and reaction time.

Extensive work has been invested in understanding gas phase ion-ion reactions,[22,23] briefly, if the reagent ion population is in excess of the precursor ions we can model the reaction as,

$$N_p(t) = N_p(0)e^{-k[R]t} \qquad \textbf{(Equation 1)}$$

where $N_p(0)$ represents the initial precursor population, $N_p(t)$ is the amount of precursor remaining at time *t*, [R] is the average number density of the portion of the reagent ion cloud that overlaps with the precursor ion cloud, and k is the ETD ion-ion reaction rate constant. This rate constant can be expressed as,

$$k = c(|v|)\, Z_p^2\, Z_r^2\, ((m_p+m_r)/\, m_p\, m_r)^2 \qquad \textbf{(Equation 2)}$$

where Z_p is precursor charge, m_p is precursor mass, Z_r is reagent charge, m_r is reagent mass, and the quantity $c(|v|)$ is a function of the distribution of the magnitude of differential velocities, $|v|$, of precursor and reagent ions in the overlapping ion clouds.

Using these two equations as guides we can determine which parameters to optimize. If the number of precursor ions and reaction vessel ($N_p(0)$ and [R] respectively) are held constant, **Equation 1** dictates that the two parameters guiding the ion-ion reaction are k and *t*. Thus, standardizing the ETD rate constant, k, would enable the optimization of reaction times, *t*, to create the highest product ion yield. From **Equation 2**, we learn that k will scale based on the square of the precursor charge, Z_p^2, as well as the density or number of precursor and reagent ions, $c(|v|)$. Practically, this means ETD reactions of higher charge state precursors occur more quickly, lowering the required reaction time. In addition to decreasing scan duration, lower ETD reaction times limit the occurrence of secondary ETD reactions of product ions that decrease overall product ion signal-to-noise. Thus, for every charge state there is an ETD reaction time that will result in an optimal creation of product ions.

Here we describe a method to calibrate the ETD reaction rate constant (k) enabling the determination of optimal charge state dependent ETD reaction times using both product ion yield and unique peptide identifications as figures of merit. Frist, we developed a calibration routine to determine the amount of reagent anions necessary to achieve a chosen ETD reaction rate. We then used this rate constant to calculate charge state dependent time constants (τ), which became a basis for finding optimal ETD reaction times. After analyzing a wide variety of peptides, we determined that a reaction time ~2τ provides optimal product

ion yield, while reaction times ~1-2τ produce the most unique peptide identifications in the analysis of a complex sample. Lastly, we evaluate a standard implementation of this algorithm implemented on a quadrupole-Orbitrap-quadrupole ion trap (Q-OT-qIT)^2 mass spectrometer and find the number of unique peptides increases ~50% as compared to a control experiment utilizing a static reaction time.

Results

Demonstration of Optimal Reaction Times

To date ETD has been implemented without calibrated or normalized reaction conditions, commonly leading to reaction times substantially longer than necessary.[26] Unlike collisional activation, where fragmentation varies with *m/z* and charge of the precursor as well as collisional gas pressure, ETD reaction times vary with the number of precursor ions, number of reagent ions, charge of the precursor, and characteristics of the ion-ion reaction vessel.[22,23,25] Optimal ETD reaction times leave a fraction of precursor remaining and often result in spectra that appear underreacted.[21] To demonstrate this, the triply charged precursor of angiotensin I (DRVYIHPFHL) was reacted with fluoranthene anions for times ranging from 5 to 160 ms and the signal to noise of *c*- or *z*•-type product ions was averaged over 50 MS/MS events for each reaction time (**Figure 1A**).

Doubtless the intensity of the unreacted precursor, following ETD, does not hold the same diagnostic value pertaining to dissociation completeness as it does in collisional activation. For example, an abundance of unreacted precursor remains for the 40 ms reaction that produced the highest sequence ion signal-to-noise (~9.5 S/N) (**Figure 1B**). Conventional logic suggests that longer reactions will lead to increased product ion signal-to-noise, as the precursor is further dissociated to product ions. However, longer reactions lead to secondary ETD events that ultimately degrade product ion signal. Consider the 100 ms reaction, the unreacted precursor is virtually nonexistent, but the sequence ion signal-to-noise is significantly lower (~6 S/N) than a 40 ms reaction (**Figure 1B**). Noting that optimized reaction times can drastically improve quality of MS/MS spectra, we surmised that an algorithm to calibrate ETD reaction times would increase peptide identifications for large scale analyses.

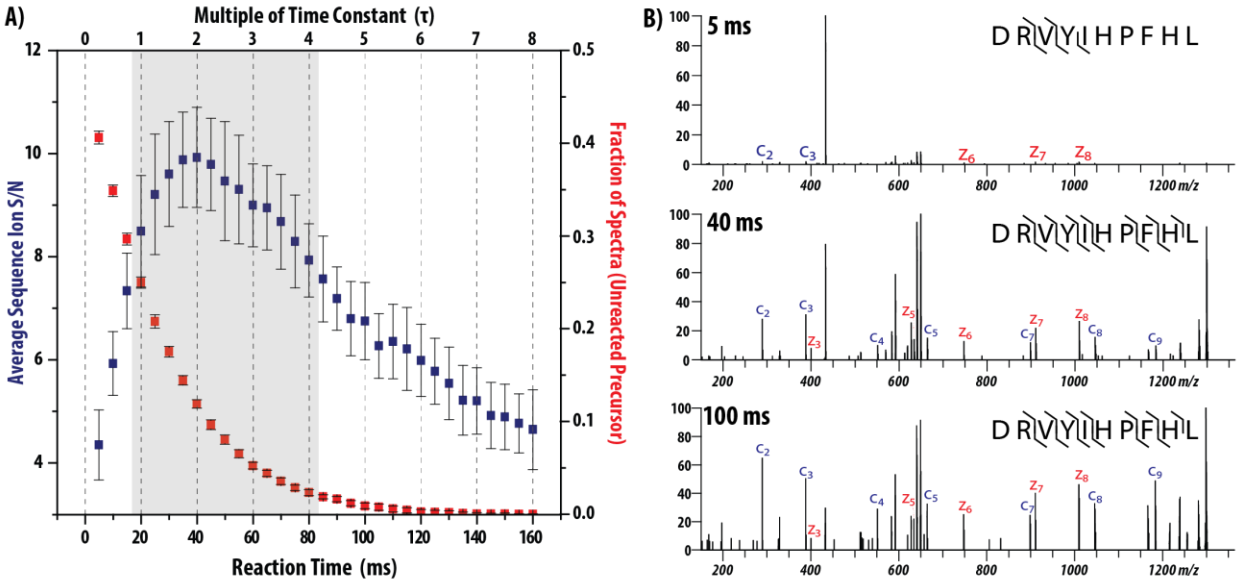


Figure 1. Reaction time dependence of product ion signal to noise for Angiotensin I.
A) The triply charged precursor of Angiotensin I (DRVYIHPFHL) was reacted with fluoranthene anions for increasing amounts of time and the average signal to noise for all product ions, as well as the unreacted precursor, was calculated. B) ETD MS2 spectra for three selected reaction times (5, 40, or 100 ms).

Calibration of ETD Reagent Anion Population

Theoretical models of ion-ion reactions have been developed[22,23] and we posited that this information could be leveraged to optimize the duration of ETD reactions. Ultimately, we aim to relate ETD reaction time to the charge state dependent time constant, which itself is related to ion-ion reaction rate constant as shown,

$$k_n = k_3 (Z_n / Z_3)^2 \qquad \textbf{(Equation 3)}$$

$$\tau_n = 1 / k_n \qquad \textbf{(Equation 4)}$$

where k_n is the charge state dependent rate constant, Z_n is the charge state of interest, Z_3 is the base charge state (3), k_3 is the base reaction rate, and τ_n is the charge state dependent time constant.

We developed a routine to calibrate the ETD reaction rate by determining the number of reagent ions necessary to achieve a desired k (**Figure 2**). To accomplish this, the triply charged precursor of angiotensin was reacted with an increasing number of reagent anions for reaction times ranging from 5 ms to 115 ms with the fraction of unreacted precursor measured for each ETD reaction time (N=4) (**Figure 2A**). This process was repeated from multiple reagent ion AGC targets (20,000 to 450,000) and the resulting data was fit with a linear function, whose slope represents the ETD rate constant. Plotting these results with all other reaction rates creates a curve with a maximum around k = 50 at a reagent anion AGC target of ~400,000 (**Figure 2B**).

Standardizing k_3 allowed us to use **Equations 3 and 4**, with $Z_3 = 3$ and $k_3 = 50$, to calculate the k_n and τ_n for charge states 2-6: $\tau_2 = 45$, $\tau_3 = 20$, $\tau_4 = 11$, $\tau_5 = 7$, and $\tau_6 = 5$ ms. We concluded above that a 40 ms reaction produced optimal product ion signal-to-noise for the triply charged precursor of angiotensin. By relating this optimal reaction time to the time constant when z = 3 ($\tau_3 = 20$ ms) we find that the optimal reaction time for angiotensin is 2τ (**Figure 1A**). We hypothesized, then, that maximum product ion yield for a precursor of any charge state would occur at a time equal to 2τ and calibrating the ETD reaction rate would enable normalized ETD reaction times across platforms and experiments.

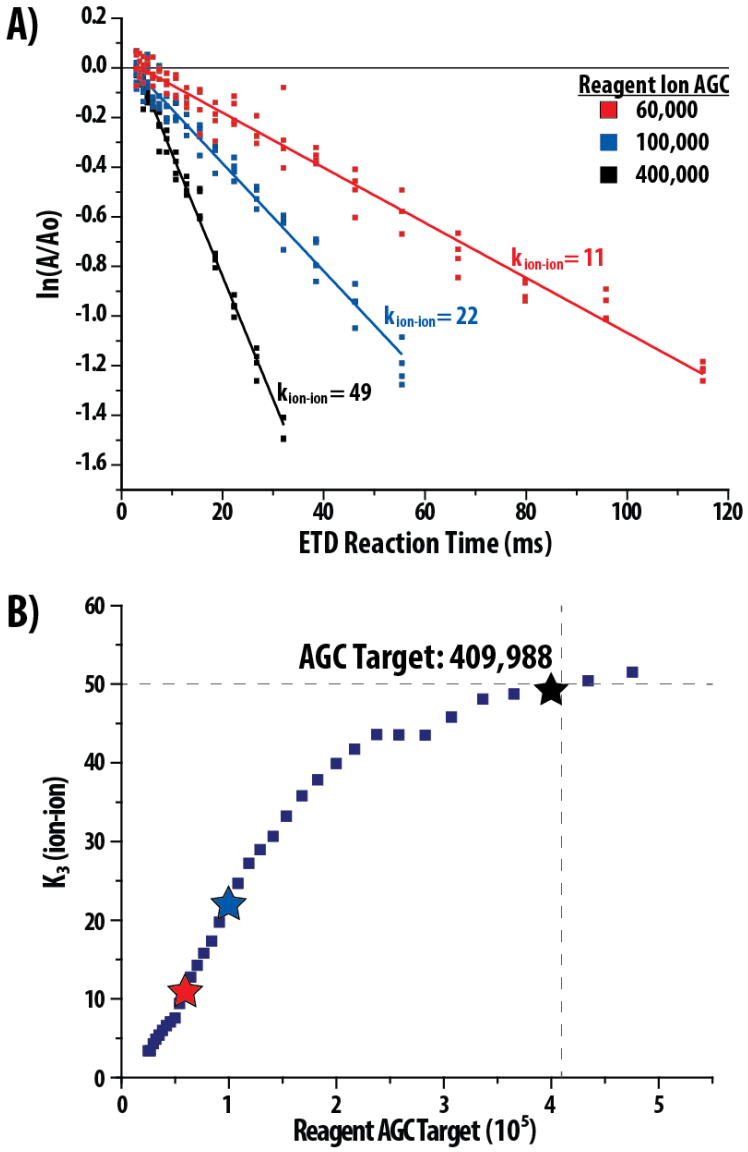


Figure 2. Calibration of ETD reagent ion target to meet the desired ETD reaction rate constant.

A) The triply charged precursor of Angiotensin I is reacted for various times at each reagent ion target to determine the rate constant. B) Plotting the rate constant as a function of reagent ion AGC enables the determination the number of reagent anion necessary to achieve a desired rate constant (50).

Large-Scale Analysis of Optimal Reaction Times

To explore the relationship between the ETD time constant (τ) and optimal reaction time we constructed a large scale experiment to test a range of reaction times and determine the optimal product ion yield (**Figure 3**). For each nLC-MS/MS experiment, precursors of a single charge state were reacted for nine different times ranging from 1-4τ with product ions analyzed in either the ion trap (**Figure 3**) or Orbitrap (**Figure 4**). This range of times (1-4τ) surrounded the maximal product ion yield in the survey experiments discussed above (**Figure 1A, shaded area**). Reaction times were sampled in five random orders to ensure peptide elution did not affect the analysis. Each order included a reaction lasting 2τ as the first, middle, and last scan event to control for peptide elution. For experiments utilizing ion-trap analysis the initial reaction at 2τ was performed in the Orbitrap to increase the probability of identification. **Figure 3B** displays six of the eleven ETD reactions of the peptide YRPNCPIILVTR. Even though they were analyzed several seconds apart, spectrum 2 is very similar to spectrum 12 validating analysis of all eleven reactions together. Longer reaction times clearly decrease unreacted precursor signal (**Figure 3**), but to determine the optimal product ion yield, a more in depth analysis is necessary.

For each MS/MS event the sequence ion intensity was summed to determine the multiple of τ that produces optimal product ion yield (**Figure 5 and 6**). To ensure proper kinetics, only MS/MS events in which the desired number of ions was reached (i.e., injection time lower than maximum) were scored. More than 100 precursors met the necessary criteria for each charge state, ensuring that a variety of peptides were sampled in all nine τ multiples (**Figure 5A and 6A**). For comparison, the summed product ion intensity for each MS/MS event was normalized to the largest summed product ion intensity for scan events 2-11 from each cycle, resulting in a score ranging from 0.0 to 1.0 for each τ multiple.

We performed this analysis for precursors with charge states ranging from 2 to 6, and encouragingly all charge states demonstrated optimal product ion intensity near 2τ for both IT and OT analysis (**Figure 5B-E and 6B-E**). These results demonstrate that product ion yield has a fixed

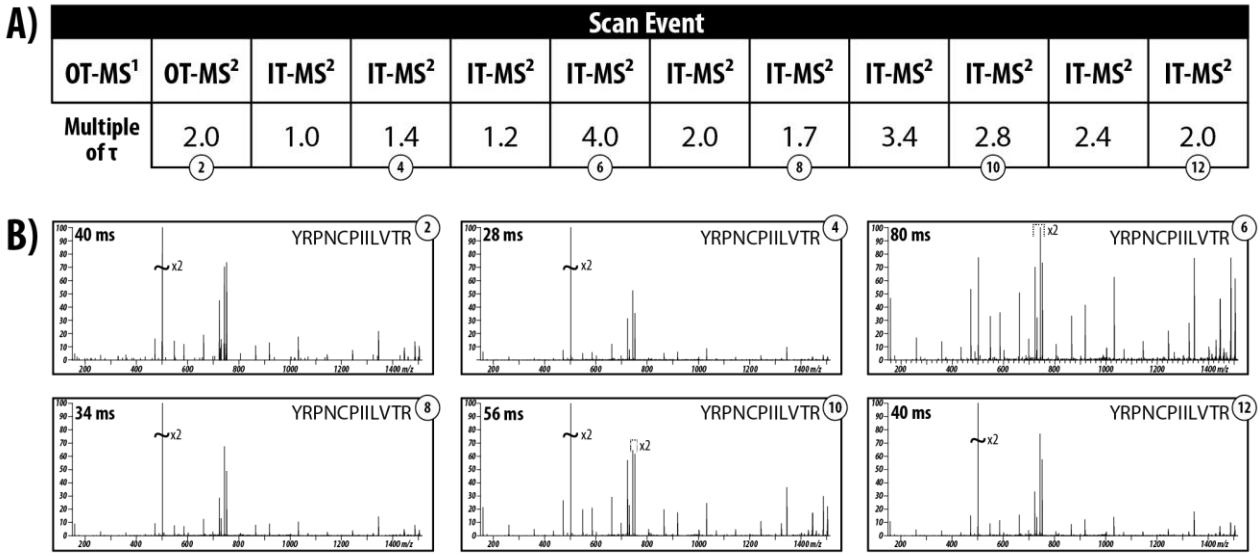


Figure 3. Experimental design to determine optimal multiple of ETD time constant.
A) Each selected precursor was reacted for nine different reaction times calculated by using various multiples of the time constant. B) Selected spectra for the precursor YRPNCIILVTR reacted for five different times based on the multiples of the time constant from (A).

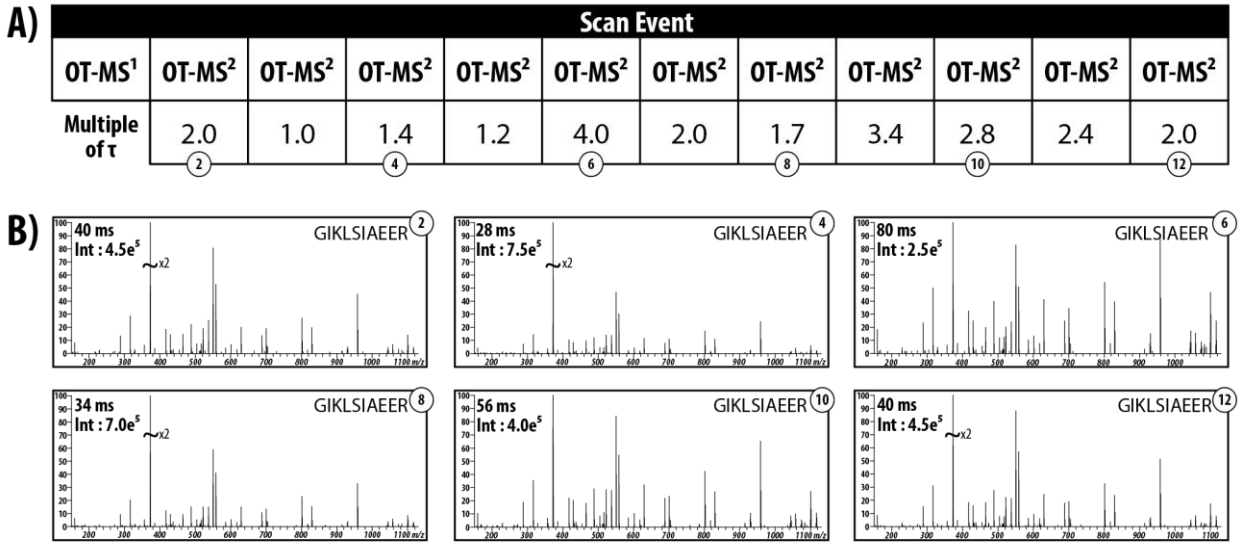


Figure 4. Experimental design to determine optimal multiple of ETD time constant.
A) Each selected precursor was reacted for nine different reaction times calculated by using various multiples of the time constant. B) Selected spectra for the precursor GIKLSIAEER reacted for five different times based on the multiples of the time constant from (A).

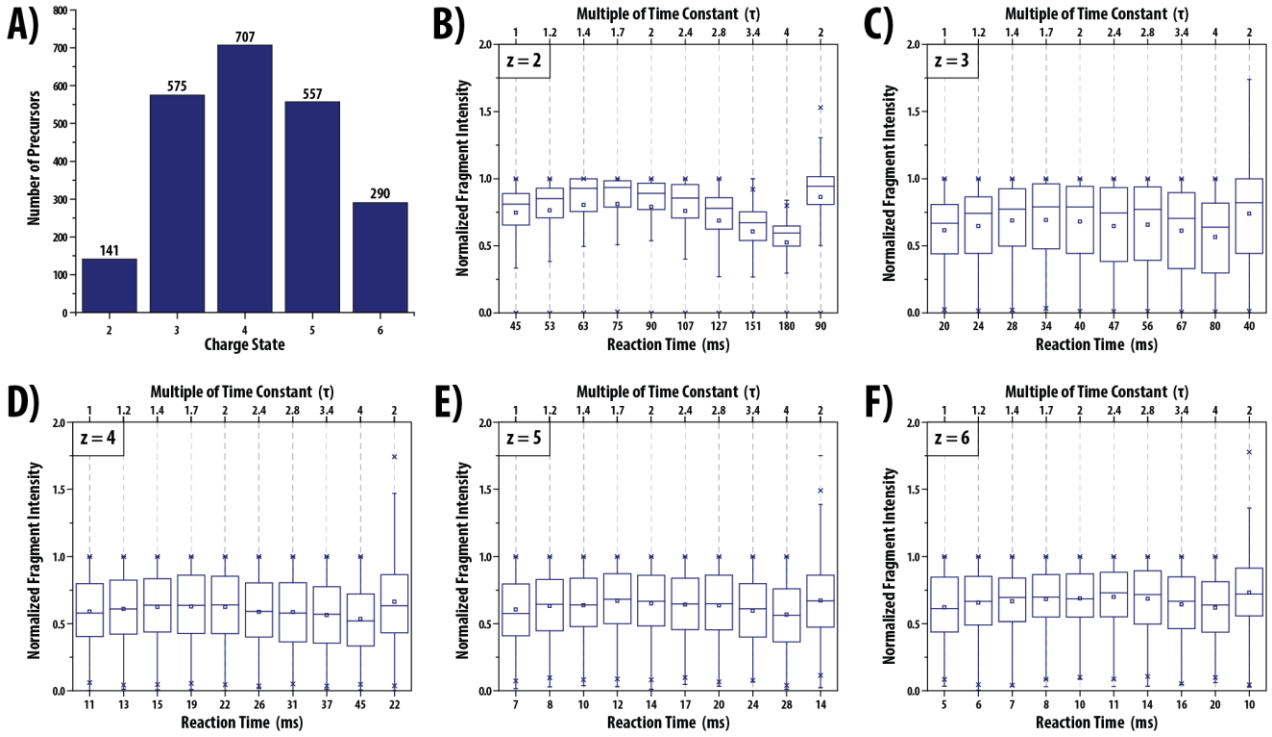


Figure 5. Global analysis of product ion intensity for each multiple of the time constant.
A) The number of precursors sampled for each charge state that included eleven measurements that reached the desired number of ions. B-F) The summed product ion intensity for each multiple of the time constant (1-4) was determined and normalized to the maximum intensity scan for each precursor.

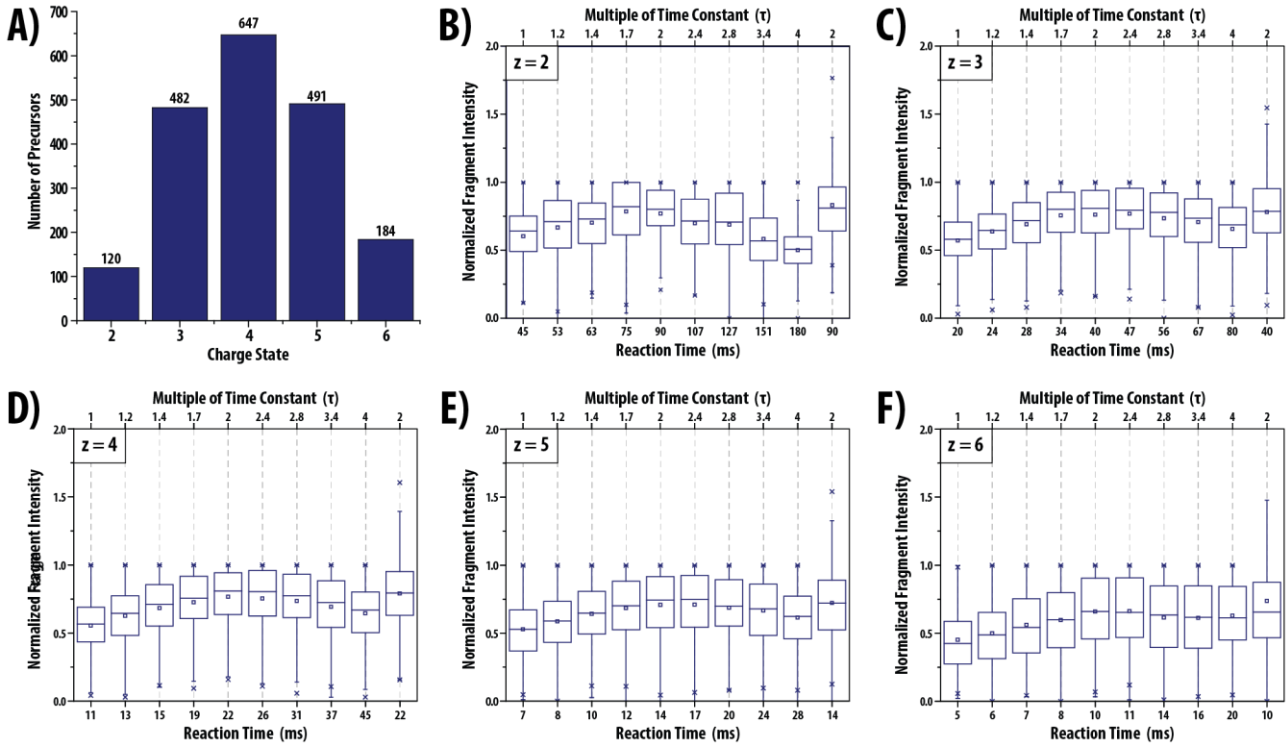


Figure 6. Global analysis of product ion intensity for each multiple of the time constant.
A) The number of precursors sampled for each charge state that included eleven measurements that reached the desired number of ions. B-F) The summed product ion intensity for each multiple of the time constant (1-4) was determined and normalized to the maximum intensity scan for each precursor. An additional scan with the time constant multiple of two was added to ensure there were no retention time effects on the calculations.

relationship to τ – suggesting the potential for a normalized ETD reaction time similar to the normalized collision energy of ion trap CAD and HCD. Interestingly, Orbitrap data demonstrates a sharp product ion intensity maximum at 2τ **(Figure 6B-E)**, whereas ion trap data displays a broader maximum stretching from 1.4τ to 2τ (**Figure 5B-E)**. This is likely explained by the greater sensitivity of ion trap and suggests that optimal reaction times are critical when analyzing product ions in the Orbitrap. These results prompted us to develop a normalized ETD reaction time algorithm to interrogate a complex sample where the benefits of optimal product ion yield are balanced with shorter scan duration.

Implementation and Evaluation of Normalized ETD Reaction Times

Collision energies for ion trap CAD and beam-type CAD (HCD) are normalized allowing users to decrease or increase the level of peptide fragmentation.[11] We surmised that a similar normalized reaction time (NRT) could be calculated for ETD, by scaling reaction times by various multiples of the time constant (**Figure 7**). Like ion trap CAD and HCD implementations, we created an algorithm that allows users to set the NRT between 0 and 100, correlating to multiples of the time constant ranging from 0 to 20 (**Figure 7A**). Using this algorithm we interrogated a complex mixture with multiple NRTs to find a balance between optimal product ion intensity and shorter overall scan duration.

A complex mixture of LysC yeast peptides was analyzed in eleven nLC-MS/MS experiments with each run conducting ETD at a normalized time ranging from 0 to 100. An additional nLC-MS/MS experiment was conducted reacting all precursors for 100 ms for comparison (**Figure 7A**). Normalized reaction times of 20-30 produce the highest number of both peptide spectral matches (PSMs) and unique peptides (**Figure 7A-B**). Specifically, NRTs of 20 and 30 produce 19% (4,344) and 16% (4,243) more unique identifications than the 100 ms control (3,655). Interestingly, a reaction lasting 1.4 to 2τ produces the highest product ion intensity (**Figure 5**); however, a NRT as low as 20 (equal to 1τ) produces the most unique identifications in a large-scale analysis of a complex mixture (**Figure 7**). This is most likely a result of additional MS/MS events performed when using an NRT of 20 as compared to the 100 ms control (19,608 vs. 15,265).

A)

Normalized Reaction Time(%)	Multiple of τ	MS^2 Scans	PSMs	Unique Peptides
0	0	20,661	30	30
10	0.3	20,432	4,926	4,021
20	1.0	19,608	5,403	4,344
30	2.0	18,532	5,301	4,243
40	3.4	17,181	5,123	4,150
50	5.2	15,723	4,805	3,888
60	7.4	14,270	4,166	3,430
70	10	12,795	3,644	3,009
80	13	11,434	3,121	2,561
90	16	10,146	2,527	2,088
100	20	9,086	2,045	1,747

B)

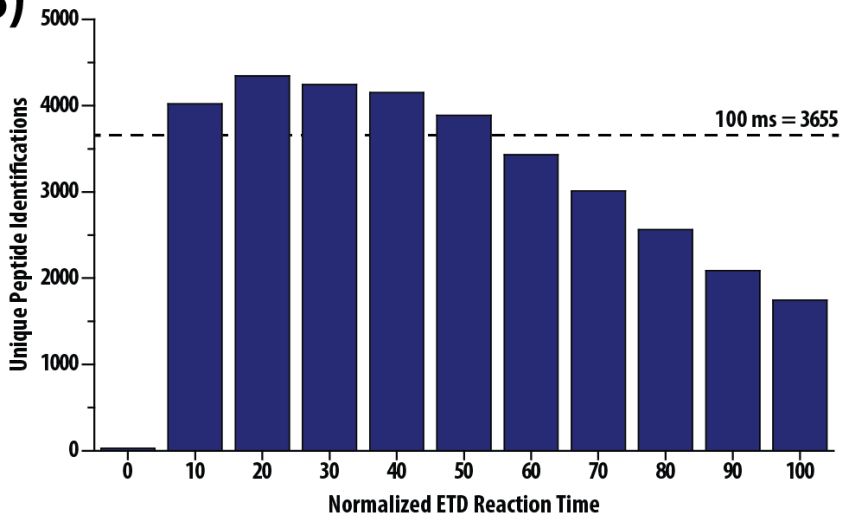


Figure 7. Implementation of a normalized ETD reaction time.

A) Number of unique peptides identified in one nLC-MS/MS experiment utilizing a normalized ETD reaction time. B) Experimental statistics for each nLC-MS/MS experiment evaluating various normalized ETD reaction times.

Identical experiments performed in the Orbitrap also produced a maximum number of PSMs and unique peptides using NRT between 20 and 30 (**Figure 8**); however, the number of unique peptides exhibits a much more distinct apex, when compared to the ion trap. Here, NRTs of 20 and 30 produce 42% and 39% more unique identifications than the 100 ms control (1227). Interestingly, the apex around 2τ matches results from the large-scale analysis of optimal reaction times (**Figure 6**) and suggests optimal ETD reaction times are critical when analyzing ETD product ions in the Orbitrap.

LysC is typically used to prepare samples for ETD because it produces peptides with higher charge states; however, in the above experiments we almost exclusively identified precursors of charge states 3 or 4 (**Figure 9**). We attribute this phenomenon to the presence of DMSO in the LC mobile phases.[1] Recently, Hahne *et al.* demonstrated that adding DMSO to LC mobile phases increases peptide identifications, partly due to charge state coalescence.[30,31] To ensure reaction time normalization holds across a larger range of precursor charge states, we performed a limited tryptic digest of yeast peptides at 4°C for 40 min and analyzed the resulting peptides using ETD (reaction time = 100 ms or NRT ranging from 10 to 100) with ion trap analysis of product ions. This analysis produced precursors with charge states ranging from 3 to 11 and demonstrated unique peptide identifications is optimal at an NRT of 20 or 30, in agreement with the above conclusions (**Figure 9**).

Implementation of Calibrated ETD Reaction Times on a Q-OT-qIT Mass Spectrometer

Based partially on this work, the calibration of ETD reaction times is a standard feature on the newest generation quadrupole-Orbitrap-quadrupole ion trap (Q-OT-qIT) mass spectrometer (Thermo Orbitrap Fusion). An algorithm analogous to that described in **Figure 2** is used to calculate the reagent anion AGC target required to achieve a desired reaction constant for the triply protonated precursor of angiotensin. Within the method editor, an option to "use calibrated ETD times" sets charge state dependent reaction times to ~2τ, corresponding to an NRT of 30 in the experiments detailed above. When used to analyze a complex mixture of peptides, calibrated ETD times result in 38% more MS/MS scans (29,457 vs. 21,261), 50% more PSMs (9,145 vs. 6,076), and 52% more unique peptides (7,915 vs. 5,210) when product

A)

Normalized Reaction Time(%)	Multiple of τ	MS^2 Scans	PSMs	Unique Peptides
0	0	17,191	0	0
10	0.3	16,888	1,355	1,239
20	1.0	14,226	2,029	1,740
30	2.0	15,976	1,899	1,707
40	3.4	14,104	1,806	1,610
50	5.2	13,588	1,464	1,323
60	7.4	12,460	1,031	943
70	10	11,381	474	442
80	13	10,206	144	138
90	16	9,284	111	110
100	20	8,391	26	25

B)

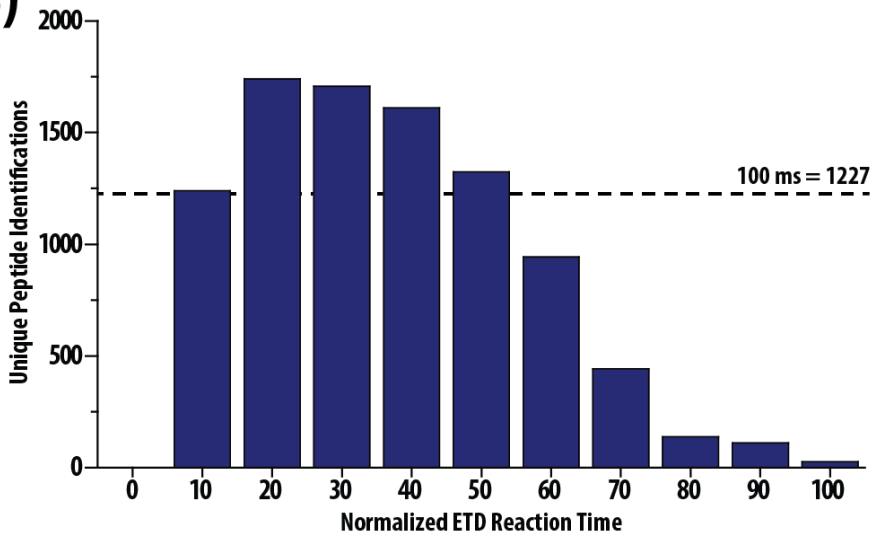


Figure 8. Implementation of a normalized ETD reaction time.
A) Number of unique peptides identified in one nLC-MS/MS experiment utilizing a normalized ETD reaction time. B) Experimental statistics for each nLC-MS/MS experiment evaluating various normalized ETD reaction times.

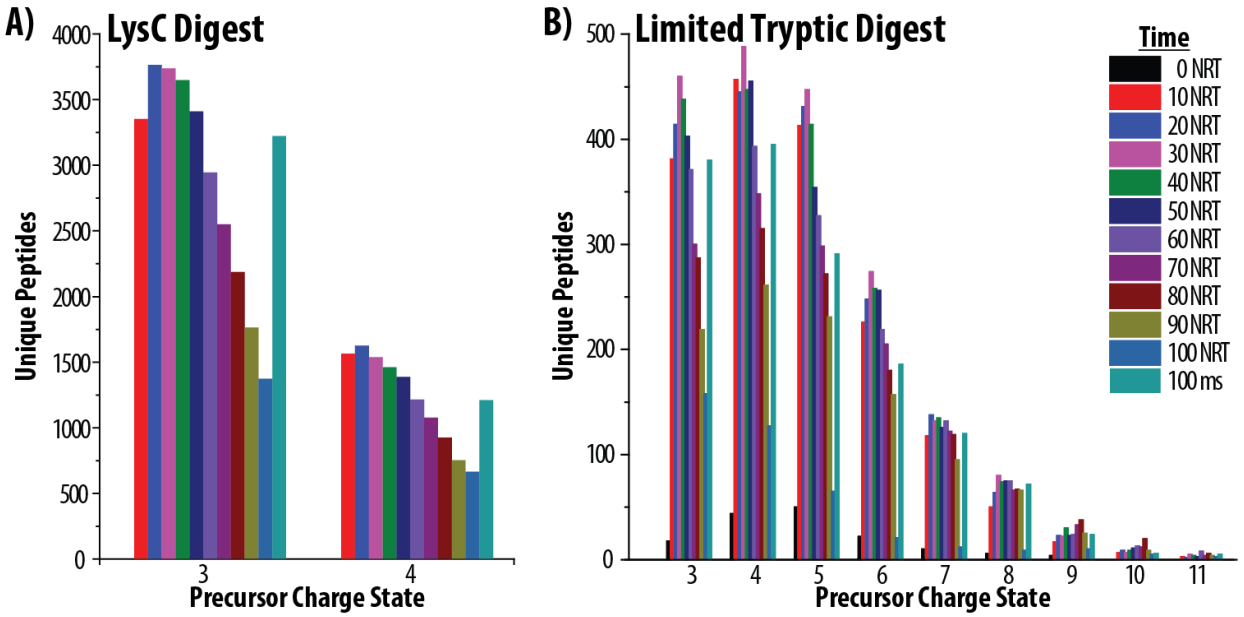


Figure 9. Unique peptide identifications for normalized ETD reaction times.
A) The number of unique peptides by precursor charge state in the analysis of a LysC digest of yeast proteins. B) The number of unique peptides by precursor charge state in the analysis of a limited tryptic digest of yeast proteins.

ions are analyzed in the ion trap (**Figure 10A**). Representative MS/MS spectra of the triply charged peptide VSIAGRIHAK demonstrate lower product ion yield in the 100 ms control (**Figure 10B top**) when compared to the calibrated reaction time of 46.5 ms (**Figure 10B bottom**).

When product ions are analyzed in the Orbitrap the benefit of calibrated times is even greater with 32% more MS/MS scans (26,461 vs. 20,017), 76% more PSMs (6,116 vs. 3,466), and 75% more unique peptides (5,418 vs. 3,090) as compared to a 100 ms control (**Figure 11A**). These results support the conclusion that optimal product ion yield is essential when analyzing ETD reactions in the Orbitrap. By examining the same peptide identified in the ion trap analysis, the benefit of using calibrated ETD times for Orbitrap analysis is demonstrated (**Figure 11B**). Here, the calibrated reaction time of 46.8 ms resulted in a spectrum whose score was superior (3.1×10^{-15}) when compared to the 100 ms control (3.3×10^{-9}). In fact, the improved product ion signal-to-noise is visually apparent when using the calibrated reaction time (**Figure 11B**).

Discussion

Increased speed of newer generation mass spectrometers underscores the importance of optimizing the length of ETD reactions to produce maximal product ion abundance and large-scale peptide identifications.[1,2] Using theoretical models of ion-ion reactions we posited that optimal reaction times could be linked to a chosen ETD reaction rate constant (k). We developed an algorithm that determines the reagent ion AGC target necessary to achieve a desired k for triply charged precursors ($k_3 = 50$). The k_3 was then used to calculate the charge state dependent time constants (τ), the basis for reaction time evaluation. Large scale analysis of more than one hundred peptides in each charge state (2-6) demonstrated that product ion intensity is optimal at ~2τ. We then used τ to implement a normalized reaction time (NRT) algorithm that enables users to vary the completeness of the ETD reactions, finding an NRT between 20 -30 (representing reaction times between 1-2τ) to be optimal. Lastly, we characterized the implementation of a similar algorithm on a quadrupole-Orbitrap-quadrupole ion trap (Q-OT-qIT) mass spectrometer and found a large increase in the number of unique identifications when we utilized calibrated ETD reaction times.

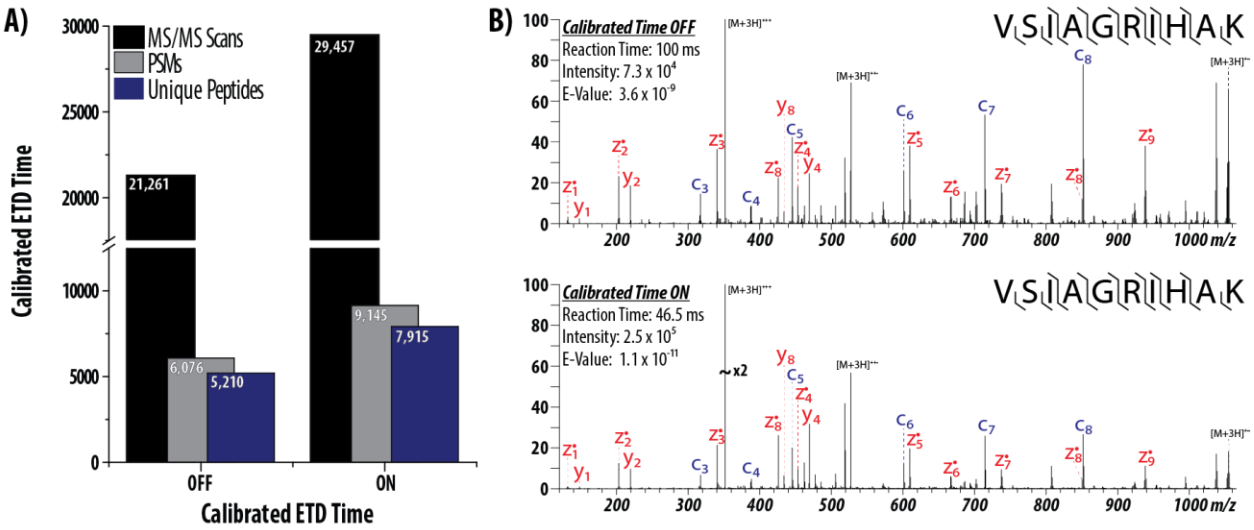


Figure 10. Implementation of calibrated ETD reaction times on a Q-OT-qIT mass spectrometer.

A) The number of MS/MS scans, peptide spectral matches, and unique peptides for each experiment. B) ETD spectra for the same peptide reacted with reagent anions for the calibrated reaction time (46.5 ms) or a control time (100 ms).

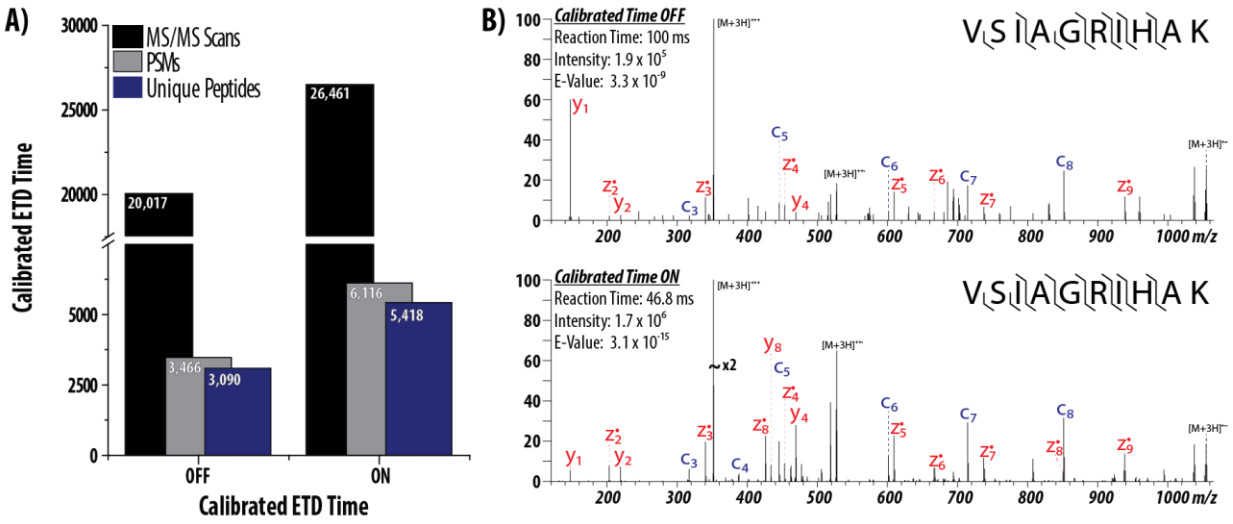


Figure 11. Implementation of calibrated ETD reaction times on a Q-OT-qIT mass spectrometer.

A) The number of MS/MS scans, peptide spectral matches, and unique peptides for each experiment. B) ETD spectra for the same peptide reacted with reagent anions for the calibrated reaction time (46.8 ms) or a control time (100 ms). The calibrated time produces a spectrum with a more significant identification score.

Mass spectrometers continue to increase with respect to the number of MS/MS events performed per second (Hz).[1] The Q-OT-qIT mass spectrometer used in this study has been demonstrated to achieve as high as 20 Hz when analyzing complex samples.[1,2] Faster instruments increase the importance of optimizing ETD reaction times to produce sufficient production ion signal, without negatively affecting duty cycle. In fact, the Q-OT-qIT mass spectrometer has many features (e.g., picking highest charge state rather than most intense precursor) specifically geared toward increasing the speed and probability of identification in ETD analysis.

Lastly, our algorithm specifically characterizes the optimal ETD reaction time for peptides identified in large-scale analysis (i.e., charge state 2-6). However, top down analysis of whole proteins is becoming more accessible and ETD has proven to be efficient for the dissociation of highly charged biomolecules.[7,25,32,33] As with peptides, determining optimal ETD conditions will be vital in ensuring that proteins are not overreacted. Whole proteins are often highly charged, which can quickly lead to overreaction – suggesting a charge state dependent adjustment in the ETD reaction rate (i.e., adjustment of the reagent ion population) might be necessary for highly charged proteins. Experiments similar to those described here could be conducted holding the reaction time (*t*) constant and varying the ETD reaction rate (k) to determine the optimal rate at a given *t*. Expanding the work demonstrated for peptides to proteins will not only simplify the use of ETD, but also aid in its growth to broader applications.

Experimental Procedures

Infusion of Angiotensin I

Angiotensin I was diluted to 2 pmol/µL in 50% ACN with 0.2% FA and directly infused using electrospray ionization. ETD was performed on the triply charge precursor for times ranging from 5 to 160 ms (in steps of 5).

Calibration of ETD Reaction Rate

The routine for calibration of ETD reaction rates was written in the instrument control language to enable automated execution of the algorithm. Briefly, the calibration procedure reacts the triply charged precursor of Angiotensin (AGC = 30,000) with various amounts of reagent ions (20,000 to 450,000) for increasing reaction times (5 to 115 ms). For each reagent AGC target, the ETD reaction rate (k) calculated by plotting $\ln(A/A_o)$, where A_0 is the precursor intensity with no reaction, and A is the precursor intensity after the ETD reaction versus the reaction time. For each reaction time sampled, four spectra were collected to account for scan to scan variation. The resulting data was fitted with a linear function where the slope represented -(k). The values for k at each reagent AGC target were then plotted on a separate graph which was used to determine the reagent AGC target necessary to achieve the desired k.

Yeast Sample Preparation

Yeast were lysed and proteins were reduced and alkylated as previously described.[34] To create large highly charged peptides for the large scale analysis of optimal ETD reaction times, proteins were digested for 40 (final analysis of NRT) or 60 (analysis of product ion yield) min at ~4 °C after the addition of 1 mM $CaCl_2$, 50 mM Tris (to decrease urea to 1 M for trypsin or 4 M for LysC digestion), and adjusting to pH 8 at an enzyme-to-substrate ratio of 1:250 of trypsin (Promega, Madison, WI, USA). For all other experiments the trypsin digestion was replaced by an overnight LysC (Wako) digestion at room temperature with an enzyme-to-substrate ratio of 1:100. All digests were quenched by the addition of TFA to a final concentration of 0.5 % (pH 2), and desalted via solid phase extraction on a 50-mg tC_{18} SepPak cartridge (Waters, Milford, MA, USA).

Mass Spectrometry and High Performance Liquid Chromatography

Large scale characterization of optimal reaction times was performed on an ETD-enabled hybrid, dual cell-quadrupole ion trap-Orbitrap mass spectrometer (Orbitrap Elite, Thermo Fisher Scientific, San Jose, CA). Eleven ETD MS^2 events were performed on the most intense precursor for reaction times varying from 1 to 4x the charge state dependent time constant (τ). Reaction times were randomized in five different

orders, with the exception of times correlating to 2τ which was always performed as the first, middle, and last MS^2 events. The instrument code was modified to facilitate the use of the five random reaction time orders. Experiments evaluating normalized ETD reaction rates were either performed on an ETD-enabled hybrid, dual cell-quadrupole ion trap-Orbitrap mass spectrometer (Orbitrap Elite)[35] or a quadrupole-Orbitrap-quadrupole ion trap (Q-OT-qIT) hybrid mass spectrometer (Orbitrap Fusion, Thermo Fisher Scientific, San Jose, CA).[2] For Large scale characterization of optimal reaction times precursors were selected from MS^1 scans comprising 1,000,000 charges and collected at 60,000 resolution. Separate experiments collecting MS^2 events in the ion trap or Orbitrap (15,000 resolution) using 30,000 or 100,000 precursor charges, respectively. A ±10 ppm window was excluded around the monoisotopic peak of each precursor for 45 sec.

Experiments evaluating normalized ETD reaction rates were either performed on an ETD-enabled hybrid, dual cell-quadrupole ion trap-Orbitrap mass spectrometer (Orbitrap Elite) or a quadrupole-Orbitrap-quadrupole ion trap (Q-OT-qIT) hybrid mass spectrometer (Orbitrap Fusion, Thermo Fisher Scientific, San Jose, CA). For experiments performed on the Orbitrap Elite, MS^1 scans at a resolution of 60,000 were used to guide selection and dissociation of the 20 most intense precursors. Precursors with a charge greater than 2 were reacted with the calibrated number of reagent anions for the specified normalized reaction time and ETD product ions were analyzed in either the Orbitrap (100,000 precursor charges, 15,000 resolution) or the ion trap (30,000 precursor charges) in separate experiments. For both ion trap and Orbitrap analysis, one experiment was performed using an ETD reaction time of 100 ms for all precursors. Precursors were isolated at 1.8 Th and an exclusion window of ±10 ppm was constructed around the monoisotopic peak of each selected precursor for 45 sec. For experiments performed on the Orbitrap Fusion, MS^1 scans guided the selection of precursors such that MS^1 scans were performed every 5 sec. Precursors with a charge greater than 2 were reacted with either the calibrated number of reagent anions or 200,000 reagent anions for the calibrated ETD reaction time or 100 ms, respectively. ETD product ions were analyzed in either the ion trap (30,000 precursor charges) or Orbitrap (100,000 precursor charges, 15,000 resolution). Precursors were

isolated using the quadrupole with an isolation window of ±0.7 Th and an exclusion window of ±10 ppm was constructed around the monoisotopic peak of each selected precursor for 45 sec.

Database Searching and FDR Estimation

Data were searched against a database containing canonical protein and protein isoform sequences from Uniprot (6,563 total entries). For large scale analysis of optimal ETD reaction times the database search was performed using a precursor mass tolerance of ± 4.5 Th, Orbitrap MS^2 product ion tolerance of ± 0.01 Th, and ion trap MS^2 product ion tolerance of ± 0.35 Th while considering up to nine missed cleavages. For all other experiments, a precursor mass tolerance of ±150 ppm was used while considering four isotopes and three missed cleavages. Peptides were considered with carbamidomethylation of cysteine as a fixed modification and oxidation of methionine as a variable modification.

Large-Scale Analysis of ETD Product Ion Intensity

Determination of optimal ETD reaction times was performed on data collected using various ETD reaction times as a function of the charge state dependent time constant (τ). For each charge state (2-6) one LC-MS/MS analysis was conducted analyzing ions in the ion trap and another analyzing ions in the Orbitrap, yielding two raw files for each charge state. Each LC-MS/MS experiment comprised an MS^1 followed by eleven MS^2 events that reacted the most intense precursor for various amounts of time ranging from 1 – 4τ. For each precursor (i.e., set of eleven MS^2 events) the top scoring peptide hit was determined to be the correct identification and the resulting sequence was used to generate a list of all c or z-dot product ions. For precursors with charge greater than 3, m/z values relating to multiply charge product ions were also considered. Product ions were considered with a tolerance of 15 ppm and 0.5 Da for Orbitrap and ion trap analysis, respectively. For each MS^2, product ions were summed and the resulting total was mapped to the appropriate multiple of τ. These values were subsequently normalized to the largest intensity within the set of reaction times (i.e., all MS^2 spectra of the same precursor) resulting in a score of 0.0 to 1.0 for each multiple of τ and enabling the comparison of all identified peptides across the LC-MS/MS experiment.

Precursors that included one or more MS^2 scans that did not reach the requested number of ions (i.e., injection time equal to max inject time) were excluded from analysis. MS^2 scans were collected at an ETD reaction time of 2τ for scan event two, seven, and twelve to ensure that elution of the peptide did not affect analysis. For ion trap analysis the first 2τ scan was collected in the Orbitrap to increase the probability of identification.

Acknowledgements

The authors gratefully acknowledge support from Thermo Fisher Scientific, NSF CAREER grant 0747990, and NIH grant R01 GM080148. C.M.R was funded by an NSF Graduate Research Fellowship and NIH Traineeship (T32GM008505). A.E.M gratefully acknowledges support from an NIH-funded Genomic Sciences Training Program (5T32HG002760).

References

1. Hebert AS, Richards AL, Bailey DJ, Ulbrich A, Coughlin EE, Westphall MS, & Coon JJ. The One Hour Yeast Proteome. *Molecular & Cellular Proteomics*, 2014, 13(1):339-347.

2. Senko MW, Remes PM, Canterbury JD, Mathur R, Song Q, Eliuk SM, Mullen C, Earley L, Hardman M, Blethrow JD, Bui H, Specht A, Lange O, Denisov E, Makarov A, Horning S, & Zabrouskov V. Novel Parallelized Quadrupole/Linear Ion Trap/Orbitrap Tribrid Mass Spectrometer Improving Proteome Coverage and Peptide Identification Rates. *Analytical Chemistry*, 2013, 85(24):11710-11714.

3. Syka JEP, Coon JJ, Schroeder MJ, Shabanowitz J, & Hunt DF. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proceedings of the National Academy of Sciences of the United States of America*, 2004, 101(26):9528-9533.

4. Coon JJ, Syka JEP, Schwartz JC, Shabanowitz J, & Hunt DF. Anion dependence in the partitioning between proton and electron transfer in ion/ion reactions. *International Journal of Mass Spectrometry*, 2004, 236(1–3):33-42.

5. Swaney DL, McAlister GC, & Coon JJ. Decision tree-driven tandem mass spectrometry for shotgun proteomics. *Nat Meth*, 2008, 5(11):959-964.

6. Swaney DL, Wenger CD, Thomson JA, & Coon JJ. Human embryonic stem cell phosphoproteome revealed by electron transfer dissociation tandem mass spectrometry. *Proceedings of the National Academy of Sciences*, 2009, 106(4):995-1000.

7. Frese CK, Altelaar AFM, Hennrich ML, Nolting D, Zeller M, Griep-Raming J, Heck AJR, & Mohammed S. Improved Peptide Identification by Targeted Fragmentation Using CID, HCD and ETD on an LTQ-Orbitrap Velos. *Journal of Proteome Research*, 2011, 10(5):2377-2388.

8. Rose CM, Venkateshwaran M, Volkening JD, Grimsrud PA, Maeda J, Bailey DJ, Park K, Howes-Podoll M, den Os D, Yeun LH, Westphall MS, Sussman MR, Ané J-M, & Coon JJ. Rapid Phosphoproteomic and Transcriptomic Changes in the Rhizobia-legume Symbiosis. *Molecular & Cellular Proteomics*, 2012, 11(9):724-744.

9. Hunt DF, Buko AM, Ballard JM, Shabanowitz J, & Giordani AB. Sequence analysis of polypeptides by collision activated dissociation on a triple quadrupole mass spectrometer. *Biological Mass Spectrometry*, 1981, 8(9):397-408.

10. Olsen JV, Macek B, Lange O, Makarov A, Horning S, & Mann M. Higher-energy C-trap dissociation for peptide modification analysis. *Nat Meth*, 2007, 4(9):709-712.

11. Zhang Y, Ficarro SB, Li S, & Marto JA. Optimized Orbitrap HCD for Quantitative Analysis of Phosphopeptides. *J. Am. Soc. Mass Spectrom.*, 2009, 20(8):1425-1434.

12. McAlister GC, Phanstiel DH, Brumbaugh J, Westphall MS, & Coon JJ. Higher-energy Collision-activated Dissociation Without a Dedicated Collision Cell. *Molecular & Cellular Proteomics*, 2011, 10(5).

13. Phanstiel D, Brumbaugh J, Berggren WT, Conard K, Feng X, Levenstein ME, McAlister GC, Thomson JA, & Coon JJ. Mass spectrometry identifies and quantifies 74 unique histone H4 isoforms

in differentiating human embryonic stem cells. *Proceedings of the National Academy of Sciences*, 2008, 105(11):4093-4098.

14. Phanstiel D, Zhang Y, Marto J, & Coon J. Peptide and protein quantification using iTRAQ with electron transfer dissociation. *J. Am. Soc. Mass Spectrom.*, 2008, 19(9):1255-1262.

15. Good D, Wenger C, McAlister G, Bai D, Hunt D, & Coon J. Post-acquisition ETD spectral processing for increased peptide identifications. *J. Am. Soc. Mass Spectrom.*, 2009, 20(8):1435-1440.

16. Phanstiel D, Unwin R, McAlister GC, & Coon JJ. Peptide Quantification Using 8-Plex Isobaric Tags and Electron Transfer Dissociation Tandem Mass Spectrometry. *Analytical Chemistry*, 2009, 81(4):1693-1698.

17. Grimsrud PA, den Os D, Wenger CD, Swaney DL, Schwartz D, Sussman MR, Ané J-M, & Coon JJ. Large-Scale Phosphoprotein Analysis in Medicago truncatula Roots Provides Insight into in Vivo Kinase Activity in Legumes. *Plant Physiology*, 2010, 152(1):19-28.

18. Xia Q, Lee MV, Rose C, Marsh A, Hubler S, Wenger C, & Coon J. Characterization and Diagnostic Value of Amino Acid Side Chain Neutral Losses Following Electron-Transfer Dissociation. *J. Am. Soc. Mass Spectrom.*, 2011, 22(2):255-264.

19. Bailey DJ, Rose CM, McAlister GC, Brumbaugh J, Yu P, Wenger CD, Westphall MS, Thomson JA, & Coon JJ. Instant spectral assignment for advanced decision tree-driven mass spectrometry. *Proceedings of the National Academy of Sciences*, 2012, 109(22):8411-8416.

20. Richards AL, Vincent CE, Guthals A, Rose CM, Westphall MS, Bandeira N, & Coon JJ. Neutron-encoded Signatures Enable Product Ion Annotation From Tandem Mass Spectra. *Molecular & Cellular Proteomics*, 2013, 12(12):3812-3823.

21. Compton PD, Strukl JV, Bai DL, Shabanowitz J, & Hunt DF. Optimization of Electron Transfer Dissociation via Informed Selection of Reagents and Operating Parameters. *Analytical Chemistry*, 2011, 84(3):1781-1785.

22. McLuckey SA, Stephenson JL, & Asano KG. Ion/Ion Proton-Transfer Kinetics: Implications for Analysis of Ions Derived from Electrospray of Protein Mixtures. *Analytical Chemistry*, 1998, 70(6):1198-1202.

23. Tolmachev AV, Udseth HR, & Smith RD. Modeling the ion density distribution in collisional cooling RF multipole ion guides. *International Journal of Mass Spectrometry*, 2003, 222(1–3):155-174.

24. Good DM, Wirtala M, McAlister GC, & Coon JJ. Performance Characteristics of Electron Transfer Dissociation Mass Spectrometry. *Molecular & Cellular Proteomics*, 2007, 6(11):1942-1951.

25. Rose C, Russell J, Ledvina A, McAlister G, Westphall M, Griep-Raming J, Schwartz J, Coon J, & Syka JP. Multipurpose Dissociation Cell for Enhanced ETD of Intact Protein Species. *J. Am. Soc. Mass Spectrom.*, 2013, 24(6):816-827.

26. Williams DK, McAlister GC, Good DM, Coon JJ, & Muddiman DC. Dual Electrospray Ion Source for Electron-Transfer Dissociation on a Hybrid Linear Ion Trap−Orbitrap Mass Spectrometer. *Analytical Chemistry*, 2007, 79(20):7916-7919.

27. Myers SA, Daou S, Affar EB, & Burlingame A. Electron transfer dissociation (ETD): The mass spectrometric breakthrough essential for O-GlcNAc protein site assignments—a study of the O-GlcNAcylated protein Host Cell Factor C1. *PROTEOMICS*, 2013, 13(6):982-991.

28. Frey B, Ladror D, Sondalle S, Krusemark C, Jue A, Coon J, & Smith L. Chemical Derivatization of Peptide Carboxyl Groups for Highly Efficient Electron Transfer Dissociation. *J. Am. Soc. Mass Spectrom.*, 2013, 24(11):1710-1721.

29. Eliuk SM, Maltby D, Panning B, & Burlingame AL. High Resolution Electron Transfer Dissociation Studies of Unfractionated Intact Histones from Murine Embryonic Stem Cells Using On-line Capillary LC Separation: DETERMINATION OF ABUNDANT HISTONE ISOFORMS AND POST-TRANSLATIONAL MODIFICATIONS. *Molecular & Cellular Proteomics*, 2010, 9(5):824-837.

30. Meyer J & A. Komives E. Charge State Coalescence During Electrospray Ionization Improves Peptide Identification by Tandem Mass Spectrometry. *J. Am. Soc. Mass Spectrom.*, 2012, 23(8):1390-1399.

31. Hahne H, Pachl F, Ruprecht B, Maier SK, Klaeger S, Helm D, Medard G, Wilm M, Lemeer S, & Kuster B. DMSO enhances electrospray response, boosting sensitivity of proteomic experiments. *Nat Meth*, 2013, 10(10):989-991.

32. Fornelli L, Damoc E, Thomas PM, Kelleher NL, Aizikov K, Denisov E, Makarov A, & Tsybin YO. Analysis of Intact Monoclonal Antibody IgG1 by Electron Transfer Dissociation Orbitrap FTMS. *Molecular & Cellular Proteomics*, 2012, 11(12):1758-1767.

33. Rhoads TW, Rose CM, Bailey DJ, Riley NM, Molden RC, Nestler AJ, Merrill AE, Smith LM, Hebert AS, Westphall MS, Pagliarini DJ, Garcia BA, & Coon JJ. Neutron-Encoded Mass Signatures for Quantitative Top-Down Proteomics. *Analytical Chemistry*, 2014, 86(5):2314-2319.

34. Merrill AE, Hebert AS, MacGilvray ME, Rose CM, Bailey DJ, Bradley JC, Wood WW, ElMasri M, Westphall MS, Gasch AP, & Coon JJ. NeuCode labels for relative protein quantification. *Molecular & Cellular Proteomics*, 2014.

35. Michalski A, Damoc E, Lange O, Denisov E, Nolting D, Müller M, Viner R, Schwartz J, Remes P, Belford M, Dunyach J-J, Cox J, Horning S, Mann M, & Makarov A. Ultra High Resolution Linear Ion Trap Orbitrap Mass Spectrometer (Orbitrap Elite) Facilitates Top Down LC MS/MS and Versatile Peptide Fragmentation Modes. *Molecular & Cellular Proteomics*, 2012, 11(3).

Chapter 7

Neutron-encoded labeling for peptide identification

CMR designed research, performed experiments, analyzed data, and wrote the paper.

This chapter has been published:

Rose C. M., Merrill A. E., Bailey D. J., Hebert A. S., Westphall M. S., Coon J. J. "Neutron Encoded Labeling for Peptide Identification." *Analytical Chemistry*, **2013**, 85(10):5129-5137.

Abstract

Metabolic labeling of cells using heavy amino acids is most commonly used for relative quantitation, however, partner mass shifts also detail the number of heavy amino acids contained within the precursor species. Here we use a recently developed metabolic labeling technique, NeuCode SILAC, which produces precursor partners spaced ~ 40 mDa apart to enable amino acid counting. We implement large scale counting of amino acids through a program, "Amino Acid Counter", which determines the most likely combination of amino acids within a precursor based on NeuCode SILAC partner spacing and filters candidate peptide sequences during a database search using this information. Counting the number of lysine residues for precursors selected for MS/MS decreases the median number of candidate sequences from 44 to 14 as compared to an accurate mass search alone (20 ppm). Furthermore, the ability to co-isolate and fragment NeuCode SILAC partners enables counting of lysines in product ions, and when the information is used, the median number of candidates is reduced to 7. We then demonstrate counting leucine in addition to lysine results in a six-fold decrease in search space, 43 to 7, when compared to an accurate mass search. We use this scheme to analyze a nanoLC-MS/MS experiment and demonstrate that accurate mass plus lysine and leucine counting reduces the number of candidate sequences to one for ~ 20% of all precursors selected, demonstrating an ability to identify precursors without MS/MS analysis.

Introduction

Shotgun proteomic analysis by mass spectrometry (MS) relies on selection of peptide precursor ions of interest from a survey scan (MS^1) followed by dissociation and product ion analysis (i.e., tandem MS, MS/MS).[1] Tens of thousands of MS/MS spectra are then searched against an *in silico* database of predicted spectra generated by an *in silico* digest of the proteome in question to produce a peptide sequence match.[2-7] Regardless of acquisition strategy, the precursor peptide sequence is primarily determined from the product ions within the MS^2 spectrum, while the MS^1 is used to determine the precursor mass to narrow the possible peptide matches from the *in silico* proteome digest.[8-10] A trend in MS instrumentation has been increased MS/MS data acquisition rates from 1 Hz several years ago to as much as 10 Hz today.[11-14] With this boosted sampling rate have come increasingly deeper proteome coverage proteome with more and more peptides identified with each newer, i.e., faster, generation of instruments. We reason that if some fraction of peptide precursors did not require MS/MS scanning to generate sequence identification, the bandwidth limited MS/MS scans could be allocated to only those precursor candidates whose sequence cannot be deduced by accurate mass alone.

Many researchers have examined the use of accurate mass to determine peptide sequence.[15-21] Even with ± 1 ppm mass accuracy only peptides smaller than 700-800 Da can unambiguously identified by mass alone.[22] From this body of work we conclude that other information must be obtained if we wish to avoid tandem MS for even a modest fraction of precursors. A decade ago, Stephenson et al. described the use of isoelectric focusing with accurate mass for peptide identification.[23] This method, however, required knowledge of precursor isoelectric point – information that is not typically available for all proteomic applications. Many modern proteomic experiments often utilize heavy amino acids for quantitative purpose, e.g., stable isotope labeling with amino acids in cell culture (SILAC).[24,25] Here MS^1 spectra contain more information than just precursor mass. Specifically, the mass shift imparted by the incorporation of heavy amino acids, typically 4 - 10 Da, discloses the number of heavy amino acids contained in the peptide, and, therefore the exact number of the labeled residues present in that particular peptide.

Amino acid counting was used with high mass accuracy Fourier transform ion cyclotron resonance (FT-ICR)[26-28] mass spectrometry to aid in whole protein identification using a variety of heavy amino acids including leucine, lysine, arginine, histidine, phenylalanine, and cysteine.[29-33] Furthermore, peptide mass fingerprinting techniques demonstrated an ability to identify proteins by accurate mass of peptides and amino acid content determined by counting isotopically labeled amino acids.[34-36] This technique was expanded to peptide sequence analysis where mass shifts in the MS^1 used to determine amino acid content and add confidence to peptide identifications, as well as infer peptide modifications.[37-40] Amino acid counting has also been used to improve *de novo* sequencing by analyzing induced mass shifts in MS^2 spectra.[41-43] More recently, amino acid counting has been employed prior to database search of peptides to limit the number of peptides considered based on the number of lysines and/or arginine residues in addition to accurate mass.[21] Obviously, increasing the number of labeled amino acids increases the informing power and the likelihood that a single peptide sequence can be mapped to a precursor *m/z* peak. A considerable limitation of the traditional SILAC approach is that adding many distinct heavy amino acids can rapidly increase MS^1 spectral complexity – using labeled amino acids with relatively large partner mass shifts (4 – 10 Da).[44] These large spacings, which are necessary to limit isotopic envelope overlap, increase MS^1 spectral complexity, limit the number of precursors that can be selected in a typical LC-MS/MS analysis, and decrease overall protein identifications.[45]

We have recently described a novel metabolic labeling strategy termed neutron encoding (NeuCode) that exploits the discrepancy of neutron binding energetics between isotopes of C, N, and H.[46] NeuCode utilizes isotopologues of amino acids that differ in mass by as little as 6 mDa and as much as ~ 40 mDa. These small mass differences are not detectable at mass resolutions lower than ~ 100,000, but are revealed under ultra-high resolution analysis that is achievable on modern FT-MS systems, including the Orbitrap analyzer.[12,14,47-51] Reducing partner spacing from 4 - 10 Da to < 40 mDa with NeuCode SILAC compresses the various precursor species in *m/z* space – space so small that only one form of the peptide is detected under typical mass resolution conditions. This approach lowers the number of precursor species

in MS^1 spectra, eliminates redundant sampling of peptide partners, increases overall protein identifications, and enables amino acid counting with multiple labeled amino acids.[46]

Here we explore the use of NeuCode SILAC to map peptide mass-to-charge (*m/z*) signals to peptide sequence without need for MS/MS analysis. A major benefit of the proposed methodology is derived from the use of NeuCode where very small mass differences permit us to use multiple amino acids, which greatly improve identification specificity with no complication of the MS1 spectra. Using theoretical calculations we determined that accurate mass measurement of the intact precursor, and knowledge of the number of Leu and Lys residues it contains (via NeuCode SILAC), would provide sufficient information to map the precursor to sequence half of the time (51%). With these enticing theoretical results we tested this idea experimentally. We developed custom software, AminoAcidCounter, and demonstrate the application of amino acid counting significantly reduces the number candidate sequences more than three-fold for lysine and six-fold for lysine and leucine counting. We also demonstrate lysine counting within product ions further reduces the number of peptides considered two-fold as compared to lysine counting alone. Finally, we use this amino acid counting strategy to analyze all precursors and find that 916 (19.5% of all selected precursors) are found to have one potential peptide after accurate mass, lysine, and leucine filtering, demonstrating the potential to identify peptides from MS^1 spectra alone.

Results

NeuCode SILAC Enables Amino Acid Counting

To test our NeuCode-facilitated amino acid counting strategy we developed a scheme to count the number of specified amino acids contained within a given precursor by analyzing the NeuCode isotopic fine structure from an MS^1 spectrum (Figure 1). To embed these NeuCode labels we began with two isotopologues of lysine (Lys-$^{13}C_6$, $^{15}N_2$ or Lys-D_8) (**Figure 1A**). Both of these amino acids are 8 Da heavier than natural Lys; however, they differ from each other by only 36 mDa and are only detected at resolutions greater than ~ 200,000 (**Figure 1B and C**). The close proximity of these NeuCode partners maintains a low complexity MS^1 scan, as compared to traditional SILAC, and allows for facile isolation of both partners

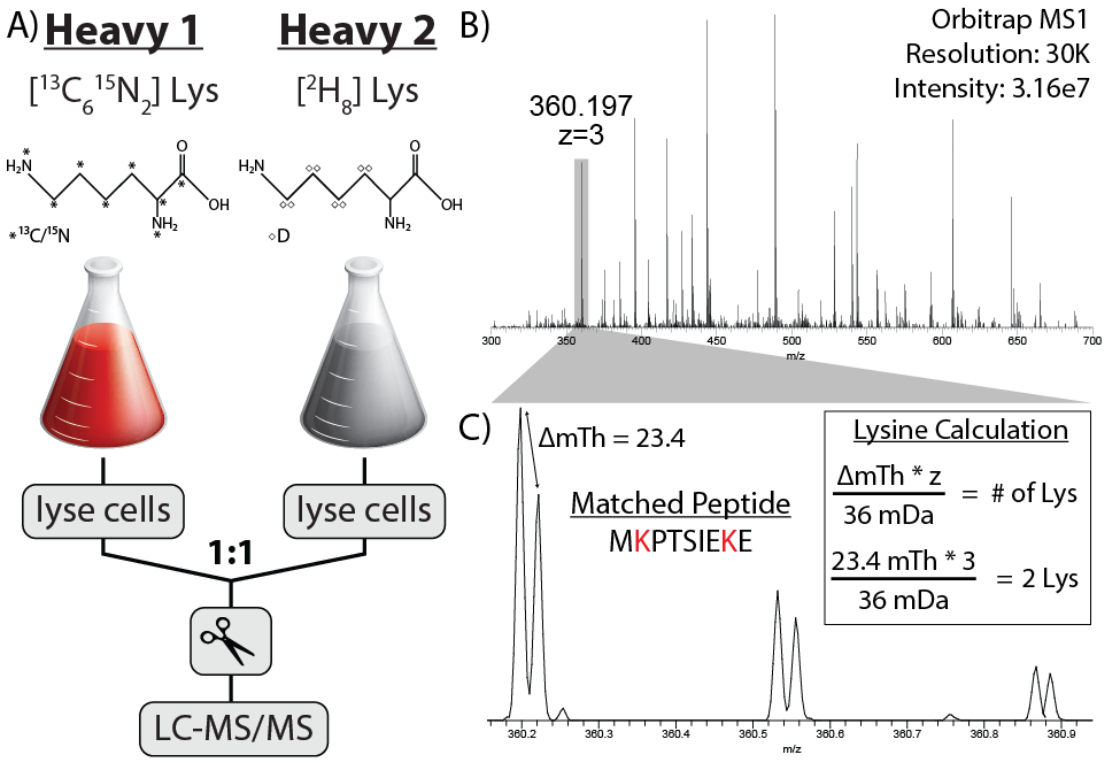


Figure 1. Amino Acid Counting with NeuCode SILAC.
A) Experimental design of lysine NeuCode SILAC utilizing either "heavy 1" lysine (Lys-$^{13}C_6$,$^{15}N_2$) or "heavy 2" lysine (Lys-D_8). B) A survey scan acquired in the Orbitrap was used to select a triply charged precursor at *m/z* 360.197. C) A precursor at *m/z* 630.197 displays the presence of two distinct peaks separated by 23.4 mTh, created by the incorporation of two lysine isotopologues. This was confirmed by a database search, which matched it to sequence, MKPTSIEKE.

for concurrent analysis in the MS/MS. Note the spacing of the two peptide isotopologues within the MS^1 spectrum still facilitates relative quantitation, provided they are resolved. Our intent here, however, is to enable the calculation of lysine count for each precursor. For example, the triply charged precursor at *m/z* 360.197 had a partner located 23.4 mTh higher in *m/z* space (**Figure 1C**). To calculate the mass difference in Da, the *m/z* difference was multiplied by the precursor *z* and divided by the mass difference imparted by one lysine (36.0 mDa, **Figure 1C**, inset). The arithmetic reveals this particular precursor peptide contains two lysine residues – a result confirmed upon database searching of the tandem MS/MS scan (MKPTSIEKE).

Theoretical Calculations of Reduction in Precursor Search Space

Traditional SILAC data analysis often begins with the assignment of peptide pairs before a database search, enabling lysine and/or arginine counting to limit the precursor search space by accurate mass (20 ppm) and then the number of counted amino acids.[21] We surmised that alternate enzyme digestion and additional heavy amino acids could further limit search space, ideally to one candidate sequence, perhaps enabling identifications of amino acids based solely on MS^1 information. To determine the best enzyme and heavy amino acid combination, 100,000 randomly selected yeast peptides were searched against all peptides from an *in silico* digest and filtered based on accurate mass, accurate mass and the count of one amino acid, or accurate mass and the count of two amino acids. For these calculations we used LysC and GluC and in addition to lysine, leucine was used due to its prevalence within the proteome as well as the availability of appropriate heavy isotopologues. Lysine and leucine are also used as heavy amino acids in yeast culture providing confidence that an experimental strategy which uses both amino acids in culture was feasible.[34,52,53]

Using LysC and accurate mass alone, the median number of candidate sequences for a precursor from 100,000 random yeast peptides is 18 (**Figure 2A**). If these peptides are filtered by accurate mass and the number of lysines or the number of lysines and leucines, the median number of candidates falls to 6 and 2, respectively (**Figure 2A**). Here, 38.2% (38,167 of 100,000) precursor *m/z* peaks have only one candidate

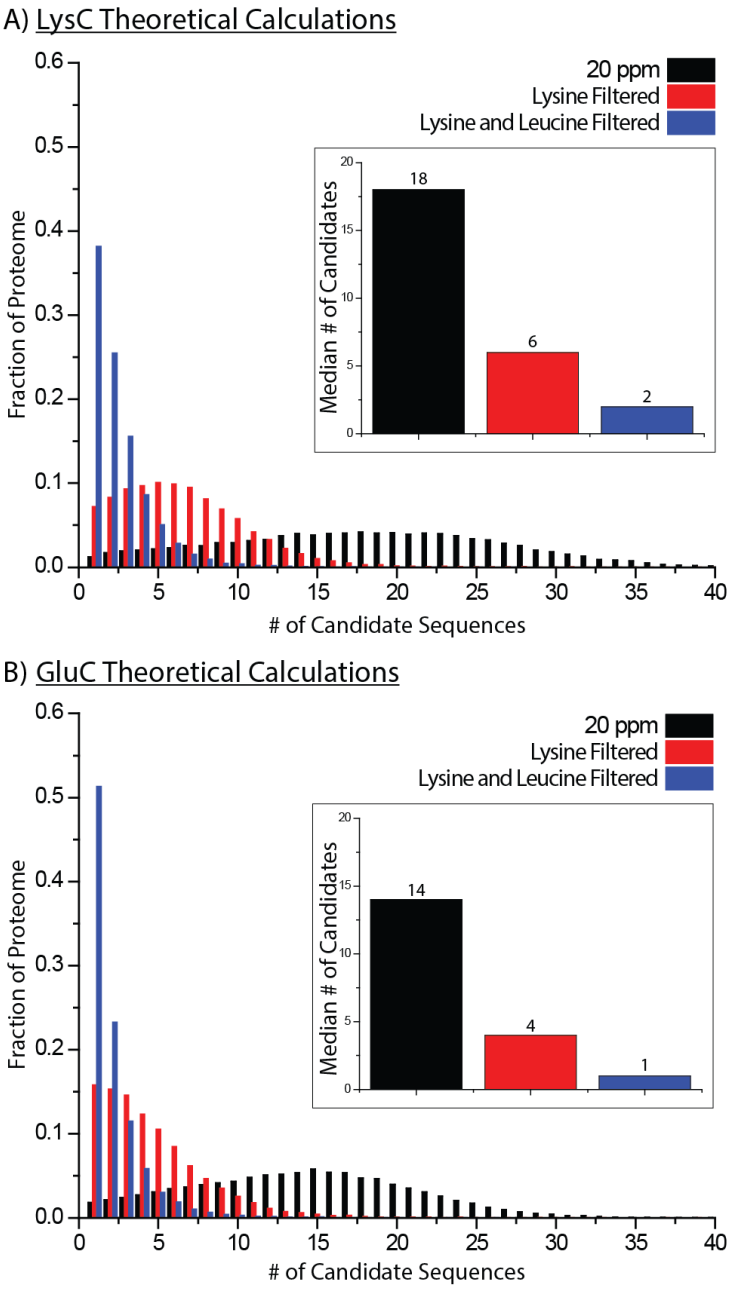


Figure 2. Theoretical Calculations Using Amino Acid Counting.
A) LysC Theoretical Calculations. 100,000 random yeast LysC peptides were searched against an entire *in silico* LysC digest of the yeast proteome with a precursor search tolerance of 20 ppm. The list of resulting peptides was then narrowed if the candidate peptide contained the same number of lysines or lysines and leucines as the peptides used for the search. B) GluC Theoretical Calculations. Identical analysis to (A) was performed, but GluC was used as the enzyme for *in silico* digest of the yeast proteome.

sequence remaining after all three filters are applied. When GluC peptides are filtered for accurate mass and the number of lysines and leucines the median number of sequences under consideration for each precursor falls to 1 and 51.2% (51,299 of 100,000) of precursor *m/z* peaks have only one candidate sequence that fits all three criteria (**Figure 2B**). This exciting calculation prompted us to proceed to test the idea experimentally. Specifically, we labeled yeast with either NeuCode Lysine or NeuCode Lysine and leucine, and digested the resulting proteins with GluC.

NeuCode SILAC to Count Lysines

To test NeuCode SILAC for MS^1 identifications, proteins from yeast cultured as shown in Figure 1A were mixed 1:1, digested by GluC, and analyzed in four nanoLC-MS/MS experiments. Tandem mass spectra from all four raw files were searched against an *in silico* GluC digest of the yeast proteome using traditional database correlation (1% FDR) to produce confident peptide assignments.[54] Automatic counting of lysines within precursors was performed by an in-house program, "AminoAcidCounter" (AACounter). AACounter uses precursor *m/z* and *z* to analyze raw data and calculate a score for potential peptide isotopologue partners separated by the mass shift imparted heavy amino acid combinations. For example, a triply charged precursor selected for MS/MS analysis at 442.970 is shown in **Figure 3A**. A zoomed view of this precursor demonstrates peptide pairs separated in the MS^1 by 23.6 mTh, a shift imparted by two lysine residues (**Figure 3B**). An accurate mass (20 ppm) search of GluC peptides within AACounter returned 11 candidate sequences, only one of which had two lysine residues, LNVPKSKALVLE – this sequence was then confirmed by database search of the corresponding MS/MS scan (high confidence match value = 3.51e-5). This result demonstrates the ability to identify a peptide based solely on information gathered from MS^1 spectra. We mapped 6,678 tandem mass spectra to sequence and used these identifications to determine how many of these could be uniquely identified using amino acid counting and accurate mass. We determined that amino acid counting reduced search space as the median number of candidate sequences per precursor falls more than three-fold, from 44 to 14 (**Figure 3D-E**), while the sensitivity (calculated by the number of times the correct peptide still remains under consideration) remains

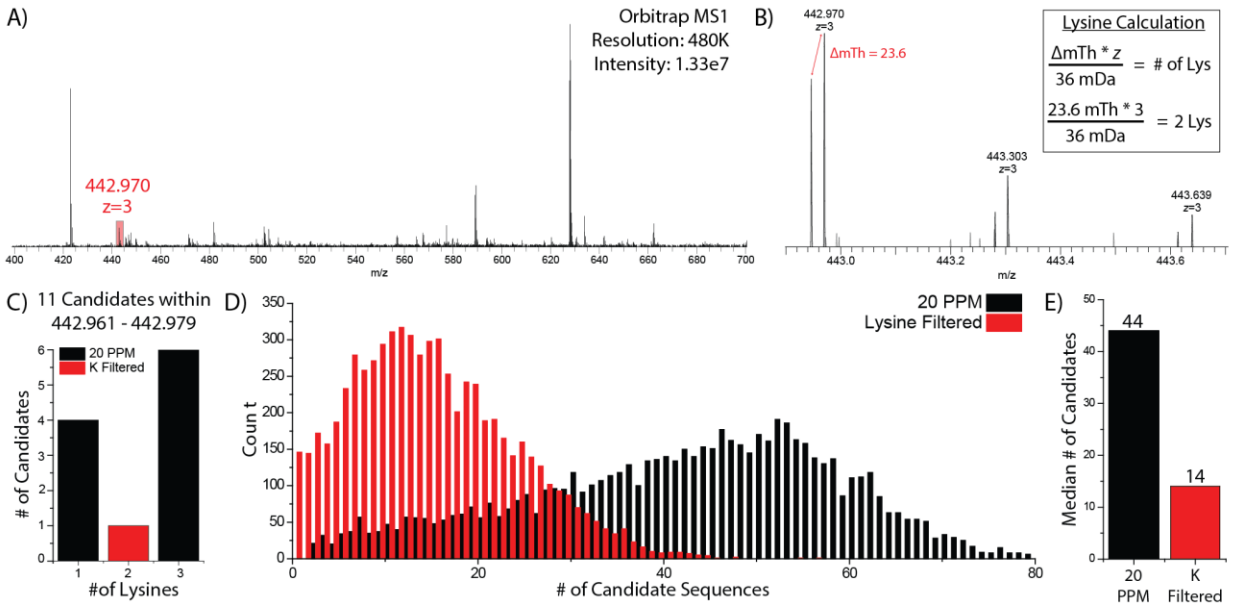


Figure 3. Lysine NeuCode SILAC Enabled Amino Acid Counting.

A) A survey scan acquired in the Orbitrap was used to select a triply charged precursor at *m/z* 442.970. B) A precursor at *m/z* 442.970 displays the presence of two distinct peaks separated by 23.6 mTh, created by the incorporation of 2 lysine isotopologues. C) A search of an *in silico* digest of the yeast proteome with a 20 ppm precursor tolerance returned 11 candidate peptides, only one of which contained two lysines, LNVPKSKALVLE. D) The number of peptides remaining after accurate mass (20 ppm) and subsequent lysine filtering of 6,678 precursors were plotted as a histogram. E) Median number of peptides considered after filtering.

high at 95.6%. Note, the FDR of AAC is slightly higher than typical peptide FDR (1%) due to incorrect scoring and higher number of false positives with low signal-to-noise precursors (**Figure 5**). From these data we conclude that NeuCode SILAC amino acid counting with a single amino acid provides significant benefits, but does not often remove the requirement for tandem MS.

Using MS^2 Information to Further Reduce Precursor Search Space

Lysine NeuCode SILAC partners are closely spaced (<40 mDa) are always co-fragmented as mass filtering devices do not have mTh resolution. Utilizing this characteristic we wondered how beneficial would the NeuCode labels be for improving identification by amino acid counting from product ion spectra. **Figure 4A** displays the MS/MS spectrum of a triply charged peptide TGVIKPGMVVTFAPAGVTTE dissociated by CAD and analyzed at high resolution (to reveal the NeuCode signatures). Both isotopologues of this peptide nearly coeluted, enabling co-fragmentation. Any product ion containing a NeuCode labeled lysine should appear as a doublet. Detection of these doublets, therefore, reveals which fragments have Lys and how many. Figure 4B displays a typical spectrum. AACounter was modified to create *in silico* fragments for each light and heavy peptide partner that was under consideration following accurate mass and lysine filtering. Putative sequences remained under consideration if at least two lysine-containing product ions contained both the "heavy 1" and "heavy 2" partners within the MS^2 spectrum. Application of this filter reduced the median number of candidate sequences considered per precursor from 13, with accurate mass and lysine filtering, to 7 with accurate mass, lysine, and MS^2 filtering with only a 5% decrease in sensitivity (**Figure 4C**). Meanwhile, the number of precursors that correctly matched to one candidate sequence increased almost three-fold from 36 with accurate mass and lysine filtering to 105 with accurate mass, lysine, and MS^2 filtering. We reason that as the time penalty for collecting high resolution MS/MS scans is reduced with newer instrumentation, this type of approach may become quite useful for aiding in confident peptide identification.

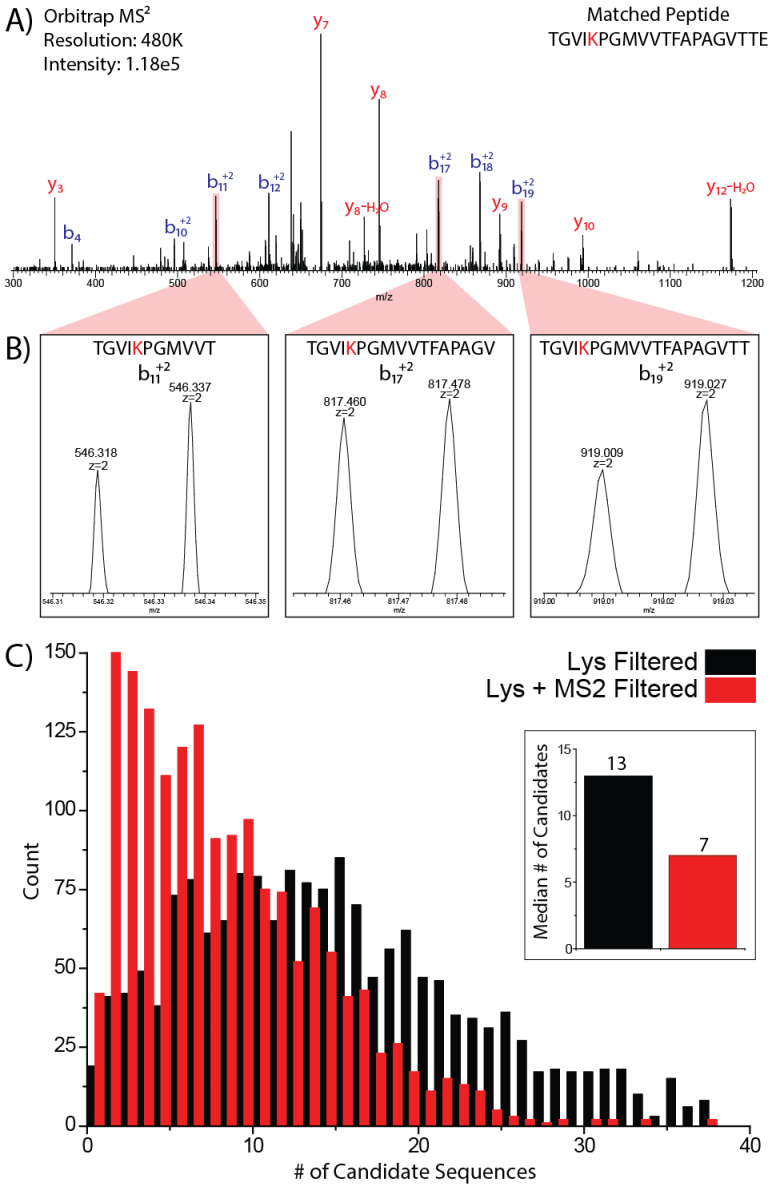


Figure 4. MS^2 Analysis Further Narrows Precursor Search Space.
A) Annotated high resolution MS^2 spectrum of the peptide TGVIKPGMVVTFAPAGVTTE. B) Enlarged view of lysine-containing product ions. Three doubly charged product ions (b_{11}, b_{17}, b_{19}) exhibit *m/z* shifts of ~18 mTh, or ~36 mDa when corrected for charge state. C) Lysine and MS^2 filtering applied to unique peptide spectral matches. For 1,652 unique peptide spectral matches the number of candidates remaining after accurate mass (20 ppm) plus lysine and subsequent MS^2 filtering were plotted as a histogram.

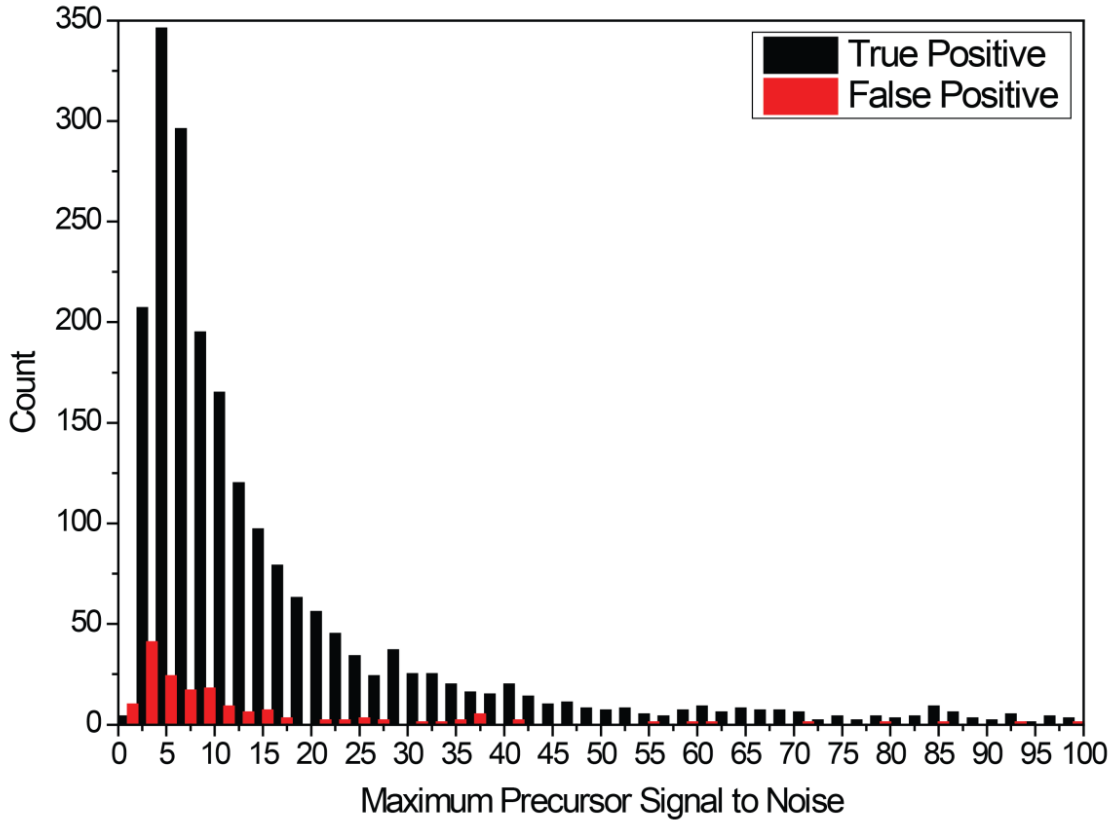


Figure 5. Number of true and false positive results as a function of maximum precursor signal to noise.

For all scored precursors AAC determined the likely number of lysine residues and filtered possible peptide matches based on that information in addition to accurate mass. True positives represent cases where the peptide identified by database search is still under consideration, whereas false positives represent cases where the correct peptide was eliminated as a choice.

NeuCode SILAC to Count Lysine and Leucine

Still hoping to demonstrate that NeuCode amino acid counting could eliminate the need for MS/MS scanning for some fraction of precursors we tested the peptides that had been labeled with both Lys and Leu (Figure 5A.). Here we combine NeuCode isotopologues of both amino acids to label yeast proteins, mixed the proteins 1:1, digested with GluC, and analyzed in five separate nanoLC-MS/MS experiments. Panel B of Figure 5 presents a MS^1 spectrum of a +4 precursor at *m/z* 525.310 that was selected for MS/MS analysis. An expanded view of this mono-isotopic *m/z* peak of the precursor reveals two partner isotopologues separated by 95.6 mDa, indicating the presence of two lysines and one leucine (**Figure 6C**). An accurate mass search of this precursor (20 ppm) returned 17 potential peptides, only one of which met the amino acid criteria AKAQGVAVQLKRQPAQPRE (**Figure 6D**). This putative match was confirmed by searching the MS/MS scan with conventional database matching algorithms. Note, the incorporation of multiple deuterated lysine and leucine residues induces a significant retention time shift of this peptide.

AACounter was modified to enable automated leucine and lysine counting for 5,571 peptide identifications across five nanoLC-MS/MS experiments (typically ~3,000 peptide identifications per run). The distribution of the number of sequence candidates remaining based on accurate mass (20 ppm) or accurate mass and lysine and leucine filtering is displayed in **Figure 6E**. Impressively, the median number of peptides under consideration for each precursor falls more than six-fold from 43 to 7, while sensitivity was high at 87% (**Figure 6F**). In addition, 276 precursors were found to have only one potential peptide after accurate mass and lysine and leucine filtering, of which 209 (75%) were found to be correct (as judged by database searching), signifying an ability to identify peptides without subjecting precursors to MS/MS analysis. Note, the FDR of AACounter is ~25% for one candidate precursors due to inaccuracies of scoring low signal-to-noise precursors, similar to lysine results in **Figure 5**, as well as the increased number of isotopologue pairs when scoring lysine and leucine combinations.

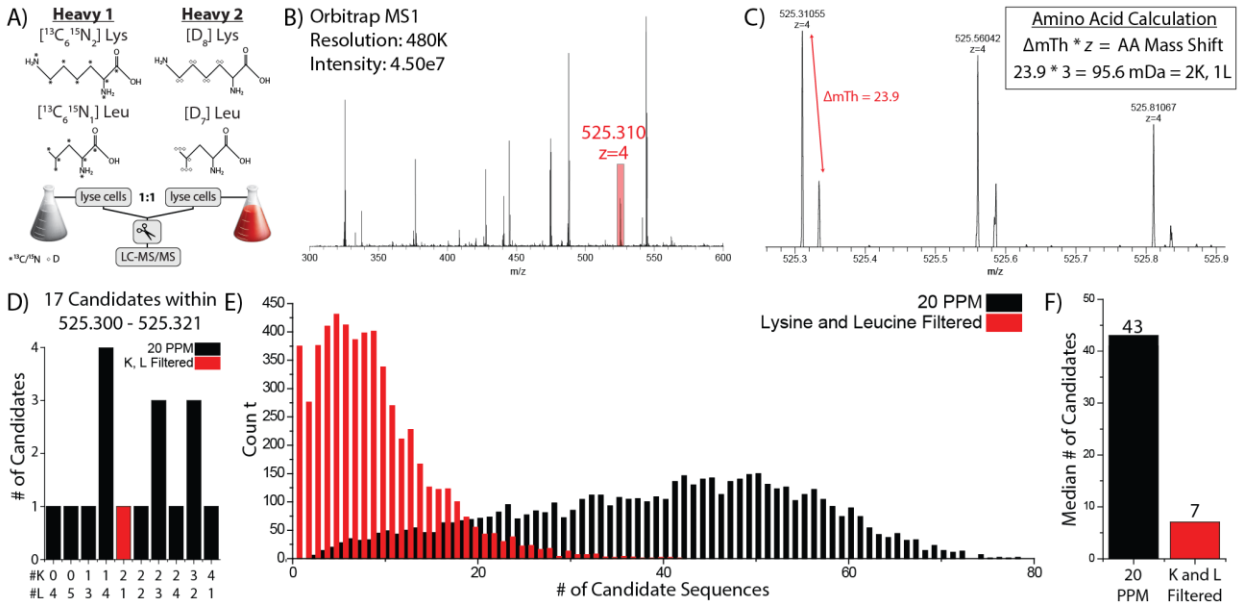


Figure 6. Lysine and Leucine NeuCode SILAC Enabled Amino Acid Counting.
A) Experimental design of lysine and leucine NeuCode SILAC. B) A survey scan acquired in the Orbitrap at a resolving power of 30,000 was used to select a +4 precursor at *m/z* 525.310. C) Enlarged view of precursor at *m/z* 525.310. The presence of two distinct peaks separated by 23.9 mTh is created by the incorporation of 2 lysines and 1 leucine residues. D) A search of an *in silico* digest of the yeast proteome with a 20 ppm precursor tolerance returned 17 candidate peptides, only one of which contained two lysines and one leucine, AKAQGVAVQLKRQPAQPRE. E) The number of candidate sequences remaining after accurate mass (20 ppm) and subsequent lysine and leucine filtering of 5,571 precursors. F) Median number of candidate sequences considered after filtering.

Application of Lysine and Leucine Counting

AACounter was then used to search all precursors chosen for MS/MS analysis using lysine and leucine counting as a search filter. Note we restricted analysis to these precursor because MS/MS scanning provided a second means by which to establish sequence. For this analysis 400 unique precursors from a nanoLC-MS/MS experiment were analyzed with varying precursor search tolerances (20 ppm to 5 ppm) within AACounter (**Figure 7A**). The narrowing of the precursor search window lowers the median number of candidate sequences considered from 42 to 11, and 7 to 2 for accurate mass and accurate mass plus amino acid filtering, respectively. The reduction in search space only slightly impacts sensitivity from 20 to 7.5 ppm (94% vs. 91%), while reduction to 5 ppm severely lowers sensitivity (79%). In addition, decreasing the precursor search tolerance increases the number of precursors with only one candidate sequence from 23 (20 ppm) to 103 (5 ppm) (**Figure 7B**). We are encouraged that peptide sequences identified by our MS^1 analysis method are correct more than 92% of the time across all search tolerances, demonstrating high confidence when these precursors have one suitable match.

A search tolerance of 7.5 ppm was used to analyze all precursors in the same nanoLC-MS/MS experiment, as its high sensitivity (91%) and single precursor match fidelity (97%) offered the best opportunity to identify peptides from an MS^1 alone (**Figure 7C)**. The median number of peptides under consideration drops seven-fold (14 to 2) when amino acid filtering is applied, while the 916 precursors which matched to one peptide (representing 19.5% of all selected precursors) is 14% higher than the number of peptide spectral matches returned through a database search (805). Of the 916 precursors that AACounter matched to one peptide, 86.5% were in agreement with traditional database searching. This provides confidence that precursors not identified by database search, but found to have one match by AACounter are correct. Knowing that many tandem mass spectra are chimeric,[55] we postulate that a considerable fraction of the 14% that disagree may stem from such mixed spectra. One of these precursors, a triply charged peptide at *m/z* 620.023 was uniquely mapped to the sequence GERAKTKDNNLLGKE by AACounter (**Figure 7D**). This spectrum was not mapped to sequence by database search, but the peptide was manually confirmed by calculating an MS^1 mass difference of 156.9 mDa, corresponding to the

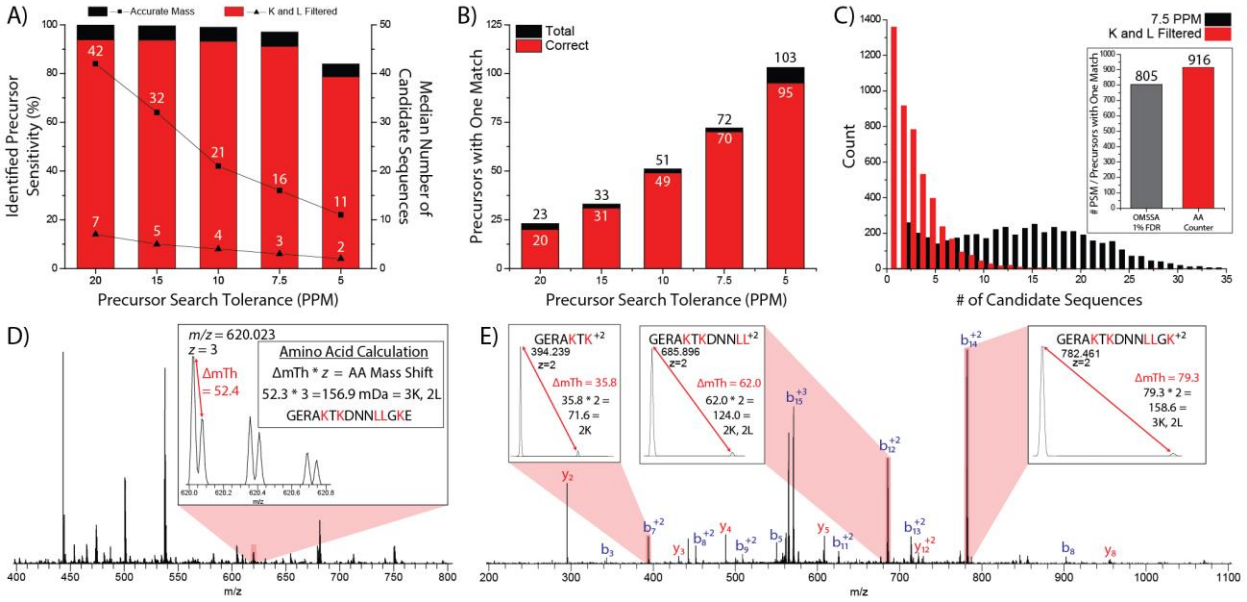


Figure 7. Applying Lysine and Leucine Counting to a nanoLC-MS/MS experiment.
A) Sensitivity and median number of candidate sequences considered vs. precursor search tolerance. B) Number of candidate sequences with one match vs. precursor search tolerance. The number of these precursors that identified the same peptide as a database search is plotted in red. C) Lysine and leucine counting applied to all precursors selected in a nanoLC-MS/MS experiment. AAC returned one sequence candidate for 916 precursors, while returning zero sequence candidates for 1,358 precursors. D) High resolution MS^1 spectrum of a one match precursor at *m/z* 620.023. Two peaks spaced 52.4 mTh, or 156.9 mDa apart signified the presence of three lysines and 2 leucine residues, mapping this precursor to the peptide GERAKTKDNNLLGKE. E) An annotated high resolution MS^2 belonging to the precursor at *m/z* 620.023 confirms the peptide GERAKTKDNNLLGKE. Product ions produced by HCD containing lysine or leucine demonstrate the appropriate partners in the MS^2 spectrum (insets).

presence of 3 lysine and 2 leucine residues (**Figure 7D, inset**) as well as confirming the number of lysines and leucines in each product ion (**Figure 7E**). The simple introduction of two amino acids using NeuCode SILAC, combined with GluC digestion, enabled the identification of peptides from MS^1 spectra alone and provides a proof-of-concept for a NeuCode MS^1 identification method.

Discussion

Here we describe an experimental strategy utilizing NeuCode SILAC peptide isotopologues and high mass accuracy to count the number of lysine and/or leucine residues within precursor species. These data can assist with traditional database correlation sequencing method or can, in some cases, permit identification directly from the MS^1 spectral information. Theoretical calculations demonstrate this strategy could potentially identify more than 50% of the peptides generated by a GluC digestion of the yeast proteome without MS/MS analysis. To implement this strategy we developed a software program, AACounter, to determine possible combinations of lysine or lysine and leucine residues. Counting lysine residues, in addition to high mass accuracy, decreased the median number of candidates for each precursor more than three-fold (44 to 14). The additional filter of requiring lysine containing product ions decreased the median number of peptides considered per precursor from 13 to 7. The addition of leucine to lysine and accurate mass filtering decreased the median number of candidates more than six-fold from 43 to 7. Finally, when used on all selected precursors in a "shotgun" proteomics analysis of GluC yeast digest, accurate mass along with lysine and leucine filtering returns only one peptide candidate for 916 (19.5%) of precursors.

We note that the current implementation used deuterium labeled amino acids for NeuCode SILAC. Deuterium can induce chromatographic elution time shifts,[56] but these are relatively small and inconsequential for most peptides (1-2).[57] That said, samples containing peptides with multiple deuterium labeled amino acids displayed greater elution discrepancies. To account for chromatographic shifts multiple MS^1 spectra were examined to adequately determine the number of amino acids in each precursor.

We envision multiple NeuCode isotopologues, with less reliance on deuterium, of several amino acids being commercially available in the near future.

Accurate selection of the correct monoisotopic precursor *m/z* peak also partly confounded our analysis and reduced our impact from theory. Theoretical calculations were performed assuming knowledge of the precursor monoisotopic mass, which is not always determined correctly by the instrument.[58] Identification of the monoisotopic peak would reduce search space and improve our uniquely identified sequences from the 20% we currently report. Here, these challenges were overcome with post-acquisition analysis, but this technology is ideally implemented in real-time so that up to half of all precursors could be avoided for MS/MS analysis. Recently, we and others described methods to enable real-time searching of precursor species using a modified instrument control methodology.[59,60] Current work is aimed to improve the above limitations and to implement this technology into the real-time acquisition methodologies, improving instrument duty cycle and allowing deeper sampling of complex mixtures.

Experimental Procedures

Yeast Growth

For lysine NeuCode SILAC, *Saccharomyces cerevisiae* strain BY4741 Lys1Δ was grown in defined, synthetic-complete (SC, Sunrise Science) drop out media supplemented with either "*light*" lysine (+0 Da), "*heavy 1*" $^{13}C_6/^{15}N_2$ lysine (+8.0142 Da, Cambridge Isotopes), or "*heavy 2*" $^{2}H_8$ (+8.0502 Da, Cambridge Isotopes). Cells were allowed to propagate for a minimum of 10 doublings to ensure complete lysine incorporation. Upon reaching mid-log phase, the cells were harvested by centrifugation at 3,000 ×g for 3 minutes and washed three times with chilled ddH_2O. Cell pellets were re-suspended in 5mL lysis buffer (50mM Tris pH8, 8M urea, 75mM sodium chloride, 100mM sodium butyrate, 1mM sodium orthovanadate, protease and phosphatase inhibitor tablet), and total protein was extracted by glass bead milling (Retsch).

For lysine and leucine NeuCode SILAC, yeast was grown, harvested, and lysed as described above, but with the following changes: *Saccharomyces cerevisiae* strain BY4742 (MATα; his3Δ1; leu2Δ0; lys2Δ0;

ura3Δ0) was grown in defined, synthetic-complete drop out media supplemented with either "heavy 1" $^{13}C_6/^{15}N_2$ lysine (+8.0142 Da, Cambridge Isotopes; 76 mg/L media) and $^{13}C_6/^{15}N$ leucine (+7.01716 Da, Cambridge Isotopes; 80 mg/L media) or "heavy 2" $^{2}H_8$ lysine (+8.0502 Da, Cambridge Isotopes; 76 mg/L media) and $^{2}H_7$ leucine (+7.04393 Da, Cambridge Isotopes; 80 mg/L media).

Sample Preparation

Protein concentration of yeast lysate was measured by BCA (Pierce). Yeast proteins from respective cultures were mixed in a 1:1 ratio before reduction by 5 mM DTT, alkylation with 14 mM iodoacetamide, and capping by an additional 5 mM DTT. Prior to digestion, the sample was diluted to 2 mM Urea using 50 mM Tris and 3 mM $CaCl_2$. Digestion was carried out by adding GluC (Roche Applied Science, Indianapolis, IN) at a 1:100 enzyme to substrate ratio and incubating overnight at room temperature. Peptides were then acidified with trifluoroacetic acid (TFA) to quench the reaction and de-salted using C-18 solid phase extraction (SPE) columns (Waters, Milford, MA) before analysis.

Nano-High Performance Liquid Chromatography

Online reverse-phase chromatography was performed using a Nano-Acuity UPLC system (Waters, Milford, MA). Peptides were loaded onto a precolumn (75 µm ID, packed with 5 cm Magic C18 particles, Bruker, Michrom) for 10 min at a flow rate of 1 µL/min. Samples were then eluted over an analytical column (75 µm ID, packed with 25 cm Magic C18 particles, Bruker, Michrom) using either a 60 or 120 min linear gradient from 8% to 35% acetonitrile with 0.2% formic acid and a flow rate of 300 nL/min.

Mass Spectrometry

All experiments were performed on an LTQ Orbitrap Elite mass spectrometer (Thermo Fisher Scientific, San Jose, CA). High resolution (30,000 resolving power) survey scans (MS^1) were used to guide data dependent sampling and subsequent CAD fragmentation (NCE=35) of the top five or ten most intense peptides. Fragment ions were then analyzed either in the ion trap or Orbitrap (480,000 resolving power) to

produce MS/MS spectra (MS^2). High resolution (240,000 or 480,000 resolving power) MS^1 scans enabled the distinction of closely spaces isotopologue partners. Preview mode was enabled and precursors with unassigned or +1 charge states were not selected for MS^2 analysis. A maximum of 500 precursors were dynamically excluded for 30 s with a window -0.55 Th and +2.55 Th surrounding the precursor. MS^1 and MS^2 target ion accumulation values were set to 1×10^6, 4×10^4, and 1×10^5 for MS^1, IT-MS^2, and FT-MS^2, respectively.

Database Searching and FDR Estimation

MS^2 data were analyzed using the Coon OMSSA Proteomic Analysis Software Suite (COMPASS).[54] The Open Mass Spectrometry Search Algorithm, OMSSA[5], was used to search spectra against a concatenated target-decoy database consisting yeast protein sequences from Uniprot.[61] GluC peptides were created *in silico* allowing up to three missed cleavages. The precursor tolerance was set to 200 ppm, while the product ion tolerance was set to +/- 0.015 Da and +/- 0.5 Da for FT-MS^2 and IT-MS^2 spectra, respectively. Carbamidomethylation of cysteines were included as a fixed modification, whereas oxidation of methionine was set as a variable modification. Fixed modifications representing the average mass increase of the $^{13}C_6/^{15}N_2$ and $^2H_8$ isotopologues (+8.0322) compared to unmodified lysine and the average mass increase of the $^{13}C_6/^{15}N$ and $^2H_7$ isotopologues (+7.03054) compared to unmodified leucine were used when appropriate. Database search results were then filtered to 1% FDR.[54]

Data Analysis with AminoAcidCounter (AACounter)

Amino acid counting was facilitated by in an in-house program, *AminoAcidCounter* (AACounter). For every precursor, the peak selected for MS^2 in the previous high resolution (>= 240,000 resolving power) spectrum was located and used as the *m/z* for analysis of potential isotopologue partners. The *m/z* shifts of all amino acid combinations (0-5 Lys and 0-5 Leu) were calculated (**Figure 8**). Using the experimental precursor *m/z* as a center point, the tallest peak within a 10 ppm window of each amino acid combination was summed over five MS^1 scans preceding and fifteen MS^1 succeeding the sampling of the

Mass Shift mDa		# of Leucines					
		0	1	2	3	4	5
# of Lysines	0	0.0	26.8	53.5	80.3	107.1	133.9
	1	36.0	62.8	89.6	116.3	143.1	169.9
	2	72.0	98.8	125.6	152.4	179.1	205.9
	3	108.0	134.8	161.6	188.4	215.1	241.9
	4	144.1	170.8	197.6	224.4	251.2	277.9
	5	180.0	206.8	233.6	260.4	287.2	313.9

Figure 8. Mass Shifts of NeuCode Lysine and Leucine Combinations.

precursor and normalized to the summed intensity of the precursor over the same range of scans. Amino acid combinations were ranked based on their absolute distance from a ratio of 1 and only those pairs that received a score of 2 (lysine filtering) or 5 (lysine and leucine filtering) were saved for use in a database search. These score thresholds were determined as they decreased the number of peptides considered, while maintaining high sensitivity.

An *in silico* GluC digest of yeast proteins was created by AACounter (0-3 missed cleavages, carbamidomethylation of cysteines). Peptides within a 20 ppm window of the neutral mass of the selected precursor mass were saved as accurate mass matches. For each precursor scored and searched, AACounter output the spectrum in which the precursor was selected, number of matched peptides based on accurate mass filtering, number of matched peptides based on amino acid filtering of the accurate mass list, peptide sequences remaining after accurate mass filtering, and peptide sequences remaining after amino acid filtering of the accurate mass list.

For MS^2 filtering, AACounter performed an *in silico* fragmentation of each peptide remaining after accurate mass and lysine filtering. For each product that contained a lysine residue, AACounter calculated the *m/z* corresponding to both the "heavy 1" and "heavy 2" isotopologue product peaks and then analyzed the MS^2 spectrum for the presence of these peaks within a 15 ppm window. If both the "heavy 1" and "heavy 2" product isotopologue peaks were present for at least two product ions the peptide remained under consideration as accurate mass, lysine, and MS^2 filtered matches. If a peptide did not contain a lysine residue, no filtering was applied and the peptide remained under consideration.

Acknowledgements

This work was funded by NIH grant R01 GM080148. CMR was funded by an NSF Graduate Research Fellowship and NIH Trainee-ship (T32GM008505). A.E.M. gratefully acknowledges support from a National Institutes of Health-funded Genomic Sciences Training Program (5T32HG002760).

REFERENCES

1. Coon JJ, Syka JEP, Shabanowitz J, & Hunt DF. Tandem mass spectrometry for peptide and protein sequence analysis. *Biotechniques*, 2005, 38(4):519-+.

2. Eng JK, McCormack AL, & Yates Iii JR. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *Journal of the American Society for Mass Spectrometry*, 1994, 5(11):976-989.

3. Perkins DN, Pappin DJC, Creasy DM, & Cottrell JS. Probability-based protein identification by searching sequence databases using mass spectrometry data. *ELECTROPHORESIS*, 1999, 20(18):3551-3567.

4. Craig R & Beavis RC. TANDEM: matching proteins with tandem mass spectra. *Bioinformatics*, 2004, 20(9):1466-1467.

5. Geer LY, Markey SP, Kowalak JA, Wagner L, Xu M, Maynard DM, Yang X, Shi W, & Bryant SH. Open Mass Spectrometry Search Algorithm. *Journal of Proteome Research*, 2004, 3(5):958-964.

6. Cox Jr, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, & Mann M. Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment. *Journal of Proteome Research*, 2011, 10(4):1794-1805.

7. Wenger CD & Coon JJ. A Proteomics Search Algorithm Specifically Designed for High-Resolution Tandem Mass Spectra. *Journal of Proteome Research*, 2013.

8. Sadygov RG, Cociorva D, & Yates JR. Large-scale database searching using tandem mass spectra: Looking up the answer in the back of the book. *Nat Meth*, 2004, 1(3):195-202.

9. Hsieh EJ, Hoopmann MR, MacLean B, & MacCoss MJ. Comparison of Database Search Strategies for High Precursor Mass Accuracy MS/MS Data. *Journal of Proteome Research*, 2009, 9(2):1138-1143.

10. Eng JK, Searle BC, Clauser KR, & Tabb DL. A Face in the Crowd: Recognizing Peptides Through Database Search. *Molecular & Cellular Proteomics*, 2011, 10(11).

11. Olsen JV, Schwartz JC, Griep-Raming J, Nielsen ML, Damoc E, Denisov E, Lange O, Remes P, Taylor D, Splendore M, Wouters ER, Senko M, Makarov A, Mann M, & Horning S. A Dual Pressure Linear Ion Trap Orbitrap Instrument with Very High Sequencing Speed. *Molecular & Cellular Proteomics*, 2009, 8(12):2759-2769.

12. Michalski A, Damoc E, Hauschild J-P, Lange O, Wieghaus A, Makarov A, Nagaraj N, Cox J, Mann M, & Horning S. Mass Spectrometry-based Proteomics Using Q Exactive, a High-performance Benchtop Quadrupole Orbitrap Mass Spectrometer. *Molecular & Cellular Proteomics*, 2011, 10(9).

13. Bereman MS, Canterbury JD, Egertson JD, Horner J, Remes PM, Schwartz J, Zabrouskov V, & MacCoss MJ. Evaluation of Front-End Higher Energy Collision-Induced Dissociation on a Benchtop Dual-Pressure Linear Ion Trap Mass Spectrometer for Shotgun Proteomics. *Analytical Chemistry*, 2011, 84(3):1533-1539.

14. Michalski A, Damoc E, Lange O, Denisov E, Nolting D, Müller M, Viner R, Schwartz J, Remes P, Belford M, Dunyach J-J, Cox J, Horning S, Mann M, & Makarov A. Ultra High Resolution Linear

Ion Trap Orbitrap Mass Spectrometer (Orbitrap Elite) Facilitates Top Down LC MS/MS and Versatile Peptide Fragmentation Modes. *Molecular & Cellular Proteomics*, 2012, 11(3).

15. Jensen ON, Podtelejnikov A, & Mann M. Delayed Extraction Improves Specificity in Database Searches by Matrix-assisted Laser Desorption/Ionization Peptide Maps. *Rapid Communications in Mass Spectrometry*, 1996, 10(11):1371-1378.

16. Clauser KR, Baker P, & Burlingame AL. Role of Accurate Mass Measurement (±10 ppm) in Protein Identification Strategies Employing MS or MS/MS and Database Searching. *Analytical Chemistry*, 1999, 71(14):2871-2882.

17. Goodlett DR, Bruce JE, Anderson GA, Rist B, Pasa-Tolic L, Fiehn O, Smith RD, & Aebersold R. Protein Identification with a Single Accurate Mass of a Cysteine-Containing Peptide and Constrained Database Searching. *Analytical Chemistry*, 2000, 72(6):1112-1118.

18. Smith RD, Anderson GA, Lipton MS, Pasa-Tolic L, Shen YF, Conrads TP, Veenstra TD, & Udseth HR. An accurate mass tag strategy for quantitative and high-throughput proteome measurements. *PROTEOMICS*, 2002, 2(5):513-523.

19. Zubarev R & Mann M. On the Proper Use of Mass Accuracy in Proteomics. *Molecular & Cellular Proteomics*, 2007, 6(3):377-381.

20. Bakalarski C, Haas W, Dephoure N, & Gygi S. The effects of mass accuracy, data acquisition speed, and search algorithm choice on peptide identification rates in phosphoproteomics. *Anal Bioanal Chem*, 2007, 389(5):1409-1419.

21. Cox J & Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotech*, 2008, 26(12):1367-1372.

22. Zubarev RA, Håkansson P, & Sundqvist B. Accuracy Requirements for Peptide Characterization by Monoisotopic Molecular Mass Measurements. *Analytical Chemistry*, 1996, 68(22):4060-4063.

23. Cargile BJ & Stephenson JL. An Alternative to Tandem Mass Spectrometry: Isoelectric Point and Accurate Mass for the Identification of Peptides. *Analytical Chemistry*, 2003, 76(2):267-275.

24. Ong S-E, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, & Mann M. Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. *Molecular & Cellular Proteomics*, 2002, 1(5):376-386.

25. Zhu H, Pan S, Gu S, Bradbury EM, & Chen X. Amino acid residue specific stable isotope labeling for quantitative proteomics. *Rapid Communications in Mass Spectrometry*, 2002, 16(22):2115-2123.

26. Comisarow MB & Marshall AG. The Early Development of Fourier Transform Ion Cyclotron Resonance (FT-ICR) Spectroscopy. *Journal of Mass Spectrometry*, 1996, 31(6):581-585.

27. Marshall AG, Hendrickson CL, & Jackson GS. Fourier transform ion cyclotron resonance mass spectrometry: A primer. *Mass Spectrometry Reviews*, 1998, 17(1):1-35.

28. Bogdanov B & Smith RD. Proteomics by FTICR mass spectrometry: Top down and bottom up. *Mass Spectrometry Reviews*, 2005, 24(2):168-200.

29. Chen X, Smith LM, & Bradbury EM. Site-Specific Mass Tagging with Stable Isotopes in Proteins for Accurate and Efficient Protein Identification. *Analytical Chemistry*, 2000, 72(6):1134-1143.

30. Veenstra TD, Martinović S, Anderson GA, Paša-Tolić L, & Smith RD. Proteome analysis using selective incorporation of isotopically labeled amino acids. *Journal of the American Society for Mass Spectrometry*, 2000, 11(1):78-82.

31. Martinović S, Veenstra TD, Anderson GA, Paša-Tolić L, & Smith RD. Selective incorporation of isotopically labeled amino acids for identification of intact proteins on a proteome-wide level. *Journal of Mass Spectrometry*, 2002, 37(1):99-107.

32. Du Y, Parks BA, Sohn S, Kwast KE, & Kelleher NL. Top-Down Approaches for Measuring Expression Ratios of Intact Yeast Proteins Using Fourier Transform Mass Spectrometry. *Analytical Chemistry*, 2005, 78(3):686-694.

33. Collier TS, Sarkar P, Rao B, & Muddiman DC. Quantitative Top-Down Proteomics of SILAC Labeled Human Embryonic Stem Cells. *Journal of the American Society for Mass Spectrometry*, 2010, 21(6):879-889.

34. Pratt JM, Robertson DHL, Gaskell SJ, Riba-Garcia I, Hubbard SJ, Sidhu K, Oliver SG, Butler P, Hayes A, Petty J, & Beynon RJ. Stable isotope labelling in vivo as an aid to protein identification in peptide mass fingerprinting. *PROTEOMICS*, 2002, 2(2):157-163.

35. Horn DM, Peters EC, Klock H, Meyers A, & Brock A. Improved protein identification using automated high mass measurement accuracy MALDI FT-ICR MS peptide mass fingerprinting. *International Journal of Mass Spectrometry*, 2004, 238(2):189-196.

36. Volchenboum SL, Kristjansdottir K, Wolfgeher D, & Kron SJ. Rapid Validation of Mascot Search Results via Stable Isotope Labeling, Pair Picking, and Deconvolution of Fragmentation Patterns. *Molecular & Cellular Proteomics*, 2009, 8(8):2011-2022.

37. Hunter TC, Yang L, Zhu H, Majidi V, Bradbury EM, & Chen X. Peptide Mass Mapping Constrained with Stable Isotope-Tagged Peptides for Identification of Protein Mixtures. *Analytical Chemistry*, 2001, 73(20):4891-4902.

38. Zhu H, Hunter TC, Pan S, Yau PM, Bradbury EM, & Chen X. Residue-specific Mass Signatures for the Efficient Detection of Protein Modifications by Mass Spectrometry. *Analytical Chemistry*, 2002, 74(7):1687-1694.

39. Gu S, Pan S, Bradbury EM, & Chen X. Precise peptide sequencing and protein quantification in the human proteome through in vivo lysine-specific mass tagging. *J Am Soc Mass Spectrom*, 2003, 14(1):1-7.

40. Pan S, Gu S, Bradbury EM, & Chen X. Single Peptide-Based Protein Identification in Human Proteome through MALDI-TOF MS Coupled with Amino Acids Coded Mass Tagging. *Analytical Chemistry*, 2003, 75(6):1316-1324.

41. Gu S, Pan S, Bradbury EM, & Chen X. Use of Deuterium-Labeled Lysine for Efficient Protein Identification and Peptide de Novo Sequencing. *Analytical Chemistry*, 2002, 74(22):5774-5785.

42. Zhong H, Marcus SL, & Li L. Two-Dimensional Mass Spectra Generated from the Analysis of 15N-Labeled and Unlabeled Peptides for Efficient Protein Identification and de novo Peptide Sequencing. *Journal of Proteome Research*, 2004, 3(6):1155-1163.

43. Shui W, Liu Y, Fan H, Bao H, Liang S, Yang P, & Chen X. Enhancing TOF/TOF-based de Novo Sequencing Capability for High Throughput Protein Identification with Amino Acid-Coded Mass Tagging. *Journal of Proteome Research*, 2004, 4(1):83-90.

44. Hilger M & Mann M. Triple SILAC to Determine Stimulus Specific Interactions in the Wnt Pathway. *Journal of Proteome Research*, 2011, 11(2):982-994.

45. Ong S-E. The expanding field of SILAC. *Anal Bioanal Chem*, 2012, 404(4):967-976.

46. Hebert AS, Merrill AE, Bailey DJ, Still AJ, Westphall MS, Streiter ER, Pagliarini DJ, & Coon JJ. Neutron-encoded mass signatures for multi-plexed proteome quantification. *Nature Methods*, 2012:In Press.

47. Makarov A, Denisov E, & Lange O. Performance Evaluation of a High-field Orbitrap Mass Analyzer. *Journal of the American Society for Mass Spectrometry*, 2009, 20(8):1391-1396.

48. Bruce JE, Anderson GA, Lin C-Y, Gorshkov M, Rockwood AL, & Smith RD. A novel high-performance Fourier transform ion cyclotron resonance cell for improved biopolymer characterization. *Journal of Mass Spectrometry*, 2000, 35(1):85-94.

49. Syka JEP, Marto JA, Bai DL, Horning S, Senko MW, Schwartz JC, Ueberheide B, Garcia B, Busby S, Muratore T, Shabanowitz J, & Hunt DF. Novel Linear Quadrupole Ion Trap/FT Mass Spectrometer: Performance Characterization and Use in the Comparative Analysis of Histone H3 Post-translational Modifications. *Journal of Proteome Research*, 2004, 3(3):621-626.

50. Kaiser N, Quinn J, Blakney G, Hendrickson C, & Marshall A. A Novel 9.4 Tesla FTICR Mass Spectrometer with Improved Sensitivity, Mass Resolution, and Mass Range. *Journal of the American Society for Mass Spectrometry*, 2011, 22(8):1343-1351.

51. Denisov E, Damoc E, Lange O, & Makarov A. Orbitrap mass spectrometry with resolving powers above 1,000,000. *International Journal of Mass Spectrometry*, 2012, 325–327(0):80-85.

52. Jiang H & English AM. Quantitative Analysis of the Yeast Proteome by Incorporation of Isotopically Labeled Leucine. *Journal of Proteome Research*, 2002, 1(4):345-350.

53. Gruhler A, Olsen JV, Mohammed S, Mortensen P, Færgeman NJ, Mann M, & Jensen ON. Quantitative Phosphoproteomics Applied to the Yeast Pheromone Signaling Pathway. *Molecular & Cellular Proteomics*, 2005, 4(3):310-327.

54. Wenger CD, Phanstiel DH, Lee MV, Bailey DJ, & Coon JJ. COMPASS: A suite of pre- and post-search proteomics software tools for OMSSA. *PROTEOMICS*, 2011, 11(6):1064-1074.

55. Houel S, Abernathy R, Renganathan K, Meyer-Arendt K, Ahn NG, & Old WM. Quantifying the Impact of Chimera MS/MS Spectra on Peptide Identification in Large-Scale Proteomics Studies. *Journal of Proteome Research*, 2010, 9(8):4152-4160.

56. Boutilier JM, Warden H, Doucette AA, & Wentzell PD. Chromatographic behaviour of peptides following dimethylation with H2/D2-formaldehyde: Implications for comparative proteomics. *Journal of Chromatography B*, 2012, 908(0):59-66.

57. Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, & Aebersold R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotech*, 1999, 17(10):994-999.

58. Mayampurath AM, Jaitly N, Purvine SO, Monroe ME, Auberry KJ, Adkins JN, & Smith RD. DeconMSn: a software tool for accurate parent ion monoisotopic mass determination for tandem mass spectra. *Bioinformatics*, 2008, 24(7):1021-1023.

59. Bailey DJ, Rose CM, McAlister GC, Brumbaugh J, Yu P, Wenger CD, Westphall MS, Thomson JA, & Coon JJ. Instant spectral assignment for advanced decision tree-driven mass spectrometry. *Proceedings of the National Academy of Sciences*, 2012, 109(22):8411-8416.

60. Graumann J, Scheltema RA, Zhang Y, Cox J, & Mann M. A Framework for Intelligent Data Acquisition and Real-Time Database Searching for Shotgun Proteomics. *Molecular & Cellular Proteomics*, 2012, 11(3).

61. Consortium TU. Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic Acids Research*, 2012, 40(D1):D71-D75.

Chapter 8

Neutron-encoded mass signatures for quantitative top-down proteomics

CMR designed research, performed experiments, analyzed data, and wrote the paper.

This chapter has been published:

Rhoads T. W.*, **Rose C. M.***, Bailey D. J., Riley N. M., Molden R. C., Nestler A. J., Merrill A. E., Smith L. M., Hebert A. S., Westphall M. S., Pagliarini D. J., Garcia B. A., Coon J. J. "Neutron-encoded Mass Signatures for Quantitative Top-Down Proteomics." *Analytical Chemistry*, **2014**, 86(5):2314-2319.

*co-first author

Abstract

We describe a multiplexed *in vivo* metabolic labeling strategy that utilizes neutron encoded mass shifts (NeuCode) imparted by lysine isotopologues used for the stable isotope labeling of mammals (SILAM). Using two lysine isotopologues whose mass differs by 36 mDa we demonstrate similar rates of incorporation that enable accurate quantification of proteins in the fastest incorporating tissue after only three days of labeling. An analysis of eight additional tissues demonstrated sufficient lysine isotopologue incorporation to perform accurate duplex quantitation after 10 to 30 days of labeling. We applied this strategy to determine the tissue-specific protein changes following the deletion of BAP1, a deubiquitinase whose loss is associated with myelodysplastic syndrome and cancer. Using four lysine isotopologues that differ in mass by 12 mDa we constructed 4-plex and 2-plex experiments that analyzed nine tissues and yielded quantitative data for 8,154 proteins, revealing the diverse role of BAP1 tumor suppressor and its effects on metabolic stress response pathways.

Introduction

Top down proteomics methodologies continue to evolve and can provide a useful alternative to the more commonly used bottom up strategies for protein characterization by mass spectrometry.[1] Top down experiments offer a number of advantages, including the ability to characterize the entire primary sequence of a given protein and identify combinatorial patterns of post-translational modifications.[2,3] The analysis of intact protein samples, however, is inherently more complex, complicating experimental outcomes such as quantitation.[4,5] Even though many of the same methods used for quantifying peptides have been used in top down workflows, there are unique challenges to quantifying intact proteins.

Label-free methods are the most accessible form of quantitation for top down experiments. Here, quantitative information is acquired through the use of spectral counts or extracted ion chromatograms.[6] Label-free experiments, however, require many technical and biological replicates to account for run-to-run variability.[7] Label-free approaches also lack the ability to multiplex several samples in a single run, a feature of some label-based quantitation methods that can considerably reduce instrument run-time requirements.[8,9]

Chemical labeling techniques such as acrylamide labeling and tandem mass tags have been used, but the complexity of intact protein samples often leads to incomplete labeling and side reactions that complicate the data analysis.[10,11] Stable Isotope Labeling of Amino acids in Cell culture (SILAC), a metabolic labeling technique, is considered the gold standard for quantitative proteomics, and has been used with affinity-purified proteins with some success.[9,12,13] A drawback of this method, however, is that even a small amount of incomplete labeling results in broad isotope distributions that challenge quantitation.[14] Further, the spectral complexity from multiple isotopic distributions for each protein hinders the ability to multiplex.

We have recently introduced a new strategy for protein quantification: neutron encoding (NeuCode).[15] Though initially applied to shotgun approaches, NeuCode has the potential to address many of the difficulties the aforementioned quantitative strategies suffer from when used for top down experiments. NeuCode is similar in structure to a traditional SILAC experiment, except that the distance

between forms of the protein is greatly compressed such that they are indistinguishable during a medium-resolution scan. Only upon using a very high resolution (>120k) scan are the separate peaks revealed. NeuCode provides quantitative accuracy similar to SILAC, but permits considerably higher multiplexing because the quantitative channels do not add to spectral complexity. In addition, comparisons between heavy labels, rather than between light and heavy as in SILAC, alleviates the need to correct for isotopic broadening and minimizes concerns about sub-unity reagent purities of the heavy labels. Here we present a first look at NeuCode quantification of proteins from a top down perspective in yeast.

Results

As we did with peptides, we first explored our ability to resolve NeuCode SILAC-labeled intact proteins using a recently published dataset.[16] A set of 1,206 top down protein identifications (Kelleher) were examined to calculate the theoretical number of resolvable NeuCode proteins at different label spacings. Briefly, for each protein identification the number of lysine residues (N) were counted and the expected *m/z* difference between NeuCode pairs with 12, 18, and 36 mDa label spacing (Δm) at the identified *m/z* value and charge state (z) were computed as follows

$$\Delta\frac{m}{z}_{exp} = (N \times \Delta m)/z$$

Next, assuming two Gaussian peaks of equal intensities, we calculated the theoretical *m/z* spacing needed to resolve them at a full width 10% (*f*) maximum (FWTM) for resolving powers (RP) between 15,000 and 1,000,000 (steps of 1,000, RP defined at 400 *m/z* and scaled for identified *m/z*). For each protein identification and its identified *m/z* value, the theoretical *m/z* spacing needed to resolve a NeuCode pair was calculated by:

$$\Delta\frac{m}{z}_{theo} = \frac{\sqrt{\ln f}}{\sqrt{\ln 2}}\,\frac{m/z}{RP\sqrt{400/m/z}}$$

The percentage of resolvable proteins (i.e., $\Delta\frac{m}{z}_{exp} \geq \Delta\frac{m}{z}_{theo}$) was plotted as a function of resolving power in **Figure 1**. Using the largest label spacing (36 mDa), over 84% of the proteins are theoretically resolvable,

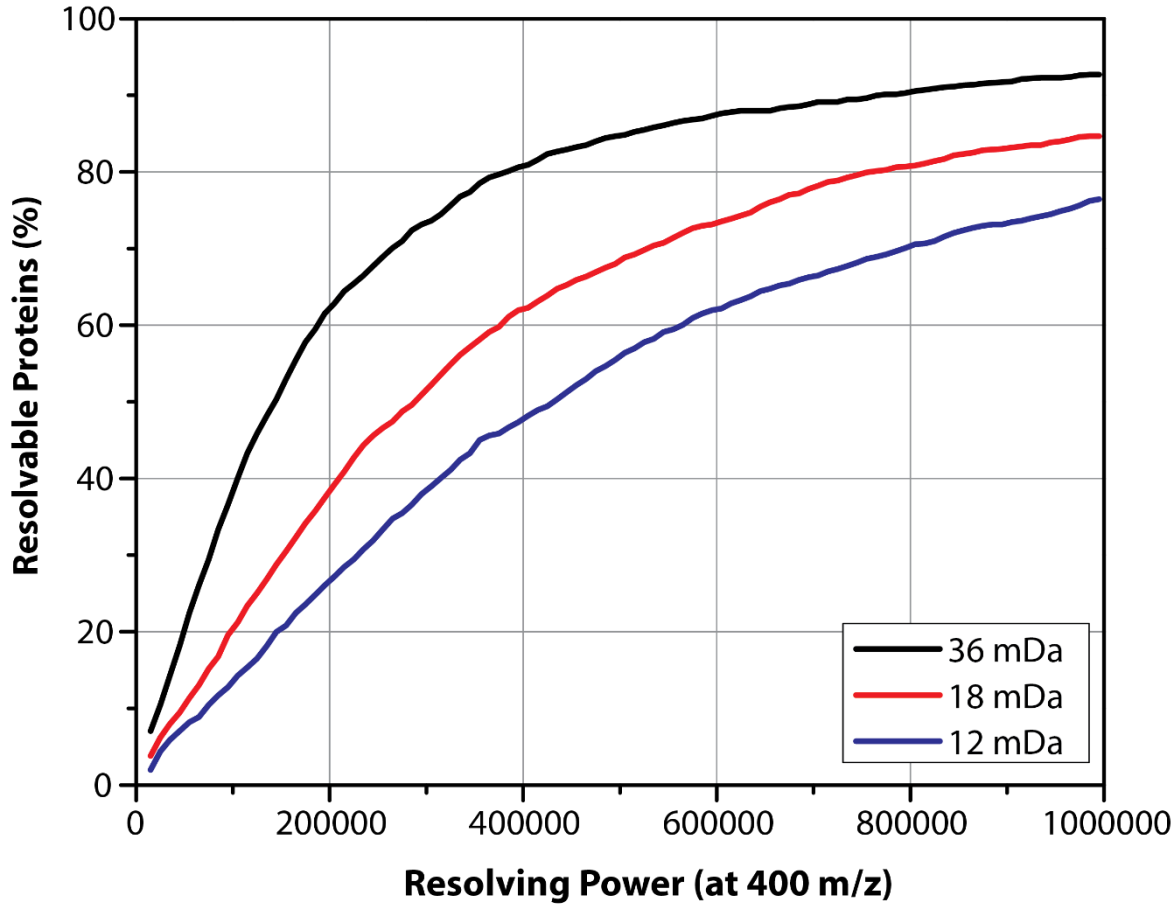


Figure 1. Theoretical resolvability of intact proteins incorporating lysine NeuCode pairs from a top down dataset.

Using a top down data set of 1,206 proteins, we calculated the fraction of resolvable peptides assuming NeuCode spacings of 12, 18, and 36 mDa at the resolution required to resolve the two peaks at full-width 10% maximum, taking into account the decrease in resolution as *m/z* increases.

and thus quantifiable, at 480,000, a level achievable on Orbitrap Elite mass spectrometers. Lysines spaced 36 mDa apart enable duplex quantitation. We also plotted all of the proteins as well as the resolved proteins as a histogram of molecular weight, demonstrating that we can theoretically resolve proteins even over 100 kDa (**Figure 2**). More tightly spaced lysine forms require higher resolving powers. Approximately 77% of the proteins with NeuCode peaks spaced 12 mDa apart in the dataset would be resolvable at a resolution of 960,000. This resolution can be achieved on Orbitraps[17] and FT-ICR instruments. Furthermore, ICR mass spectrometers are a common platform for top down proteomics, and many are capable of resolutions in excess of 1 million, permitting routine use of the 12 mDa spaced lysines and allowing 4-plex quantitative comparisons.[18] These data confirm that NeuCode SILAC quantitation of whole proteins is obtainable for a large percentage of the identified top down proteome.

We next used K_{602} and K_{080} lysine to label yeast from which we prepared top down samples at ratios of 1:1, 3:1, and 5:1 (K_{602}:K_{080}). **Figure 3** illustrates a protein that was identified from an unlabeled yeast sample and then matched to the NeuCode sample by the number of lysines and the retention time. The +9 charge state of the 1-47 fragment of ribosomal protein L26A was identified, yielding a protein isotope distribution at an *m/z* value of 570 and consisting of approximately 9 peaks at 30,000 resolution. At this resolution, there was no indication of multiple forms of the protein that would convey quantitative information. However, upon acquisition of a 240,000 resolution scan, each single peak of the isotope distribution was revealed to consist of two peaks corresponding to the two forms of lysine used (**Figure 3A**). The spacing of the isotopologue pairs revealed the number of lysines in the sequence, aiding in the identification of the protein (**Figure 3B**).[19] In this case the distance is 19.8 mTh and the charge state is +9, indicating that there are five lysines. This information, when matched with the retention time to an unlabeled yeast sample, indicates that it is a fragment of ribosomal protein L26A.

As shown in **Figure 3C**, the spacing in Th between the isotopologue peaks in a fragmentation spectrum collected at sufficient resolution aids in annotation by revealing the number of lysines present in the sequence, as well as providing another avenue through which to acquire quantitative information.[20] **Figure 3C** gives several examples of fragments that contain two or three lysines. This information verifies

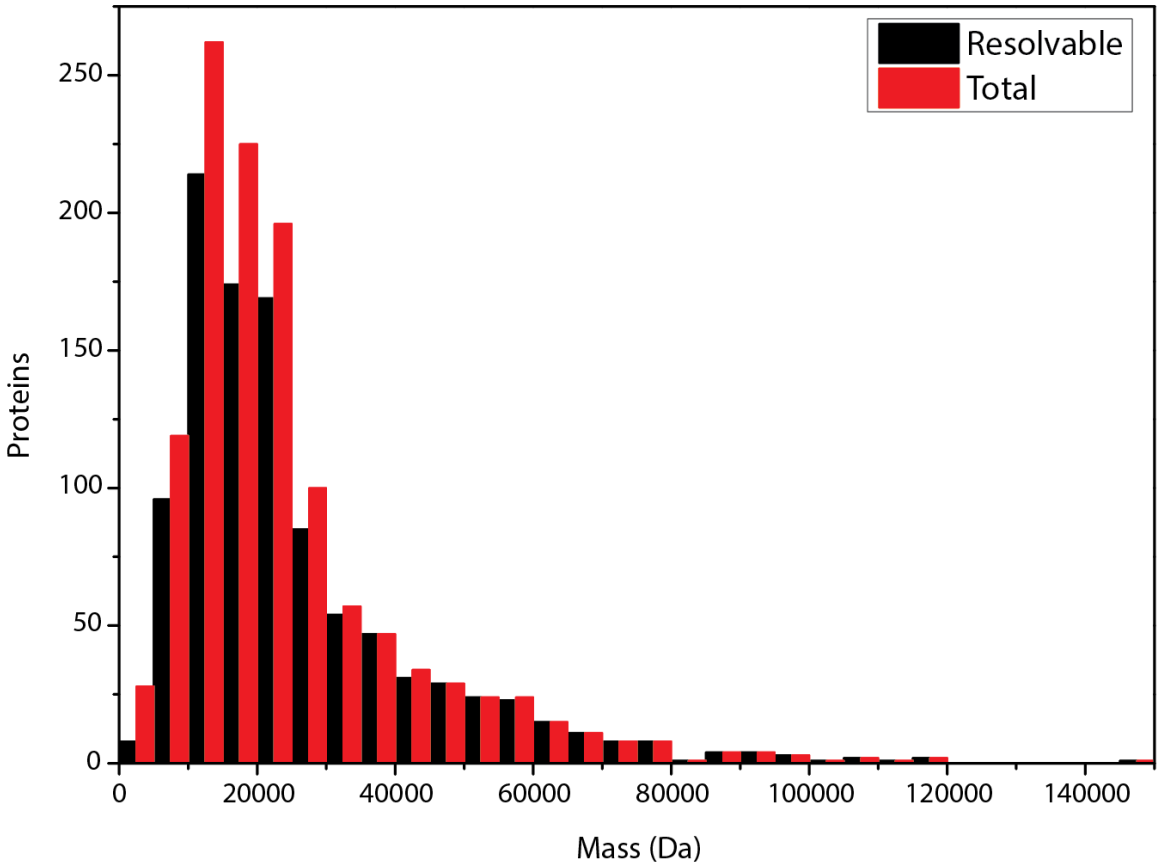


Figure 2. Histogram of the molecular mass of all of the proteins from the Kelleher data set and the proteins that could be resolved.
The resolved proteins were determined by using 36 mDa spaced lysine isotopologues and 480,000 resolution.

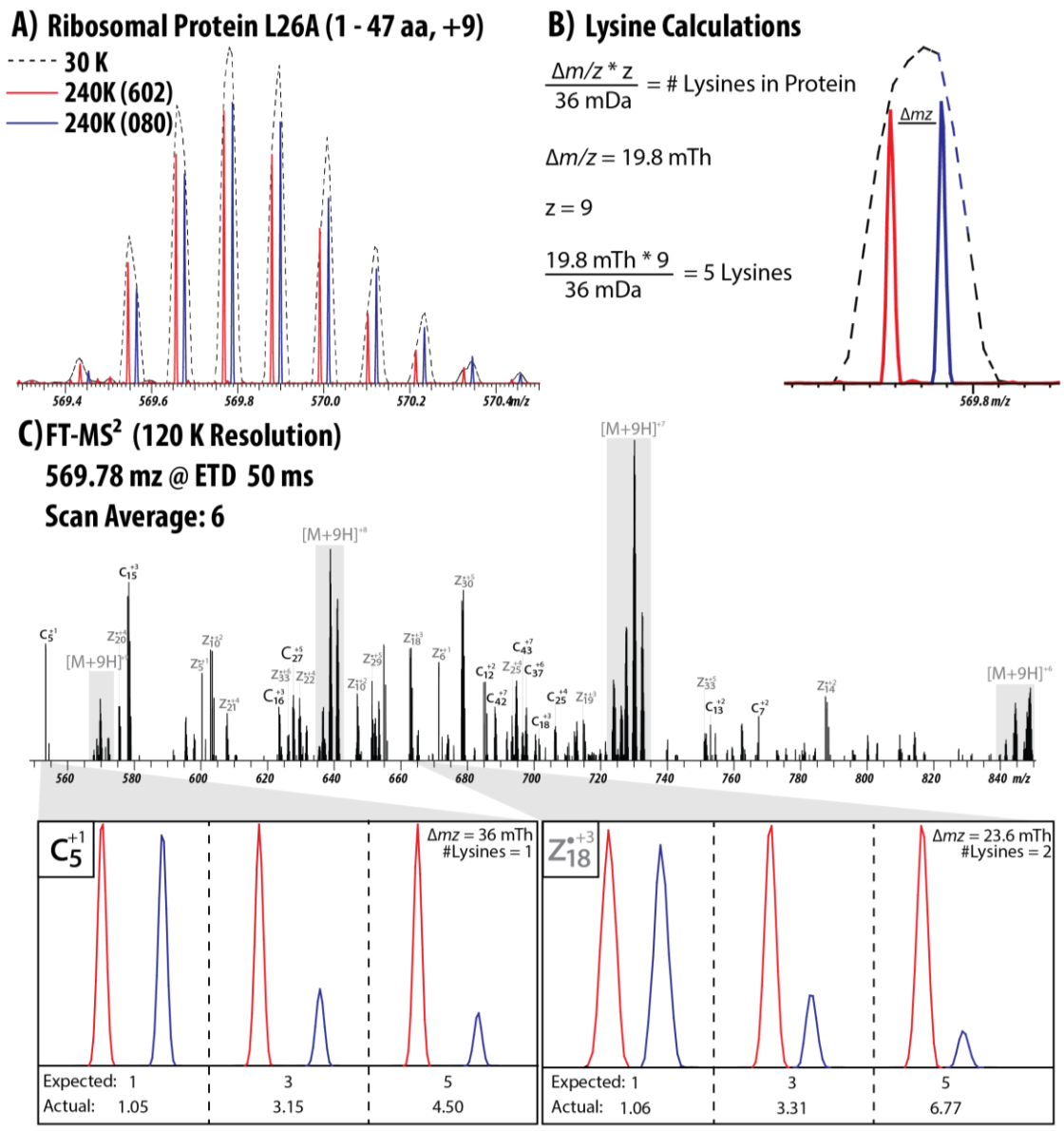


Figure 3. Quantitation of the 1-47 fragment of Ribosomal Protein L26A from yeast using NeuCode.

A) The +9 charge state of the protein (Uniprot B3RHL4) was analyzed first at 30,000 resolution which shows one distinct isotope distribution. A scan at 240,000 resolution, however, reveals the presence of two forms of the protein. B) The spacing between the isotopologue peaks can be used to calculate the number of lysines present in the protein. The protein is carrying 9 charges and the peaks are spaced 19.8 Th apart, indicating 5 lysines. C) Annotated fragmentation spectrum of the L26A precursor. The fragment ions that contain a lysine also show pairs of peaks that can be used to calculate the number of lysines. Two fragments (c5 and z18, ppm errors of 6.40 and 6.56, respectively) were used as examples for quantitation in all three samples, demonstrating excellent quantitative accuracy.

the protein identification from the MS^1 scan, as well as matching the quantitation of the intact protein. We also used this to aid in the identification of Elongin-C from yeast in a previous experiment, as well as annotate an ETD fragmentation spectrum of the +18 charge state precursor of histone H2B (**Figure 4 and 5**).

We observed several hundred intact protein isotope distributions in the defined ratio yeast samples. We currently lack the ability to perform a database search with NeuCode labeled samples, so we are unable to identify all but a few that we can match by hand to an unlabeled yeast sample. However, that does not prevent us from extracting quantitative information from the distributions. Using a list of unique isotopic clusters and the predicted number of lysine residues we calculated the ratio of partners in each known ratio sample. **Figure 6A** summarizes the quantitation of all of the peaks extracted from each sample, demonstrating median ratios of 1.01, 3.02, and 4.58 for the 1:1, 3:1, and 5:1 respectively. Encouragingly, our simple quantitative algorithm achieved excellent accuracy even though we did not have the benefit of the protein sequence. More than 500 species were quantified in the 1:1 and 3:1 samples; however, fewer clusters were quantified in the 5:1 sample where the effect of stringent filters employed during lysine prediction limited the number of detected NeuCode SILAC partners. Future implementations of protein quantification will utilize identified protein sequences removing the requirement to predict the number of lysines and enabling the development of more advanced partner picking algorithms. Using the same identification strategy as above, we were able to identify the +12 charge state of ubiquitin based on its retention time and the number of lysines present. Shown in **Figure 6B**, summing together the isotopologue peaks yields ratios of 0.77, 2.09 and 4.27 for the 1:1, 3:1 and 5:1 samples, in relatively good agreement with the box plots above.

Discussion

Although originally developed for shotgun-based proteomics methods, we demonstrated the feasibility of using NeuCode to enable multiplexed-quantitation for top down experiments. First, we demonstrated that the various lysine isotopologues for NeuCode SILAC incorporate sufficiently (at the

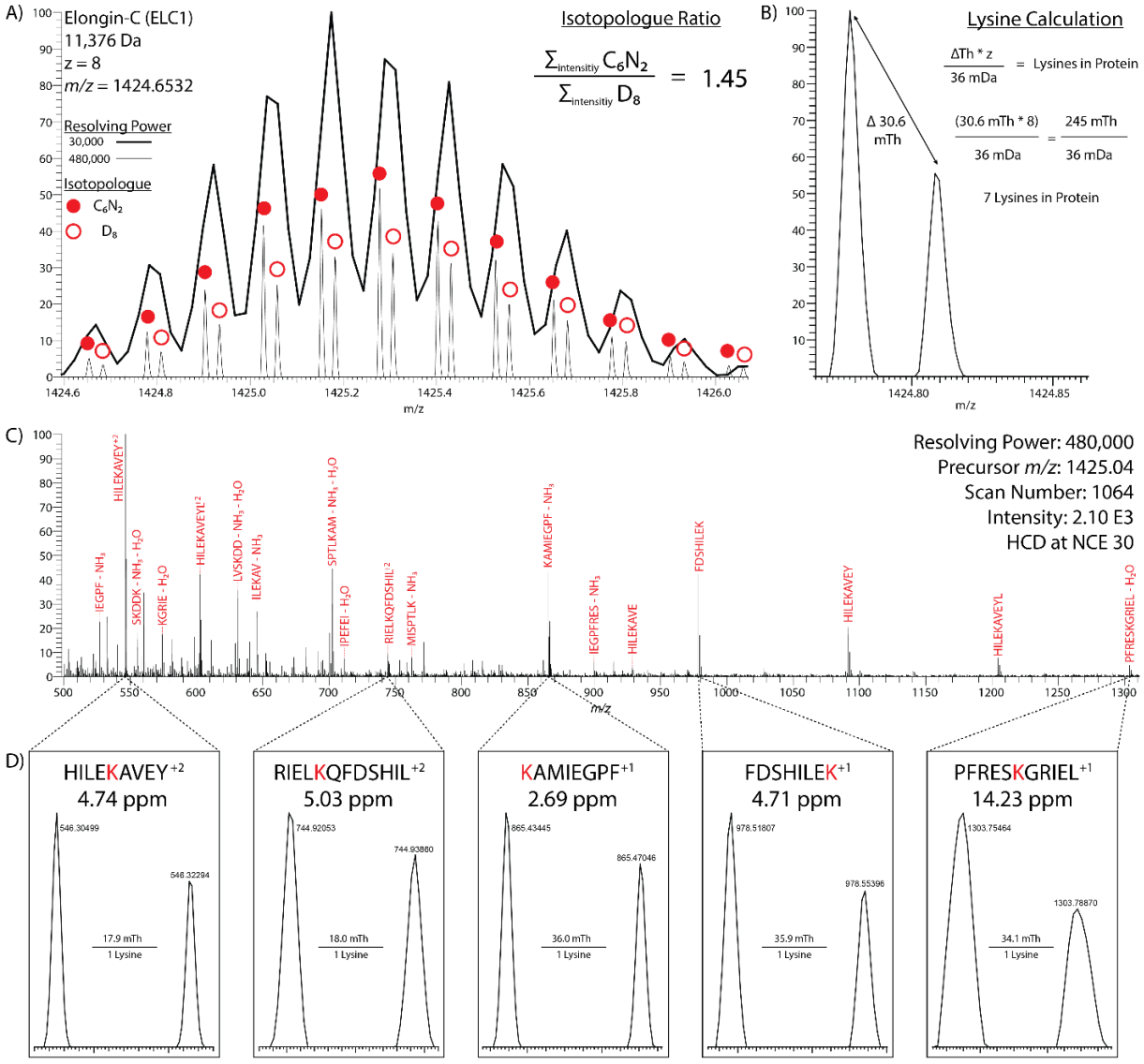


Figure 4. **Identification and Quantitation of Elongin-C from yeast using NeuCode.**
A) The +8 charge state of Elongin-C was isolated and analyzed first at 30,000 resolution which shows one distinct isotope distribution. A scan at 480,000 resolution, however, reveals the presence of two forms of the protein. The $^{13}C_6{}^{15}N_2$ form of lysine is represented by a closed circle, while the $^{2}H_8$ form is represented by an open circle. The sum of the intensities of the two different forms provides the quantitative data, yielding a ratio of 1.45:1. B) The spacing between the isotopologue peaks can be used to calculate the number of lysines present in the protein. The protein is carrying 8 charges and the peaks are spaced 30.6 Th apart, indicating 7 lysines. C) Annotated fragmentation spectrum of the Elongin-C precursor. The sequence of each product ion peak (within 10 ppm) is listed above the peak. D. Selected examples of product ions containing a lysine, showing both isotopologue peaks.

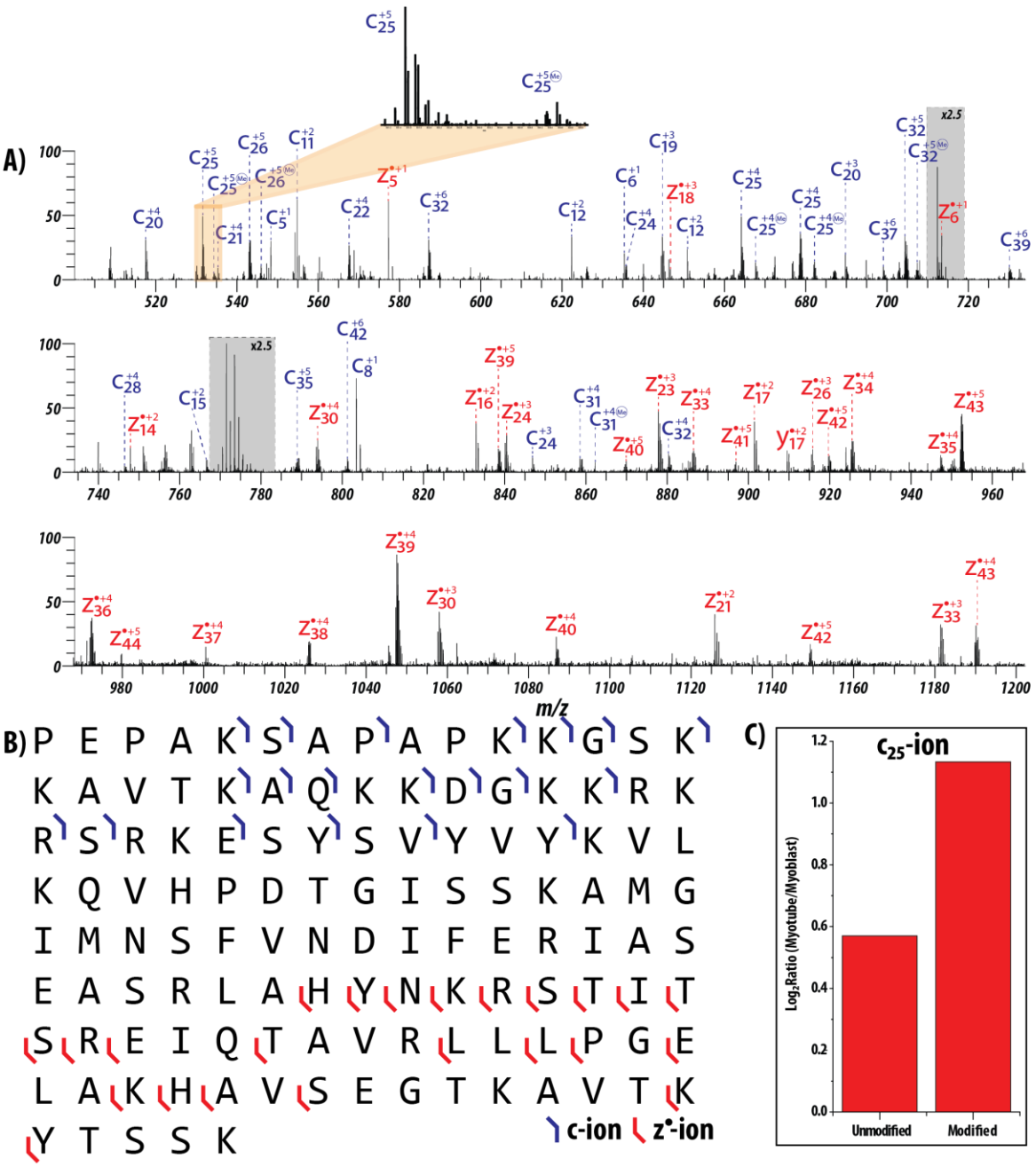


Figure 5. Identification and Quantification of histone H2B from an ETD fragmentation spectrum.

Histone H2B purified from differentiating murine myoblasts was infused into the mass spectrometer and fragmented with ETD. A) The annotation of the fragments from 500-1200 Th. The inset displays the resolved isotopologues of the modified and unmodified forms of the c25 fragment used to derive the quantitation in C. B) Sequence of H2B, with the coverage represented by carrots at each fragment site (blue for c-ions, red for z· ions). C) Integration of the two forms of the c25 ion demonstrate a change due to differentiation.

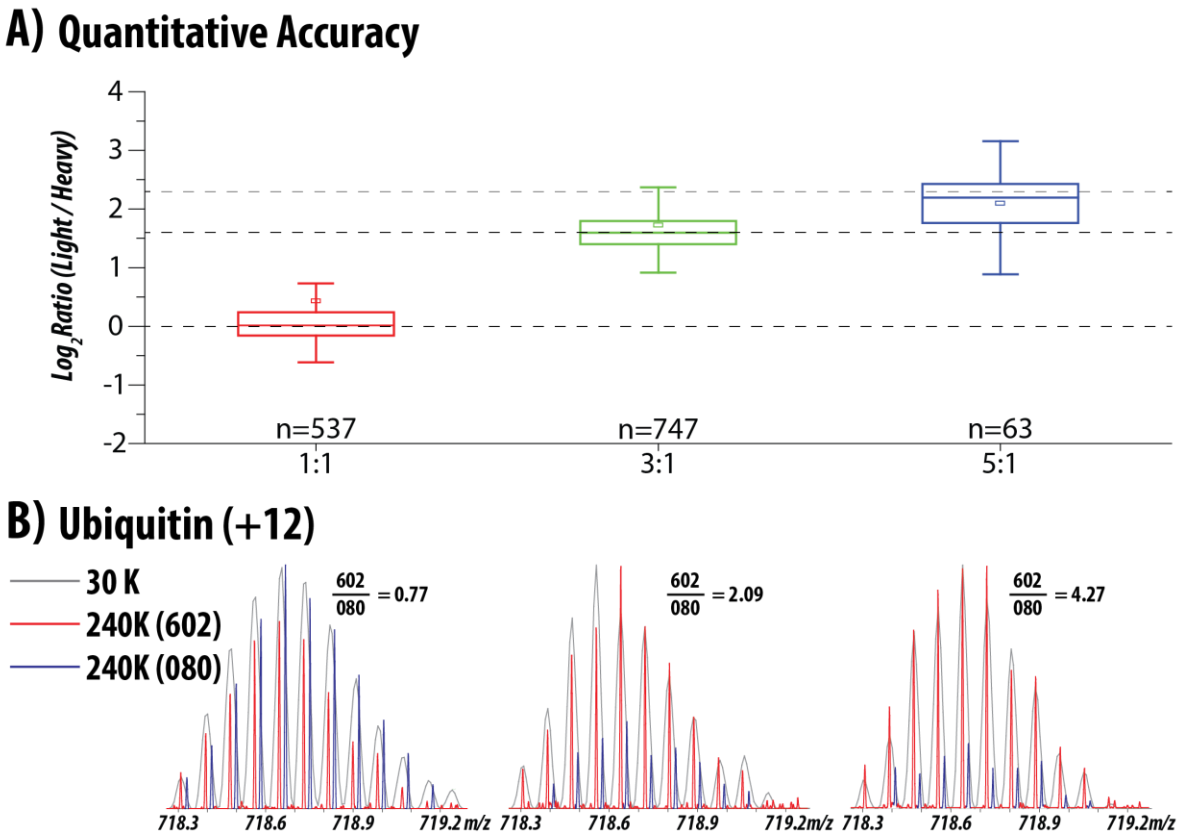


Figure 6. **Quantitation of isotopic distributions from yeast lysate.**
A) Overall quantitation of the distributions (after filtering for a minimum signal/noise of 15) in yeast lysates that were mixed 1:1, 3:1, and 5:1 (K_{602}:K_{080}). Measured (box and whiskers) and true (dotted lines) ratios for all three samples is displayed with the median (stripe), mean (square), interquartile range (25^{th} to 75^{th}, box), and 1.5X interquartile range (whiskers). B) Quantitation of the +12 charge state of ubiquitin (Uniprot P0CG63) from the same mixtures, demonstrating ratios of 0.77, 2.09 and 4.27, in good agreement with the mixing ratios of 1:1, 3:1, and 5:1.

intact protein level) in two disparate cellular systems. Second, we provided evidence that NeuCode SILAC works identically to that observed in our prior shotgun experiments – i.e., the multi-plexed signals are concealed under low to medium resolution scans and only revealed upon analysis under high resolution. We conclude that these benefits will circumvent the problem of MS^1 spectral complication that occurs with traditional SILAC for intact protein analysis. This work provides a basis from which to continue NeuCode development for top-down methodology. A current bottleneck is that top-down spectral searching routines must be modified to accommodate the mass differences and isotopic envelope shifts imparted by NeuCode labels; note this lack of searching capability limited the current study to only examples we could manually identify. With these examples we demonstrate that the embedded NeuCode signals can be fully resolved, even in the context of a protein isotopic distribution. We also note that the high resolution MS requirement of NeuCode SILAC is a condition that is typically met in top down workflows. In sum, NeuCode SILAC has great potential as a robust method for obtaining quantitative data in top down experiments and bypasses many of the drawbacks of current top down quantitative strategies.

Experimental Procedures

Yeast samples

For lysine NeuCode SILAC, *Saccharomyces cerevisiae* strain BY4741 Lys1Δ was grown in defined, synthetic-complete (SC, Sunrise Science) drop out media supplemented with either "K_{602}" $^{13}C_6/^{15}N_2$ lysine (+8.0142 Da, Cambridge Isotopes), or "K_{080}" $^2H_8$ lysine (+8.0502 Da, Cambridge Isotopes). Cells were allowed to propagate for a minimum of 10 doublings to ensure complete lysine incorporation. Upon reaching mid-log phase, cells were harvested by centrifugation at 3,000 × g for 3 minutes and washed three times with chilled ddH_2O. Cell pellets were re-suspended in 5 mL lysis buffer (50 mM Tris pH 8, 8 M urea, 75 mM sodium chloride, 100 mM sodium butyrate, 1 mM sodium orthovanadate, protease and phosphatase inhibitor tablet), and total protein was extracted by glass bead milling (Retsch). Protein concentration of yeast lysate was measured by BCA (Pierce). The 1:1, 3:1 and 5:1 (K_{602}:K_{080}) yeast samples

were mixed in the defined ratios based on the BCA and all yeast lysates were desalted via a tC2 sep-pak (Waters). The samples were dried down and then resuspended in 0.2% formic acid.

Myoblast/myotube samples

C2C12 cells were grown in DMEM dropout culture media (Cambridge Isotopes) supplemented with 602 or 080 lysine. Cells were harvested after six passages, with the exception of the cells grown on 602 lysine, which were placed in differentiation media and allowed to differentiate for an additional five days before harvesting.

The cells were resuspended in Nuclei Isolation Buffer (NIB; 15 mM Tris-HCl pH 7.5, 60 mM KCl, 15 mM NaCl, 5 mM $MgCl_2$, 1 mM $CaCl_2$, 250 mM sucrose) with 0.2% NP-40 at a final ratio of 10:1 buffer:sample, and the NeuCode channels were combined 1:1 based on total protein amounts (BCA, Pierce). Nuclei were isolated and histones extracted and separated as previously described.[21]

Intact histone variants were resuspended in 50% ACN/H2O with 0.2% formic acid and infused via a static nanospray capillary into an LTQ Orbitrap mass spectrometer at 3 µL/min. Spectra were collected and averaged for approximately 2 minutes. Analysis was performed at 480,000 resolution (2 microscans), and target ion accumulation values were set to 2×10^6 and 8×10^5 for MS^1 and MS^2 scans, respectively. For MS^2 analysis, the species at 776 Th was isolated (10 Th window) and fragmented with ETD (10 ms reaction time).

Mass spectrometry and high-performance liquid chromatography

Online reverse-phase chromatography was performed using a Nano-Acuity UPLC system (Waters, Milford, MA). Proteins were eluted over an analytical column (75 µm ID, packed with 30 cm of 5 µm, 300 Å Magic C4 particles, Bruker, Michrom) at 300 nL/min using a 96 min gradient of solvent A (94.8% water, 5% DMSO, 0.2% formic acid) and solvent B (99.8% acetonitrile, 0.2% formic acid): 5% to 8% B from 0 to 1 min, 8% to 22% B from 1 to 45 min, 22% to 30% B from 45 to 60 min, and 30% to 90% B from 60 to 96 min, followed by a 4 min wash at 90% B.

Data were collected on an LTQ Orbitrap Elite mass spectrometer (Thermo Fisher Scientific, San Jose, CA). The nitrogen flow to the Orbitrap chamber was altered such that the increase in pressure (as measured by a Penning ionization gauge) was ~0.15 X 10^{-10} Torr, as compared to the pressure in the absence of nitrogen. Medium resolution survey scans (30,000 resolving power; 4 microscans) were used to guide data dependent sampling of the most intense peaks. Before acquiring the MS^2 scans, a high resolution (240,000 or 480,000, 4 microscans) MS^1 was acquired for quantitation purposes. Precursors were fragmented with ETD (reaction time 50 ms) and detected in the orbitrap at 120,000 resolution with 6 microscans. Target ion accumulation values were set to 3 x 10^6 and 5 x 10^5 for MS^1 and MS^2 scans, respectively. For all scan functions the precursor ions were isolated ±2.5 Th and peaks with assigned charge states of 1-3 were excluded from analysis. Dynamic exclusion was turned off for the duration of the run.

Sample preparation for Elongin-C data

Proteins were precipitated by the addition of 0.4 N H_2SO_4 to 300 µL of 10 mg/mL yeast lysate and incubated at 4 °C on a rotator overnight. Precipitated protein was pelleted by spinning at 16,000 X g at 4 °C for 10 min. The supernatant was removed before 1 part 100% TCA was added to 3 parts supernatant and cooled on ice for 1 hour to precipitate proteins that remained in solution. Proteins of interest were pelleted by spinning at 16,000 X g at 4 °C for 10 min. The supernatant was discarded and the pellet washed with acetone followed by spinning at 16,000 X g at 4 °C for 5 min for a total of three wash cycles. The resulting pellet was air dried for 20 min before the sample was resuspended in 0.2 % formic acid and passed through a 30 kDa molecular weight cutoff filter to remove larger proteins.

Online reverse-phase chromatography was performed using a Nano-Acuity UPLC system (Waters, Milford, MA). Peptides were loaded onto a precolumn (75 µm ID, packed with 5 cm of 3 µm, 300 Å Magic C18 particles, Bruker, Michrom) for 10 min at a flow rate of 1 µL/min. Samples were then eluted over an analytical column (50 µm ID, packed with 15 cm of 3 µm, 300 Å Magic C18 particles, Bruker, Michrom) at 300 nL/min using a 120 min gradient of solvent A (99.8% water, 0.2% formic acid) and solvent B (99.8% acetonitrile, 0.2% formic acid): 8% to 35% B from 0 to 60 min and 35% to 85% B from 60 to 120 min.

Data were collected on an LTQ Orbitrap Elite mass spectrometer (Thermo Fisher Scientific, San Jose, CA). The nitrogen flow to the Orbitrap chamber was altered such that the increase in pressure (as measured by a Penning ionization gauge) was ~0.15 X 10^{-10} Torr, as compared to the pressure in the absence of nitrogen. Medium resolution survey scans (30,000 resolving power; 3 microscans) were used to guide data dependent sampling of the top two most intense peaks. Three scan functions were performed for each precursor: (1) isolation and analysis at 30,000 resolution (4 microscans), (2) isolation and analysis at 480,000 resolution (4 microscans), and (3) HCD activation (NCE = 30) of the precursor followed by FT analysis (480,000 resolution; 4 microscans) of product ions with m/z values > 500. Target ion accumulation values were set to 3 x 10^6 and 1 x 10^5 for MS^1 and MS^2 scans, respectively. For all scan functions the precursor ions were isolated ±2.5 Th and peaks with assigned charge states of 1-3 were excluded from analysis. A maximum of 500 precursors were dynamically excluded for 20 s with a window -1.5 Th and +1.5 Th surrounding the precursor.

Data Analysis

To determine protein identity, the mass shift (36 mDa) induced by the presence of isotopically labeled lysine was used to determine the number of lysines within a precursor. The masses of all yeast proteins were then calculated assuming lysines were isotopically labeled with eight deuterium atoms, and candidate proteins were filtered based on the deviation of the intact protein mass of the isolated precursor ion. Corresponding MS^2 spectra were then manually validated to confirm protein sequence.

Quantification of unidentified isotopic clusters was performed by constructing a list of unique MS^1 features with the corresponding peak apex retention time. This was accomplished by considering each peak in each MS^1 in order of decreasing intensity. Peaks above a signal to noise ratio of 15 and assigned a charge state were saved for future use. To ensure unique clusters within an MS^1, after the peak was saved all peaks +/- 5 Da of the saved peak were excluded. Then, the apex retention time for each saved peak was calculated by assessing the intensity of the peak within a window of +/- 15 MS^1 scans. All

peaks representing unique clusters from each MS^1 were then ranked in order of increasing apex retention time. Starting with the earliest eluting peak, peaks were saved to a final list of unique clusters. To ensure clusters were unique, a +/- 5 Da exclusion window was applied for all precursors with an apex retention time within 30 sec of the peak that was added to the final unique cluster list.

For each unique cluster, the precursor intensity for the K_{602} partner was calculated using five high resolution MS^1 scans proceeding and following the apex retention time. The *m/z* values for potential NeuCode partners were calculated assuming the number of lysine residues in the protein was less than or equal to the charge state. The intensity of the peak with the smallest deviation from the expected mass within a 10 ppm window was summed for each potential partner over the same MS^1 range as the light species. Due to the lack of sequence information, confidence in assigning correct partners was increased by ranking partners by the deviation from the expected ratio (e.g., 1:1, 3:1, or 5:1). This automated algorithm was manually validated to ensure accurate quantitation and partner assignment.

Acknowledgements

We are grateful to Adam Catherman and Neil Kelleher for the details of the top down data set we used for the theoretical calculations. We also thank Rebecka Manis for critical proofreading of the manuscript. This work was supported by the National Institutes of Health grant R01GM080148 to J.J.C, R01DK098672 to D.J.P., an Innovator grant DP2OD007447 to B.A.G., and a training grant 5T32GM007215-37 to A.J.S. The authors also acknowledge support from the National Science Foundation: NSF graduate fellowship DGE-1256259 to A.J.S. and a National Science Foundation Early Faculty Career grant from the office of the director to B.A.G. C.M.R was funded by an NSF Graduate Research Fellowship and NIH Traineeship (T32GM00850

References

1. Siuti N & Kelleher NL. Decoding protein modifications using top-down mass spectrometry. *Nature methods*, 2007, 4(10):817-821.

2. Garcia BA. What does the future hold for Top Down mass spectrometry? *Journal of the American Society for Mass Spectrometry*, 2010, 21(2):193-202.

3. Cui W, Rohrs HW, & Gross ML. Top-down mass spectrometry: recent developments, applications and perspectives. *The Analyst*, 2011, 136(19):3854-3864.

4. Russell JD, Scalf M, Book AJ, Ladror DT, Vierstra RD, Smith LM, & Coon JJ. Characterization and quantification of intact 26S proteasome proteins by real-time measurement of intrinsic fluorescence prior to top-down mass spectrometry. *PLoS One*, 2013, 8(3):e58157.

5. Gordon EF, Mansoori BA, Carroll CF, & Muddiman DC. Hydropathic influences on the quantification of equine heart cytochrome c using relative ion abundance measurements by electrospray ionization fourier transform ion cyclotron resonance mass spectrometry. *Journal of mass spectrometry : JMS*, 1999, 34(10):1055-1062.

6. Asara JM, Christofk HR, Freimark LM, & Cantley LC. A label-free quantification method by MS/MS TIC compared to SILAC and spectral counting in a proteomics screen. *Proteomics*, 2008, 8(5):994-999.

7. Elliott MH, Smith DS, Parker CE, & Borchers C. Current trends in quantitative proteomics. *Journal of mass spectrometry : JMS*, 2009, 44(12):1637-1660.

8. Thompson A, Schafer J, Kuhn K, Kienle S, Schwarz J, Schmidt G, Neumann T, Johnstone R, Mohammed AK, & Hamon C. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem*, 2003, 75(8):1895-1904.

9. Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, & Mann M. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics*, 2002, 1(5):376-386.

10. Du Y, Parks BA, Sohn S, Kwast KE, & Kelleher NL. Top-down approaches for measuring expression ratios of intact yeast proteins using Fourier transform mass spectrometry. *Analytical chemistry*, 2006, 78(3):686-694.

11. Hung CW & Tholey A. Tandem mass tag protein labeling for top-down identification and quantification. *Analytical chemistry*, 2012, 84(1):161-170.

12. Waanders LF, Hanke S, & Mann M. Top-down quantitation and characterization of SILAC-labeled proteins. *Journal of the American Society for Mass Spectrometry*, 2007, 18(11):2058-2064.

13. Collier TS, Sarkar P, Rao B, & Muddiman DC. Quantitative top-down proteomics of SILAC labeled human embryonic stem cells. *Journal of the American Society for Mass Spectrometry*, 2010, 21(6):879-889.

14. Collier TS, Hawkridge AM, Georgianna DR, Payne GA, & Muddiman DC. Top-down identification and quantification of stable isotope labeled proteins from Aspergillus flavus using online nano-flow

reversed-phase liquid chromatography coupled to a LTQ-FTICR mass spectrometer. *Analytical chemistry*, 2008, 80(13):4994-5001.

15. Hebert AS, Merrill AE, Bailey DJ, Still AJ, Westphall MS, Strieter ER, Pagliarini DJ, & Coon JJ. Neutron-encoded mass signatures for multiplexed proteome quantification. *Nature methods*, 2013, 10(4):332-334.

16. Catherman AD, Durbin KR, Ahlf DR, Early BP, Fellers RT, Tran JC, Thomas PM, & Kelleher NL. Large-scale top down proteomics of the human proteome: membrane proteins, mitochondria, and senescence. *Molecular & cellular proteomics : MCP*, 2013.

17. Denisov E, Damoc E, Lange O, & Makarov A. Orbitrap mass spectrometry with resolving powers above 1,000,000. *Int J Mass Spectrom*, 2012, 325:80-85.

18. Bogdanov B & Smith RD. Proteomics by FTICR mass spectrometry: top down and bottom up. *Mass spectrometry reviews*, 2005, 24(2):168-200.

19. Rose CM, Merrill AE, Bailey DJ, Hebert AS, Westphall MS, & Coon JJ. Neutron encoded labeling for peptide identification. *Analytical chemistry*, 2013, 85(10):5129-5137.

20. Richards AL, Vincent CE, Guthals A, Rose CM, Westphall MS, Bandeira N, & Coon JJ. Neutron-encoded signatures enable product ion annotation from tandem mass spectra. *Molecular & cellular proteomics : MCP*, 2013, 12(12):3812-3823.

21. Young NL, DiMaggio PA, Plazas-Mayorca MD, Baliban RC, Floudas CA, & Garcia BA. High throughput characterization of combinatorial histone codes. *Molecular & cellular proteomics : MCP*, 2009, 8(10):2266-2284.

Chapter 9

Multiplexed neutron encoded metabolic labeling of mammals reveals tissue specific effects of BAP1-KO

CMR designed research, performed experiments, analyzed data, and wrote the paper.

This chapter is in preparation for submission:

Rose C. M.*, Baughman J. M.*, Rhoads T. W., Williams C. E., Merrill A. E., Stapleton D., Keller M., Hebert A. S., Westphall M. W., Attie A., Kirkpatrick D. S., Coon J. J., Dey A. "Multiplexed neutron encoded metabolic labeling of mammals reveals tissue specific effect of BAP1-KO." **2014**, *in preparation*.

*co-first author

Abstract

We describe a multiplexed *in vivo* metabolic labeling strategy that utilizes neutron encoded mass shifts (NeuCode) imparted by lysine isotopologues used for the stable isotope labeling of mammals (SILAM). Using two lysine isotopologues whose mass differs by 36 mDa we demonstrate similar rates of incorporation that enable accurate quantification of proteins in the fastest incorporating tissue after only three days of labeling. An analysis of eight additional tissues demonstrated sufficient lysine isotopologue incorporation to perform accurate duplex quantitation after 10 to 30 days of labeling. We applied this strategy to determine the tissue-specific protein changes following the deletion of BAP1, a deubiquitinase whose loss is associated with myelodysplastic syndrome and cancer. Using four lysine isotopologues that differ in mass by 12 mDa we constructed 4-plex and 2-plex experiments that analyzed nine tissues and yielded quantitative data for 8,246 proteins, revealing the diverse role of BAP1 tumor suppressor and its effects on metabolic and stress response pathways.

Introduction

Mammalian model systems are the preferred method to probe the *in vivo* effects of protein knockout and study human disease mimics.[1-3] Large scale analysis of protein and post-translational modifications (PTMs) of these systems has emerged as an important addition to the genomics and transcriptomic analysis typically performed.[4] Recent developments in mass spectrometry (MS) have enabled the comprehensive analysis of the yeast proteome in one hour, clearing a path toward profiling of biologically important systems.[5,6]

Analysis of biological systems by MS has increased due to the development of quantitative MS methods.[7,8] Generally, these approaches fall into two categories: label-free and stable isotope labeling. Label-free methods consist of running multiple, replicate samples in succession and relying on algorithms to compare the peptide abundances across multiple MS experiments.[9-11] Stable isotope labels, introduced metabolically or through peptide labeling, provide multiplexed relative quantitation in one analysis.[8] Isobaric labels (e.g., TMT or iTRAQ) are peptide tags that utilize stable isotopes to enable quantitation in MS/MS spectra (MS^2).[12,13] These methods have grown in popularity due to multiplexing capabilities (up to 10 samples)[14-16] and compatibility with samples derived from either tissue or cell culture.[17,18] Recently, the analysis of precursor interference in MS^2 quantitation strategies has demonstrated these methods suffer from decreased dynamic range.[19] Methods have been developed to purify precursor peptides from interfering species to increase the accuracy of mixed ratio samples, but these methods reduce MS/MS duty cycle leading to a decrease in peptide identifications.[20-22]

Stable isotope labeling by amino acids in cell culture (SILAC) is considered the gold standard for accuracy in quantitative MS measurements.[23,24] Typically, a SILAC experiment consists of two cultures with media containing natural lysine ("light culture") or isotopically labeled +8 Da lysine ("heavy culture"). Proteins from each culture are mixed, digested with LysC (ensuring each peptide will contain a lysine), and analyzed in one MS experiment.[25] This results in two MS^1 features for each peptide, a heavy and light partner, and the MS^1 intensities of these two species is compared to yield relative quantitation. Quantifying from MS^1 spectra results in excellent accuracy, as precursor interference is now a minor issue. MS^1 based

quantification has traditionally suffered from limited multiplexing (≤3) due to the increased MS^1 spectral complexity with each additional sample.[26] Recently, we introduced a method to increase the multiplexing capacity of SILAC by using neutron encoded (NeuCode) lysine isotopologues that differ in mass by 6 to 36 mDa.[27-30] Metabolic introduction of NeuCode lysine reagents (NeuCode SILAC) enabled higher levels of multiplexing as compared to traditional SILAC, while providing more accurate measurements with a larger dynamic range when compared to isobaric labeling methods.[26] A key difference of NeuCode SILAC is that quantitation is performed on multiple heavy peptide partners, as opposed to a light and a heavy partner, decreasing MS^1 spectral complexity increasing the number of proteins that can be identified in one analysis.[26,27]

Since the introduction of SILAC, isotope labeling using amino acids has expanded to stable isotope labeling in mammals (SILAM).[31-35] In these experiments mammals, typically mice, are fed diet containing isotopically heavy lysine to create a heavy reference sample.[35] This tissue is then used in future experiments that quantify experimental samples relative to the reference sample in multiple 2-plex experiments, enabling comparison of experimental samples as a ratio of ratios.[36] The accessibility of SILAM has been limited due to the requirement that the reference material be completely labeled, as unincorporated light peptides will interfere with experimental samples. Full incorporation in mice can take as long as three generations, increasing the time and cost necessary to complete a SILAM experiment.[35] Unlike traditional methods NeuCode SILAC dos not compare light and heavy peptides, but rather compares multiple heavy labeled versions of peptides.[27] We surmised that a SILAM workflow utilizing NeuCode lysine isotopologues would enable the direct comparison of experimental samples by comparing the incorporated heavy peptides following a relatively short labeling period.

Here we describe a multiplexed, quantitative *in vivo* metabolic labeling strategy that reduces labeling time and enables simultaneous analysis of up to four mammals. Analyzing the incorporation at 3, 10, 20, 30 days, demonstrates that two lysine isotopologues incorporate at similar rates without phenotypic effects. Using partially incorporated samples mixed in known ratios we demonstrate that accurate quantitation is possible in fast incorporating tissues after just 3 days of labeling, and possible in all tissues

within the time frame of the study (30 days). Using two more lysine isotopologues, the multiplexing capabilities of NeuCode SILAM were expanded to 4-plex. Using this technology we characterized tissue specific protein changes following deletion of BAP1, a deubiquitinase whose loss is associated with myelodysplastic syndrome and cancer. Combining 4-plex and 2-plex experiments 8,246 proteins were quantified from nine tissues of wild-type and BAP1 KO mice, revealing the diverse role of the BAP1 tumor suppressor and its effects on metabolic and stress response pathways.

Results

Neutron encoded stable isotope labeling by amino acid in mammals (NeuCode SILAM)

In traditional stable isotope labeling by amino acid in mammal (SILAM) experiments, light peptides from experimental samples are compared to heavy-labeled reference material in multiple 2-plex experiments.[35] In that context, incomplete labeling of reference sample yields peptides that interfere with unlabeled experimental peptides. For example, a light peptide from an experimental sample (**Figure 1A**, Light Sample) mixed at equal amounts with a reference peptide that exhibits 50% incorporation (**Figure 1A**, Heavy Reference) results in a ratio of 3:1 (**Figure 1A**, Mixed). Neutron encoded SILAM (NeuCode SILAM) directly compares the heavy versions of partially incorporated peptides for up to four samples. If two experimental peptides that exhibit 50% incorporation (**Figure 1A**, Heavy Sample #1-K_{602} and #2-K_{080}) are mixed at equal amounts the resulting ratio will appear 1:1; however, each peak represents a mixture of both samples (**Figure 1A**, Mixed). When analyzed with high resolution the heavy isotopologue partners are revealed and demonstrate a 1:1 ratio (**Figure 1B**), while the light partners remain one mixed peak.

To demonstrate this experimentally we analyzed peptides that resulted from a LysC digest of liver proteins extracted from mice labeled with either K_{602} (lysine – $^{13}C_6{}^{15}N_2$) or K_{080} (lysine – $^{2}H_8$) for 10 days (**Figure 1C**). An Orbitrap MS^1 spectrum collected at a medium resolution (30,000) demonstrates that each peptide has both a light and heavy partner (**Figure 1D**). When the resolution is increased to 480,000 the light partner remains as a single peak, while the isotopologue pair is revealed in the heavy partner, enabling quantitation of the peptide (**Figure 1E**).

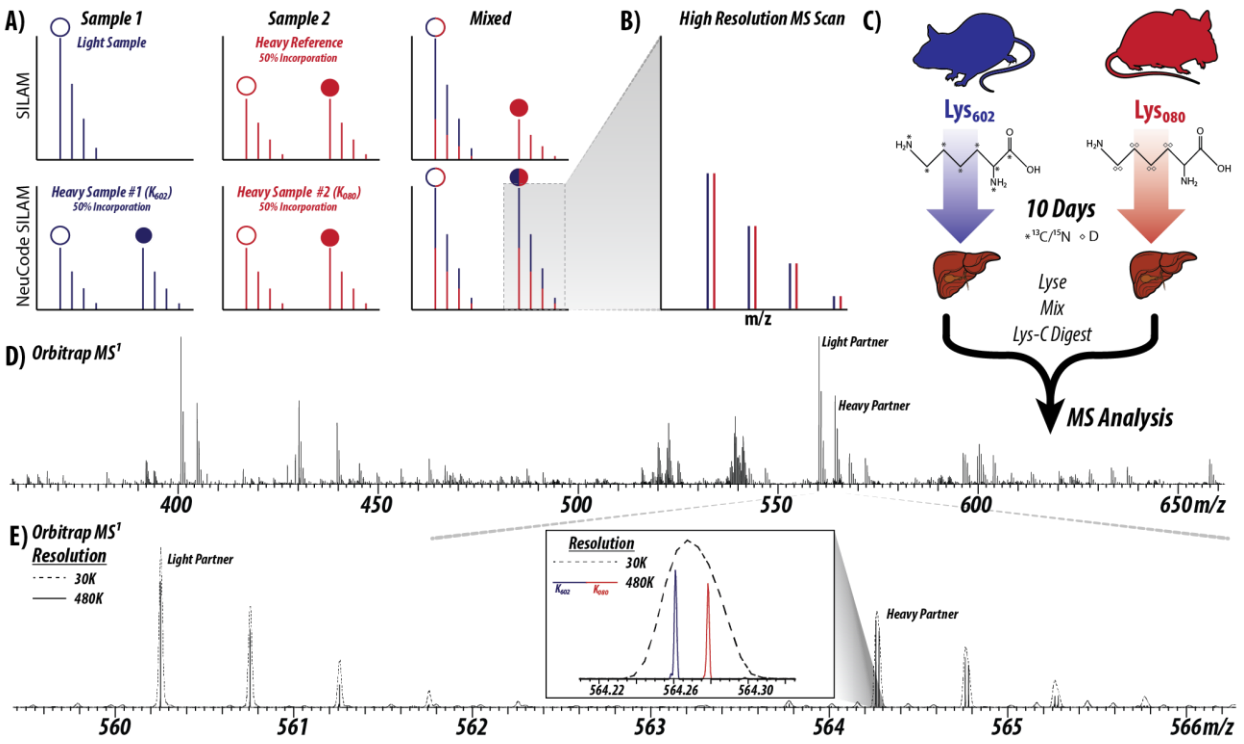


Figure 1. NeuCode Stable Isotope Labeling of Mammals (SILAM)

A) Comparison of SILAM and NeuCode SILAM analysis of partially incorporated samples. B) A higher resolution scan reveals NeuCode SILAM isotopologue partners. C) Sample preparation for NeuCode SILAM experiment. Liver from mice labeled with either K_{602} or K_{080} for 10 days was mixed and digested with LysC before MS analysis. D) Orbitrap MS^1 analysis of partner peptides from both light (unincorporated) and heavy (incorporated) peptides. E) Lysine isotopologue pair within the heavy partner is revealed with a high resolution MS^1 quantitation scan (480,000).

Incorporation rates of lysine isotopologues in various tissues

Reported examples of SILAM have utilized the $^{13}C_6$ version of lysine (K_{600}) within mouse diet to create heavy labeled reference material. NeuCode SILAM differs as the lysine isotopologues may contain a heavy nitrogen or hydrogen (deuterium). To ensure that these isotopologues would exhibit similar incorporation rates we labeled six week old mice with control laboratory diet for two weeks, followed by two weeks of lysine deficient diet supplemented with light lysine (K_{000}) and four weeks of labeling with lysine deficient diet fortified with either K_{602} or K_{080} (**Figure 2**). For this experiment three mice were fed K_{000} diet, twelve mice were fed with K_{602} diet, and twelve mice were fed with K_{080} diet. At 3, 10, 20, and 30 days nine tissues (brain, heart, intestine, islets, kidney, liver, lung, muscle, and plasma) were harvested from three mice in each of the K_{602} and K_{080} cohorts (**Figure 2A**). The three control mice were fed on K_{000} diet for the duration of the experiment. The weight (**Figure 2B**) and food consumption (**Figure 2C**) of the mice was monitored throughout the experiment for mice on both the light (K_{000}) and heavy lysine isotopologue (K_{602} and K_{080}) diets.

The mice fed heavy diet did not exhibit any phenotypic changes, as the weights are within measurement error of the control group (**Figure 2B**). The drop in weight measured at day 13 was the result of an unintended fasting period and not the result of the heavy diet as the control mice experienced the same effect. To determine the amount of lysine isotopologue incorporation, proteins extracted from each tissue, from each mouse, at each time point were digested with LysC and analyzed separately by nLC-MS/MS. The percent incorporation was determined by comparing the intensities of the heavy to the light channel for each sample and the mean percent incorporation for all proteins from each mouse was calculated for each tissue and day (**Figure 2D**). The intestine appears to incorporate lysine into the proteins rapidly, as the average incorporation percentage is greater than 40%, 60%, 70%, and 75% after 3, 10, 20 and 30 days of labeling, respectively. Encouragingly, five (intestine, plasma, liver, kidney, islets), seven (lung and heart), and all (brain, muscle) tissues reach 50% incorporation at 10, 20, and 30 days, respectively, suggesting quantitation is possible with all tissues within the time frame of this study.

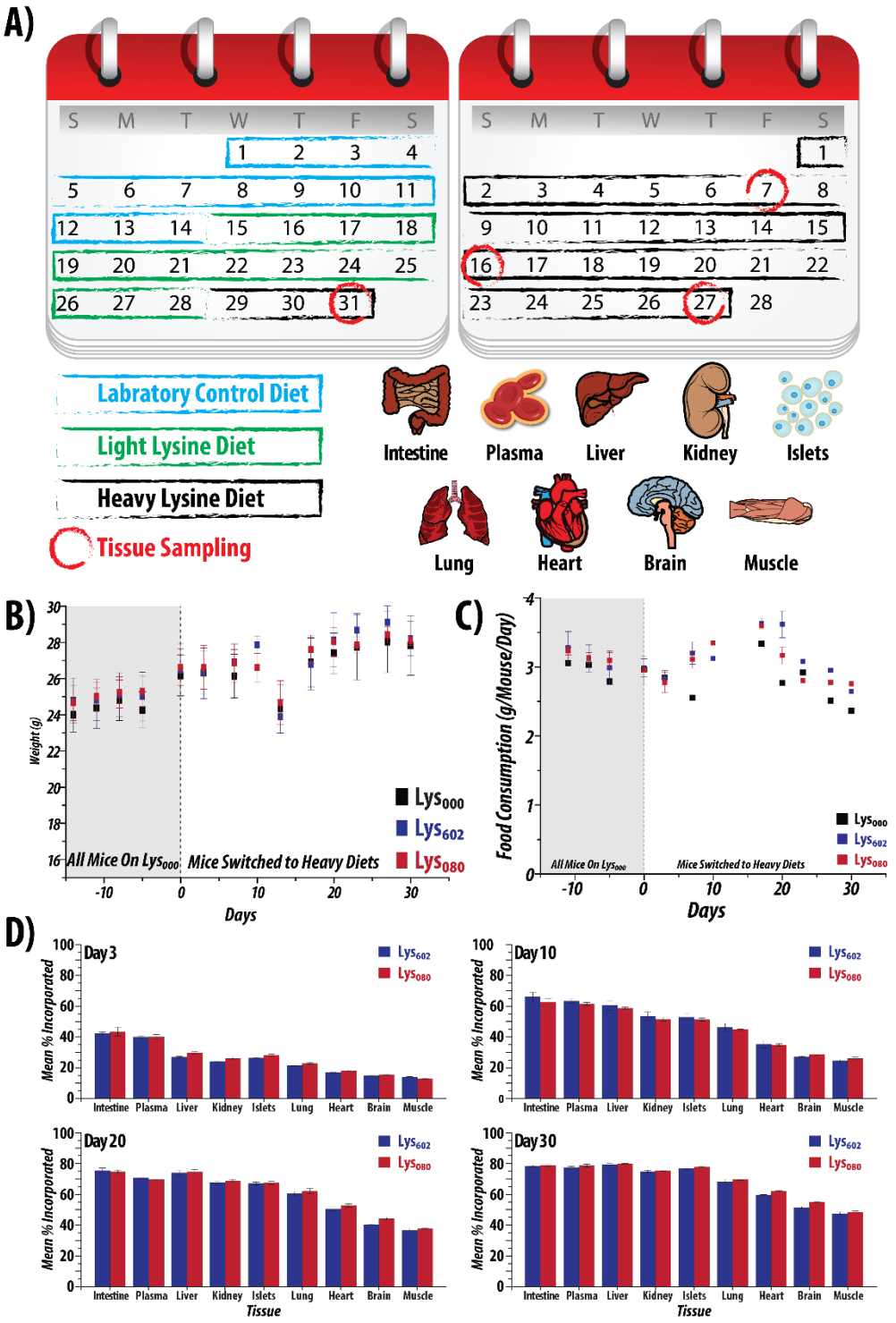


Figure 2. Characterizing incorporation of lysine isotopologues.

A) Experimental design of the 30 day study. B) Mouse weights throughout the course of the experiment. The decrease at day 13 is a result of an unintended fasting period experienced by mice on both control and heavy diets. C) Food consumption of animals housed together. D) Average lysine incorporation for all nine tissues for both heavy diets at each day.

These data also suggest proteins incorporate at the same rate, but to directly compare incorporation on a protein by protein basis we plotted the percent incorporation for all mice versus the percent incorporation for the K_{602} #1 mouse at each time point (**Figure 3**). These analyses demonstrate excellent correlation of protein incorporation for different mice and diets (**Figure 3**). Variations in protein incorporation may arise based on differences in food consumption, K_{602} #3 at day 10 exhibits slightly higher incorporation than K_{602} #1 (**Figure 3**); however a systematic shift such as this can be corrected in the same way total protein input is corrects in isobaric labeling experiments. With the knowledge that proteins exhibit similar incorporation patterns across diets and biological replicates we aimed to determine the percent incorporation (i.e., labeling time) necessary to enable protein quantitation.

Characterizing NeuCode SILAM quantitative performance

Stable isotope labeling by amino acid in cell culture (SILAC), the cell culture analogue of SILAM, has long been the gold standard for quantitative accuracy in mass spectrometry measurements. The primary limitation of SILAC quantitation is precursor ion signal-to-noise, as low signal-to-noise measurements may only detect one SILAC partner.[37] NeuCode SILAM analyzes partially incorporated peptides using only the heavy SILAM partners, whose signal-to-noise is dictated by the level of lysine incorporation. It stands to reason that there is a level of lysine incorporation necessary to achieve accurate protein quantification. To determine this threshold, peptides resulting from the LysC digestion of proteins from three tissues of varying incorporation rates (intestine, live, brain) were mixed in ratios of 1:1, 5:1, and 10:1 (K_{602}:K_{080}) and analyzed in separate nLC-MS/MS experiments (**Figure 4**).

As expected, the quantitative accuracy and percent of peptide quantified improves with the amount of lysine incorporation (**Figure 4 and 5**). After only three days of labeling the mean percent incorporation in the intestine is 40% where the MS measurements exhibit excellent quantitation (**Figure 4A**) and a high quantified peptide percentage (84%, **Figure 5B**). Alternatively, the liver exhibits accurate quantitation at three days for both 1:1 and 5:1 ratios with 65% of peptides quantified, but underestimates a 10:1 ratio (**Figure 4B**). At this time the mean incorporation is 30%, suggesting a larger incorporation percentage is

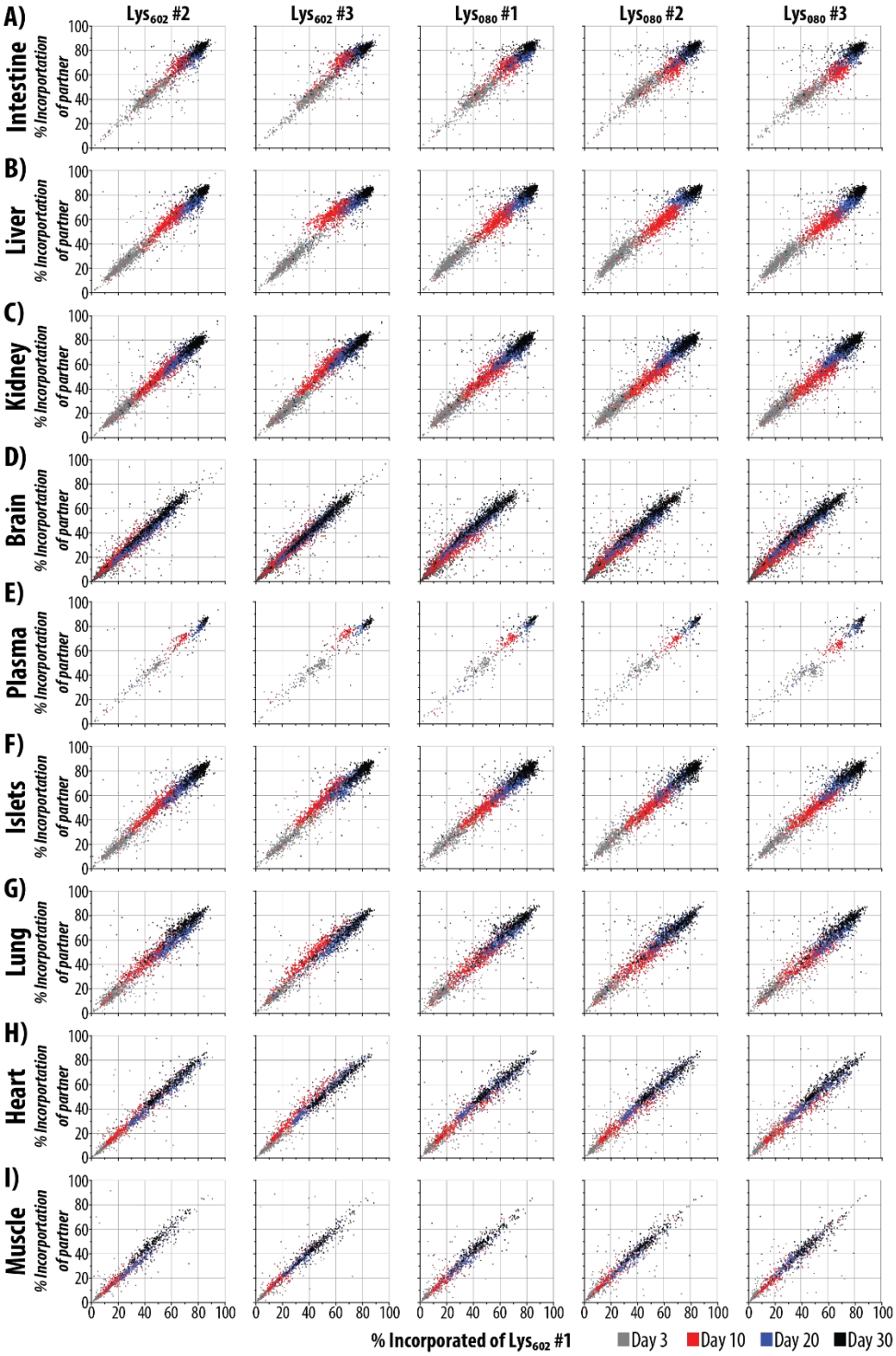


Figure 3. Protein level incorporation of lysine isotopologues.

A-I) Protein level percent incorporation at each time point for each biological replicate fed either K_{602} or K_{080} diet. Samples were compared to biological replicate #1 of K_{602}, regardless of diet.

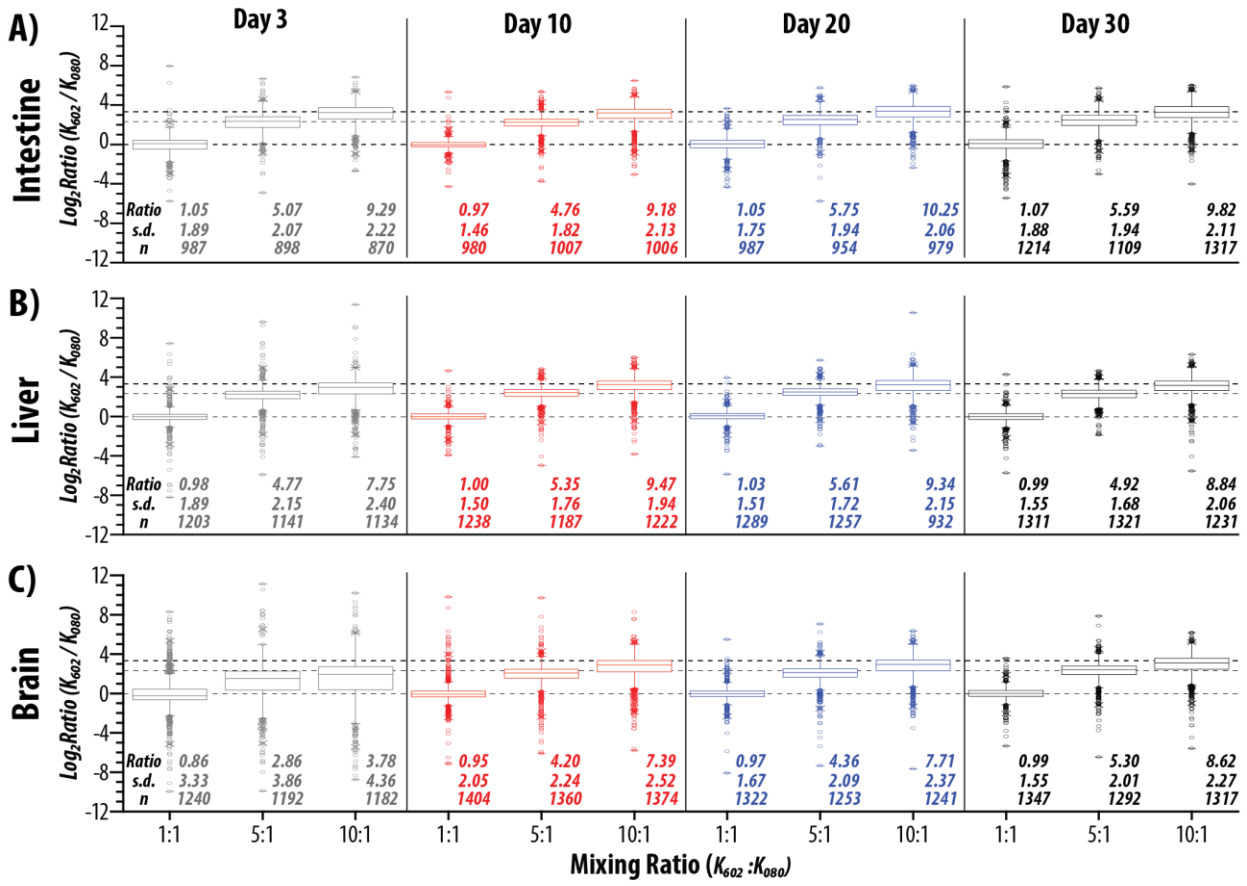


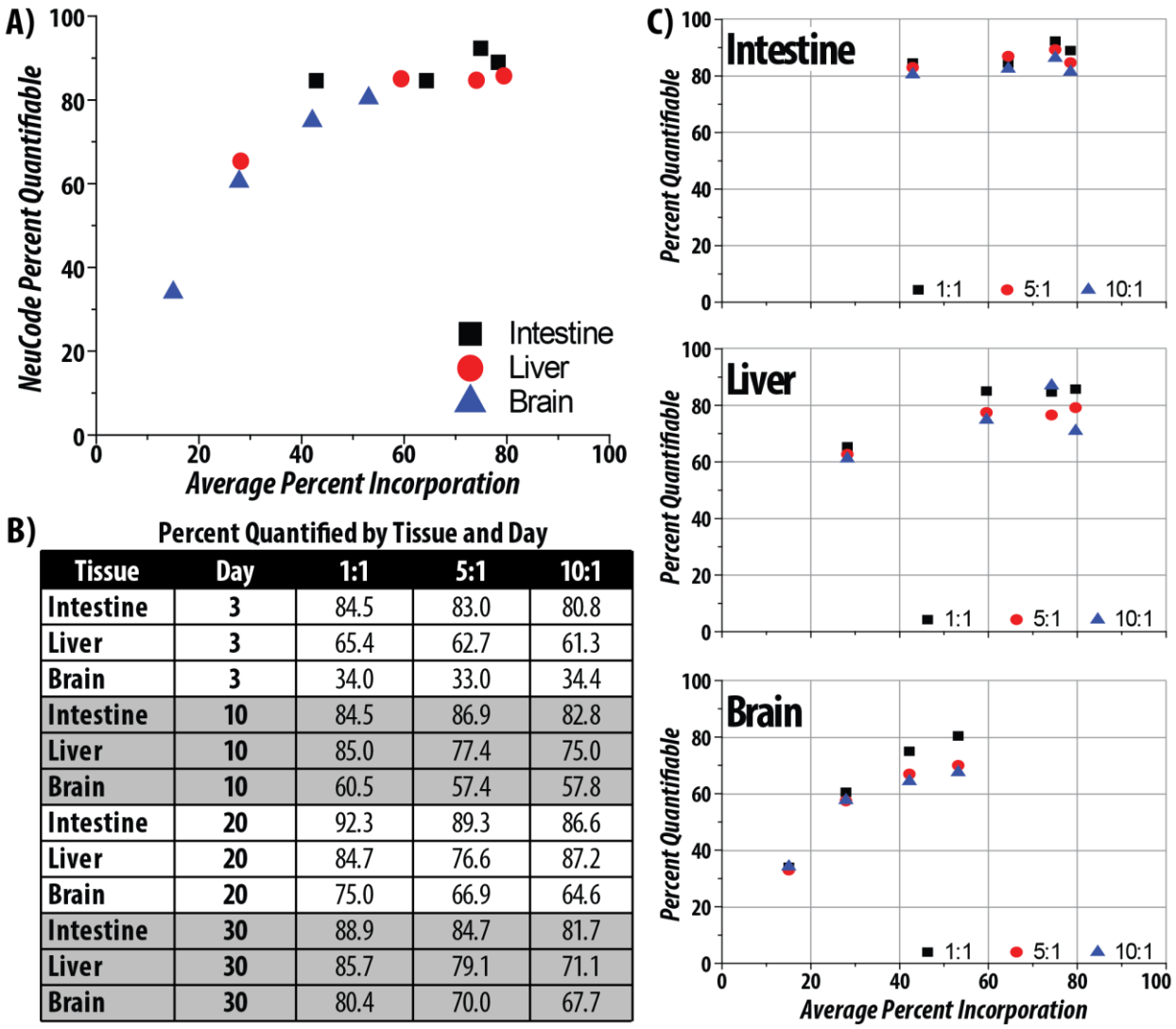
Figure 4. Characterization of NeuCode SILAM quantitation.

A-C) Analysis of mixed ratios for tissues at four time points. Proteins from mice fed either K602 or K080 were mixed in various ratios (1:1, 5:1, and 10:1) for three tissues, intestine (A), liver (B), and brain (C) for all time points. Measured (box and whiskers) and true (dotted lines) ratios are displayed with the median (stripe), interquartile range (25^{th} to 75^{th}, box), 1.5X interquartile range (whiskers), and outliers (circles). The median ratio, standard deviation, and number of measurements is given for each box plot.

necessary for accurate quantitation of higher ratios. This is confirmed at day ten when the mean incorporation is now at 60%, the measured ratios are much closer to the theoretical values, and the percent quantified raises to 85% (**Figure 4B** and **Figure 5**). The brain, one of the slowest incorporating tissues, exhibits poor accuracy and precision after three days of labeling, while quantifying only 34% of peptides (**Figure 4C** and **Figure 5B**). As the mean percent incorporation rises to 40% after 20 days of labeling, the measured ratios and precision of measurements improves to levels that would be acceptable for quantitation, with only slight decrease in dynamic range while still achieving 75% of peptides quantified. To achieve a high level of accuracy labeling slow incorporating tissues would require 30 days of labeling where the percent incorporation is above 50% and the measured ratios are much closer to the desired ratio and the percent of peptides quantified is >80% (**Figure 4C** and **Figure 5**). From these analyses we surmised that a mean percent incorporation of ~40% will enable the quantification of 75% of peptides with acceptable quantification in even the slowest incorporating tissues. This conclusion suggests that optimal labeling times fall between 3 and 20 days, depending on the target tissues of a particular study. With this information we set out to apply this method to an *in vivo* system.

NeuCode SILAM reveals tissue specific effects of deubiquitinase deletion

Tumor suppressor BAP1 is a deubiquitinase whose loss is associated with myelodysplastic syndrome and cancer.[38] Recently, a report describes the mutations of BAP1 have been found in roughly 30% of malignant mesothelioma and uveal melanoma cases as well as 15% of cutaneous melanomas that were studied.[39] BAP1 is involved in many cellular processes including cell cycle progression, DNA damage control, glucose sensing, and transcriptional and epigenetic regulation.[40] Given its importance in disease, a BAP1 deficient mouse model would afford the opportunity to study the effects of BAP1 deletion on the entire animal. Noting that germline mutations are lethal in mice we utilized a BAP1 inducible deletion strain that has been previously described.[41] Deletion of BAP1 within this mouse model is ubiquitous, quickly effects myeloid cells, and makes it difficult to follow the progression of the disease. To circumvent this shortcoming, the bone marrow of BAP1-KO mice is replaced with that of normal wild-type (WT) mice,



B)

Percent Quantified by Tissue and Day

Tissue	Day	1:1	5:1	10:1
Intestine	3	84.5	83.0	80.8
Liver	3	65.4	62.7	61.3
Brain	3	34.0	33.0	34.4
Intestine	10	84.5	86.9	82.8
Liver	10	85.0	77.4	75.0
Brain	10	60.5	57.4	57.8
Intestine	20	92.3	89.3	86.6
Liver	20	84.7	76.6	87.2
Brain	20	75.0	66.9	64.6
Intestine	30	88.9	84.7	81.7
Liver	30	85.7	79.1	71.1
Brain	30	80.4	70.0	67.7

Figure 5. Percent quantified of NeuCode peptide pairs vs. tissue percent incorporation.
A) The percent of peptides quantified for three tissues mixed 1:1 (K_{602}:K_{080}) for all four time points in the lysine isotopologue incorporation study. B) The percent quantified for each tissue at each day for three ratios. C) The percent of peptides quantified for samples at various ratios and stages of incorporations for three tissues.

ensuring that myeloid cells will contain functional BAP1 after the knock-out is induced. After establishing a proper disease model system early experiments demonstrated that BAP1-KO mice display a series of abnormalities including hypoglycemia and elevated liver damage markers. These results implicated metabolic defects in the mice and prompted a construction of a tissue atlas to determine the protein change that occurs across organs involved in metabolic processes (e.g., liver, pancreas, kidney, etc.).[42]

Using the two lysine isotopologues described above (K_{602} and K_{080}) in addition to two newly synthesized lysine isotopologues (K_{521} and K_{440})[26] we constructed two NeuCode SILAM experiments in either a 4-plex or 2-plex configuration to probe a recently described BAP1 knockout mouse model after 21 days of labeling with diet containing NeuCode lysine isotopologues. **Figure 6** displays the scan sequence utilized to analyze a 4-plex NeuCode SILAM sample. A medium resolution (30,000) survey scan was acquired in the Orbitrap (**Figure 6A**) followed by a high resolution (480,000) quantitation scan (**Figure 6B**). As demonstrated above, the peptide (TGYSFVNCK, **Figure 6D**) exists in both unincorporated and incorporated forms and when the peptide is analyzed under high resolution the incorporated partner splits into the four peaks representing each individual lysine isotopologue (**Figure 6C**). Closer inspection of the monoisotopic peak reveals the isotopologue partners that were derived from mice on each custom mixed diet.

Using NeuCode SILAM we set out to determine the tissue specific effects of BAP1 knockout by analyzing nine tissues (brown adipose, heart, kidney, liver, lung, muscle, pancreas, stomach, and white adipose) in both 4-plex and 2-plex formats, enabling quantitation in biological triplicate for proteins identified in both experiments (**Figure 7**). To accomplish this, BAP1-KO mice were aged for 5 weeks, at which point they were given a tamoxifen injection for three days, effectively deleting the 4^{th} and 5^{th} exon of BAP1, which were flanked by lox sites.[43] The mice were then aged for three months at which point they were fed diet containing one of four distinct lysine isotopologues (K_{602}, K_{521}, K_{440}, or K_{080}). After three week of metabolic labeling the tissues were harvest and protein was extracted for proteomic analysis.

Proteins from each tissue were mixed, digested with Lys-C, and separated into 16 high pH reversed phase fractions. Each fraction was analyzed over a 70 min gradient with two technical replicates. In

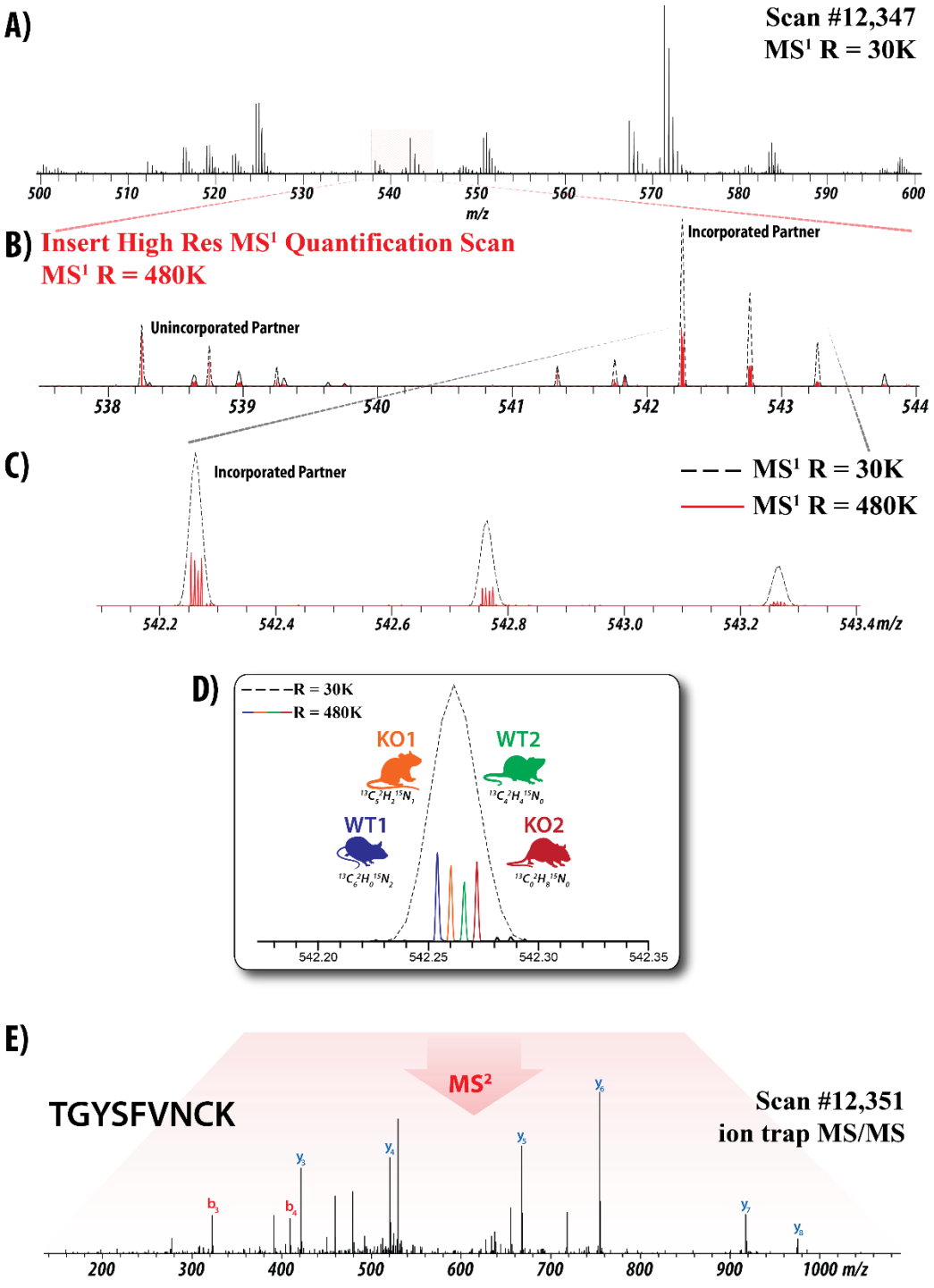


Figure 6. Analysis of NeuCode SILAM 4-plex

A medium resolution survey scan (30,000, A) was used to guide data-dependent analysis of peptide in the ion trap (E). A High resolution quantitation scan (480,000, B) was inserted to reveal isotopologue partners (D) within the heavy incorporated partner (C).

aggregate, 8,651 proteins were identified, with 8,246 proteins quantified in at least two quantitative channels in at least one tissue. Encouragingly, over 5,000 proteins were quantified in 7 of 9 tissues (**Figure 7B**) and over 5,000 proteins were quantified across 5 of the 9 tissues. Given that each tissue was analyzed with deep proteomic coverage we then focused on the number of proteins significantly changing ($p < 0.05$ and fold change >2) in the BAP1-KO as compared to the WT (**Figure 7C**). After applying this filter it is clear that the pancreas and liver experience the largest shift in protein expression with ~16.5% and ~5% of the proteins changing, respectively. Given the highly dynamic protein changes in the liver we compared the NeuCode SILAM proteomic results to transcriptomic data collected on a similar experiment but harvested at the 6 month time point, as opposed to the 12 week time point of the NeuCode samples (**Figure 7D**). Matching data was compiled for 5,046 proteins and generally good agreement was found between the two techniques, suggesting that many of the large protein differences were due to changes at the transcriptional level. Highly regulated proteins were found to be enriched for various gene ontology (GO)[44] terms as inflammation and replication (up-regulated proteins), or β-oxidation, gluconeogenesis, and liver enzymes (down-regulated proteins, **Figure7E**). The description of down regulated proteins as being involved in metabolic processes matched data collected on the level of glucose in the blood of both non-fasted and fasted mice (**Figure 7F**). Considering specific proteins elucidates players in inflammation, replication, lipid metabolism, and glucose metabolism (**Figure 7G**), some of which will be discussed in more depth below.

The large protein differences found in both the live and the pancreas demonstrated that significant changes are occurring at the 12 week time point, but we also wanted to investigate potential tissue and protein level changes at an earlier time point, 2 weeks post KO (**Figure 8**). In addition to being hypoglycemic (**Figure 7E**), both the liver and the pancreas were significantly damaged at the 12 week time point, but healthy at the 2 week time point (**Figure 8A**). Given the extent of damage to the pancreas, it stands to reason that metabolism defects could be due to dysfunctional insulin production; however, at 12 weeks insulin production is comparable to WT levels, most likely due to the health islets of Langerhans which are unaffected by BAP1-KO, as opposed to the acinar cells which are significantly degraded (**Figure**

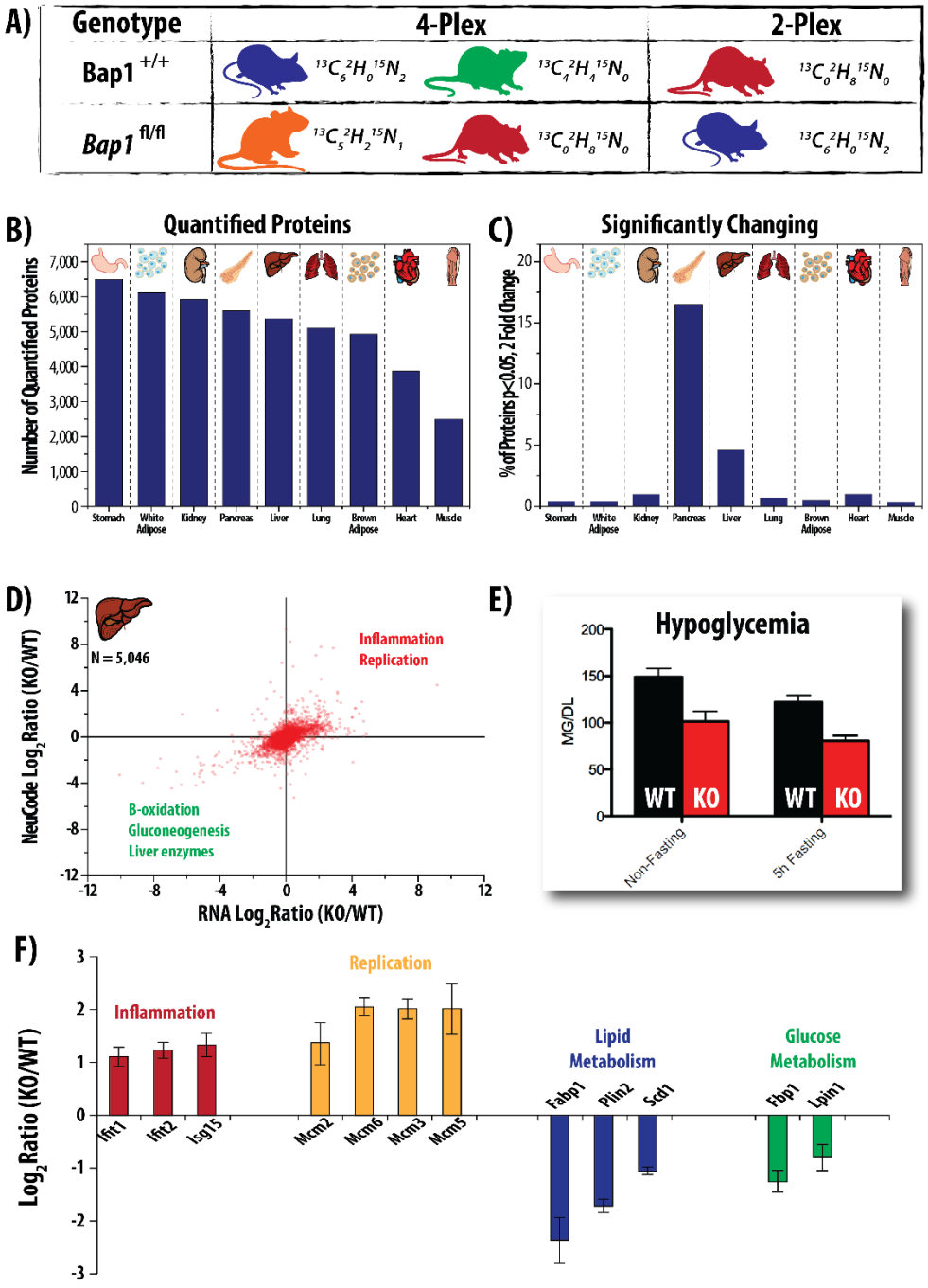


Figure 7. NeuCode SILAM analysis of BAP1-KO mice.
A) 4-plex and 2-plex experiments enabled a 6-plex comparison of WT and BAP1-KO mice. B) Quantified proteins in each tissue. C) Significantly changing ($p < 0.05$, fold change > 2) proteins. D) Comparison of mRNA and protein measurements. E) Mice exhibit hypoglycemia at 12 weeks post BAP1-KO. F) Selected proteins demonstrate large fold changes in key metabolic processes.

8B-C). With the knowledge that the islets cells are spared, we characterized metabolism in the liver at the 2 week time point, when the liver damage is not present. **Figure 8D-E** demonstrated that both glucose and pyruvate tolerance is significantly effected upon BAP1-KO, suggesting metabolism is defective even at early time points when the tissue appears healthy. Two proteins involved in gluconeogenesis, frustose-1, 6-bisphosphate (FBP1) and lipin-1 (LPIN1) were down regulated in the knock out at the three month time point, as measured with NeuCode SILAM (**Figure 8F-G, top row**). Using labeled free protein quantification techniques we determined that FBP1was down-regulated in the BAP1-KO mice in both the fasted and fed state, agreeing with the NeuCode SILAM data. LIPIN1 was also drastically down-regulated in the KO, but only detected in the fasted state in the 2 week samples. These results point to specific proteins that could be involved in defects in gluconeogenesis.

Metabolism defects could also arise from changes in protein expression of previously documented BAP1 associated proteins (**Figure 9**). By examining the change in the protein versus transcript expression in the BAP1-KO mice at three months and highlighting BAP1 associated proteins, ASXL2 arises as a protein that appears to be regulated at the protein, but not transcript level (**Figure 9A**). ASXL2 is a protein that is believed to be involved in metabolic gene transcription (**Figure 9B**), making it a plausible target of BAP1 and mediator of the metabolic dysfunction in BAP1-KO mice. To probe this relationship measured ASXL2 expression at the 2 week time point (before liver damage presents) and found that it was down-regulated, suggesting changes occurring before this early time point (**Figure 9C**). To capture the period of decrease of ASXL2, the transcript level of ASXL2 and other BAP1 associated proteins was measured at the one week mark, revealing similar transcript expression for all BAP1 associated proteins (**Figure 9D**). We also measured the presence of the ubiquitylated form of an ASXL2 peptide at the one week mark and found that ubiquitylation of ASXL2 is lower in the absence of BAP1, suggesting ASXL2 is not a substrate of BAP1, but is affected by deletion of BAP1 (**Figure 9E**). Lastly, we monitored the 1 week transcript levels of many interesting protein targets that arose in the NeuCode SILAM analysis (Figure 9F). This list contains proteins we are interested in following up on including peroxisome proliferator-activated receptor gamma coactivator 1-alpha (PCG1α) and hepatocyte nuclear factor 3-beta (HNF-3B/FOX2a).

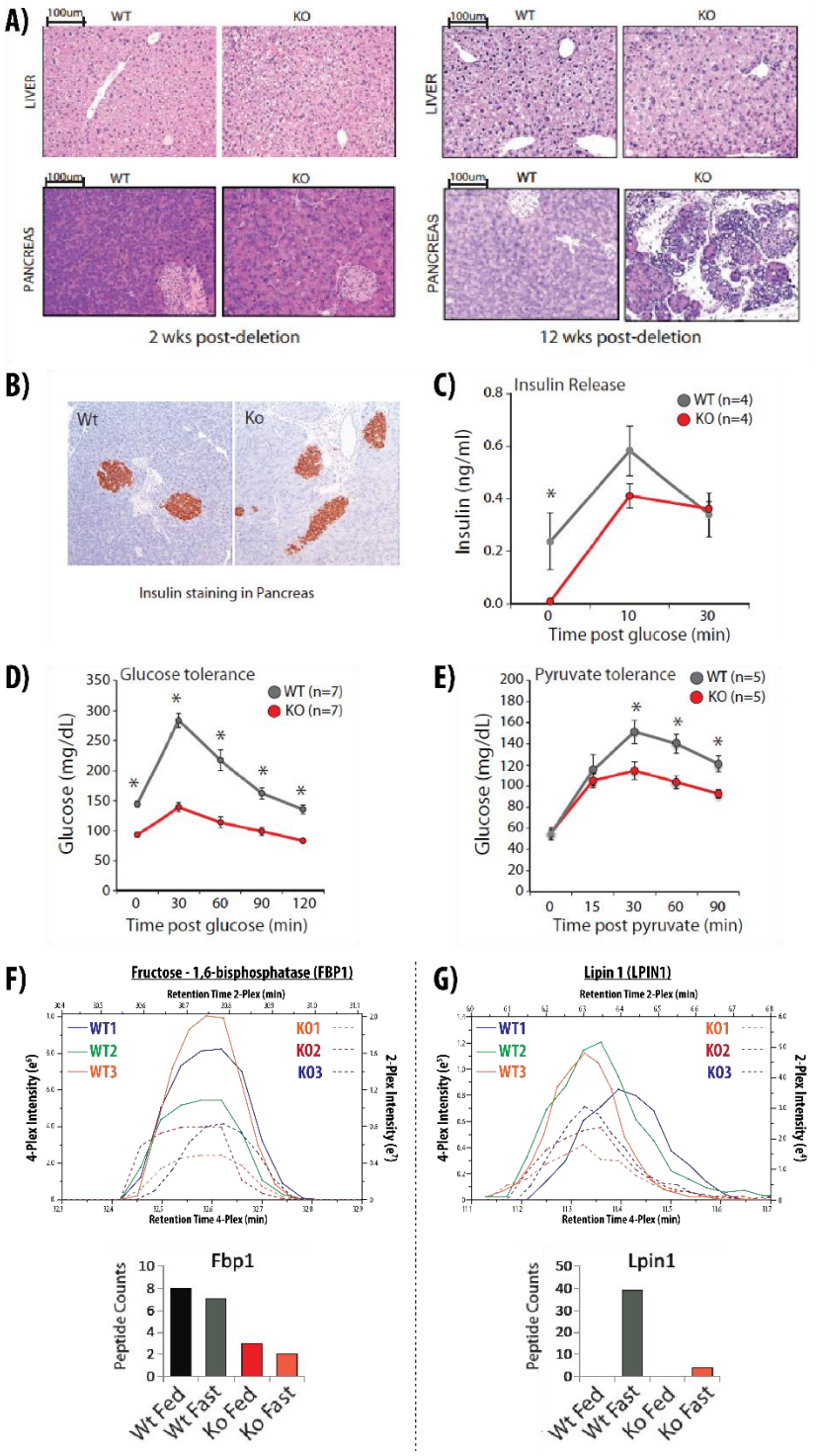


Figure 8. Metabolism Defects in BAP1-KO Mice.

A) Tissue samples from 2 and 12 weeks post BAP1-KO. B) Insulin staining in the pancreas. C) Insulin release from the pancreas. D) Glucose tolerance test at 2 weeks post KO. E) Pyruvate tolerance test at 2 weeks post KO. F-G) NeuCode SILAM data (12 weeks post KO) and label free data (2 weeks post KO) for selected proteins involved in gluconeogenesis.

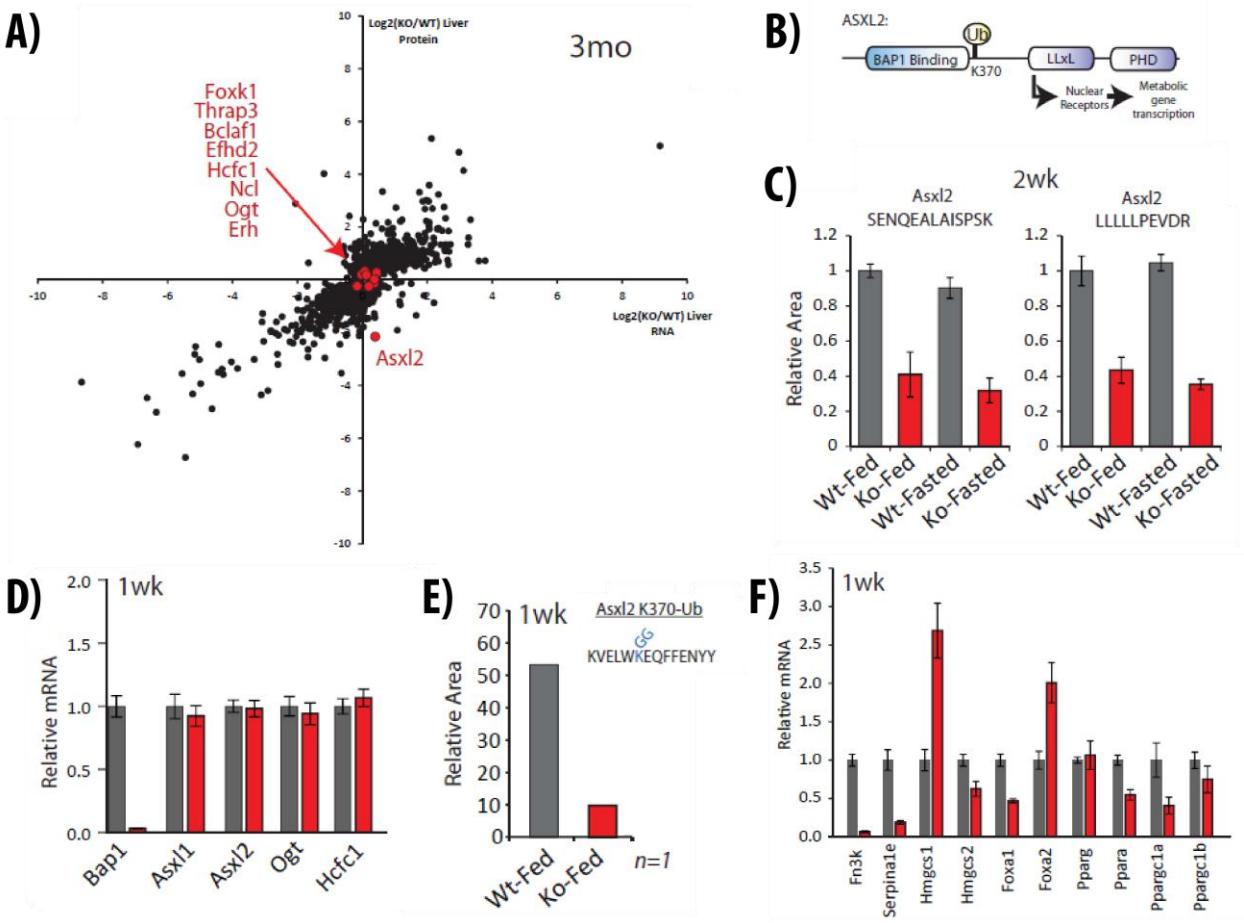


Figure 9. Analysis of BAP1 associated proteins

A) Comparison of mRNA and protein measurements for BAP1 associated proteins. B) ASXL2 gene structure. C) Quantitation of ASXL2 peptides 2 weeks post BAP1-KO. D) mRNA measurements of BAP1 associated proteins 1 week post BAP1-KO. E) Quantitation of a ubiquitylated ASXL2 peptide. F) mRNA measurements from additional protein of interest 1 week post BAP1-KO.

Discussion

In this work we demonstrate a strategy for the direct comparison of up to four mammalian samples utilizing *in vivo* labeling on a shortened times scale – enabling experiments that previously were only possible with a fully labeled reference mammal and multiple binary comparisons. We demonstrated that the key to this technology, lysine isotopologues, incorporate at similar rates within a common laboratory mouse model and enable accurate quantitation in fast incorporating tissues after only ten days of labeling. We then used NeuCode lysine isotopologues in either 2-plex or 4-plex experiments to determine the tissue specific effect of BAP1-KO in mice. We found that proteins in the pancreas and the liver are highly regulated upon BAP1-KO with both of these tissues showing significant damage at 12 weeks post knock-out. Analysis of proteins that are down-regulated in the liver demonstrated that many of these proteins are involved in metabolic processes, agreeing with data that suggested the BAP1-KO mice were hypoglycemic at 12 weeks. Tissue histology demonstrated that while the pancreas is degraded, only the acinar cells are effected by the knock-out and the insulin production of the islets of Langerhans is normal. At two weeks post BAP1-KO, both the pancreas and the liver appear healthy, but many of the same defect in metabolism are beginning to emerge, even at this early time point. Lastly, we focused on BAP1 interacting proteins and found ASXL2 to be a candidate for the mediating protein between the loss of BAP1 and downstream metabolic effect; currently, *in vivo* studies are underway to elucidate the mechanism of this relationship.

Limited labeling combined with multiplexing in NeuCode SILAM paradigm present challenges relating to the signal-to-noise of the measured precursors. At higher multiplexing (4-plex) most biological samples quantified ~10-15% less resolvable peptides than a 2-plex analysis; a penalty incurred for splitting the heavy labeled NeuCode SILAM partner peaks into four channels, thus lowering the signal-to-noise. Generally, this does not impact measurements made at the protein level, as multiple peptides are available to obtain quantitative data. This phenomenon could decrease the number of peptide level measurements of post-translationally modified species (e.g., phosphorylated or acetylated) if the amount of lysine incorporation of the peptide is low. Current efforts are underway to understand the specifics of NeuCode SILAM performance on these types of measurements.

The application presented in this work demonstrates a 6-plex experiment probing the tissue specificity of BAP1 knockout in the same time frame that a typical binary SILAM comparison would require (i.e., two nLC-MS/MS experiments). As instrument performance evolves, the multiplexing capacity of NeuCode SILAM could be expanded to single 6-plex experiments using two newly synthesized lysine isotopologues (K_{422} and K_{341})[26] that are available, but not used in this study. This 6-plex experiment would exacerbate signal-to-noise challenges and presumably require longer labeling times to increase the signal in the heavy partner channel. Additionally, higher multiplexing will also require higher resolution (1,000,000 at *m/z* 400) that is only currently available on FT-ICR instruments and select Orbitrap systems.[45-51] Despite these challenges a 6-plex *in vivo* labeling strategy would enable direct comparison of experimental and control mammals in three biological replicates in just one nLC-MS/MS experiment, with the potential to expand to six biological replicates with the same instrument demands of the current 2-plex SILAM design (i.e., two nLC-MS/MS experiments).

NeuCode SILAM represents an opportunity for biologists to compare mice within model systems directly, without the use of a reference mammal that can be costly and time consuming to fully label. Higher multiplexing allows a reduction in instrument resources and increases the overlap of identified proteins, which increases the number of replicate measurements and statistical power. Peptide overlap will always be higher when comparing two mammals with similar genetic background, and maintaining heavy mammals or generating heavy tissues for any particular mouse model could prove prohibitive.(CITE) Shorter labeling times will enable the routine application of this technology to the particular mouse model of interest, where quantitative data is available after a manner of weeks, rather than months.

This technology could also be applied to other systems were full lysine incorporation is difficult to obtain. For example, primary cell culture from tissues is often limited in the amount of growth, making full incorporation almost impossible to achieve.[52] Using NeuCode isotopologues one would need to only achieve 50% incorporation to enable quantitation. Additionally, preliminary studies in out lab suggest NeuCode reagents could be used in yeast strains that are not auxotrophic for lysine. Typically, the genes responsible for lysine synthesis are knocked-out in order to use a yeast strain for SILAC type experiments.[53]

This severely limits the types of strains that can be analyzed, or adds a significant time commitment to creating lysine auxotrophic strains of exotic yeast systems. Using lysine isotopologues in SILAC experiments would enable quantitation in the same way NeuCode SILAM does – by ignoring the light unincorporated peptides and focusing just on the heavy isotopologues partners. These aspects of NeuCode SILAM present the potential to empower biologists to make comparisons that would have not previously been feasible and greatly expand the toolset for comparing proteomes in mammalian systems.

Experimental Procedures

Lysine Incorporation Study: Isotopologue Labeling of Mice

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 for two weeks. Starting at 10 weeks of age, mice were fed custom lysine-free diet containing 1% K_{608} (n = 12), K_{080} (n = 12), or natural light lysine (n = 3) (Cambridge Isotopes, Boston, MA). The food consumption and weight was monitored throughout the entirety of the experiment and all animals were fed ad libitum and had access to water. At four time points (3, 10, 20, 30 days) mice from K_{608} (n = 3) and K_{080} (n = 3) were sacrificed, with the control group fed light lysine (n = 3) sacrificed at 30 days. After sacrificing animals by cervical dislocation, tissues were dissected, washed in phosphate-buffered saline (PBS), and frozen in liquid nitrogen. Pancreatic islets of Langerhans were isolated by collagenase digestion and a Ficoll gradient separation as previously described.[54]

Lysine Incorporation Study: Proteomic Sample Preparation

Seven tissues (liver, kidney, heart, lung, muscle, brain, and intestine) were homogenized in lysis buffer containing 8M urea, 100 mM NaCl, 50 mM Tris, 5 mM $CaCl_2$, 10 mM nicotinamide, 10 mM sodium butyrate, and phosphatase and protease inhibitor tablet (Roche). Tissues were homogenized in 1 – 2 mL of buffer for 6 – 10 cycles, followed by two washing steps with equivalent amounts of lysis buffer. Subsequently, 1 mL of the lysate was removed for further proteomic analysis. For the intestine and muscle

this aliquot was sonicated to improve protein solubility. All samples were spun at 10,000 x g prior to protein concentration measurement by BCA (Pierce). Additionally, prior to BCA analysis, islet cells were resuspended in 300 µL of lysis buffer and mechanically lysed utilizing a pestle, while plasma samples were simply diluted 10X in lysis buffer.

For all samples, 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 hours. The urea concentration was then diluted to 4 mM using 50 mM Tris, 3 mM $CaCl_2$, 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 and then desalted with tC18 Sep-Pak cartridges (Waters).

For all lysine isotopologue incorporation experiments each sample was resuspended to 1 µg/µL prior to nLC-MS/MS analysis. For mixed ratio experiments, samples from the same tissue (intestine, liver, or brain) and day from either the K_{602} or K_{080} were mixed in 1:1, 5:1, to 10:1 ratios prior to nLC-MS/MS analysis.

Lysine Incorporation Study: Mass spectrometry and high performance liquid chromatography

Online reverse-phase chromatography was performed using a nanoAcquity UPLC (Waters, Milford, MA). Peptides were eluted over an analytical column (75 µm ID) heated to 60°C and packed with 30 cm of 1.7 µm 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 electrospray ionization and analyzed on an Orbitrap Elite mass spectrometer (Thermo Scientific). For isotopologue incorporation study, a survey scan was performed in the Orbitrap at 240,000 resolving power to identify precursors to sample for data-

dependent, top 20 ion trap CAD MS/MS (rapid scan analysis). For mixed ratio experiments, an additional quantitative 480,000 resolving power scan immediately followed the survey scan. Ion trap MS/MS scans were performed while the FT transient 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. For lysine incorporation experiments dynamic exclusion was set to 40 s for -10 ppm and +10 ppm around the selected precursor and 45 s for -25 ppm and +15 ppm around the selected precursor for mixed ratio experiments.

Lysine Incorporation Study: Data Processing

Data reduction was performed with COMPASS,[55] a program that converts output files to searchable text files, as described previously. These text files were then searched against a target-decoy[56] database containing mouse protein entries from Uniprot[57] using the Open Mass Spectrometry Search Algorithm (OMSSA, 2.1.8).[58] Cysteine carbamidomethylation was set as a fixed modification, while methionine oxidation was set as a variable modification. To search NeuCode labeled data a variable modification was considered with the average lysine isotopologue mass (+8.0322 Da), as the precursor mass was determined from a medium resolution scan. Precursor mass tolerance was defined as 150 ppm and fragment ion mass tolerance was set to 0.35 Da. Search results were filtered based on precursor mass error and e-value using FDR-Optimizer within COMPASS.

Lysine Incorporation Study: Protein Quantification

Peptides were quantified using NeuQuant, as previously described.[26] To determine lysine incorporation, samples were quantified by measuring the intensity of the light or heavy peptide partners. Following quantification peptides were grouped into proteins and the ratio of heavy to light signal was converted to a percent lysine incorporation. To determine the accuracy of measurements throughout the time course study, the abundance of lysine isotopologue partners was measured at the peptide level before

protein grouping, and then the relative level of the isotopologue intensities was compared at the protein level.

BAP-1 KO Mouse: Custom diet, labeling, tissue extraction, and lysis

Custom NeuCode mouse diet was created to enable up to a 4-plex analysis. Mouse diet deficient in lysine was purchased (Harlan, Madison, WI) and fortified with 1% of a lysine isotopologue. For these diets we used the commercially available Lys_{602} and Lys_{080} isotopologues as well as custom synthesized Lys_{521} and Lys_{440} (Cambridge Isotopes, Cambridge, MA).[26] The diets were pressed into pellets and dried before being stored at -20°C until their use. To enable a 6-plex analysis wild-type three mice were fed diet containing Lys_{602}, Lys_{440}, or Lys_{080}, while three BAP1-KO mice were fed diet containing Lys_{521}, Lys_{080}, or Lys_{602}.

BAP1-KO mice were aged for five weeks, at which point they received tamoxifen injections for five days, inducing the BAP-KO. Mice were then fed the heavy diet for 21 days, at which point the organs were harvested and flash-frozen, in a method similar to that described above. Nine tissues including the brown adipose, white adipose, heart, lung, muscle, kidney, stomach, liver, and pancreas were homogenized in a method similar to that above before preparation for proteomic analysis. Following protein concentration estimation, a small amount of each lysate was mixed to provide two samples (4-plex and 2-plex) per tissue. Protein disulfide bonds were the 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 hours. The urea concentration was then diluted to 4 mM using 50 mM Tris, 3 mM $CaCl_2$, 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 and then desalted with tC18 Sep-Pak cartridges (Waters). This test sample was analyzed and, if necessary, lysate mixing ratios were adjusted to ensure total protein input was the same across all channels.

The newly mixed lysate was then prepped in an identical manner prior to fractionation. Each sample was separated into 16 high pH reversed phase fractions and dried down prior to nLC-MS/MS analysis.

BAP-1 KO Mouse: 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 µm ID) heated to 60°C and packed with 30 cm of 1.7 µm 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 electrospray ionization and analyzed on an Orbitrap Elite mass spectrometer (Thermo 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 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.

BAP-1 KO Mouse: Data processing

Data reduction was performed with COMPASS,[55] a program that converts output files to searchable text files, as described previously. These text files were then searched against a target-decoy (CITE) database containing mouse protein entries from Uniprot (CITE) using the Open Mass Spectrometry Search Algorithm (OMSSA, 2.1.8). (CITE, CITE) Cysteine carbamidomethylation was set as a fixed modification, while methionine oxidation was set as a variable modification. To search NeuCode labeled data a variable modification was considered with the average lysine isotopologue mass (+8.0322 Da), as the precursor

mass was determined from a medium resolution scan. Precursor mass tolerance was defined as 150 ppm and fragment ion mass tolerance was set to 0.35 Da. Search results were filtered based on precursor mass error and e-value using FDR-Optimizer within COMPASS.

BAP-1 KO Mouse: Quantification and data aggregation

Peptides were quantified using NeuQuant, as described previously. Quantified peptides from all samples were then grouped into proteins using Protein Hoarder within COMPASS. Grouping peptides from all samples was important to ensure accurate combination of the resulting protein quantitative data. Data from each 4-plex and 2-plex experiment for each tissue was then combined using an in-house program Procyon (**Figure 10**). Procyon was developed as a tool to combine quantitative data from multiple experiments and uses a unique identifier (e.g., protein group number or protein group number/modification site) to quickly (3 to 30 sec) create an output file that contains quantitative data from each experiment for each protein group. This represents a significant time savings as doing this manually could take many hours to days to perform. Procyon also executes significance testing by enabling users to add each unique sample to a common group. If multiple experiments are loaded (as in the case with the 4-plex and 2-plex experiments) Procyon will calculate the mean normalized value for each protein measurement prior to combining the data from each experiment. This enables Procyon to perform significance testing using either a fold change, p-value, or corrected p-value (P-value, Benjamini-Hochberg correction). Users can set custom significance thresholds and using Uniprot identifiers Procyon can determine if any gene ontology terms are enriched in proteins that are significantly changing, significantly changing and up-regulated, or significantly changing and down-regulated. Additionally, Procyon enables the user to filter the data based on any particular GO term (e.g., mitochondrial) or perform normalization on a specific subset of the data. Procyon outputs contain Log2 transformations of the normalized intensities, the mean normalized intensities, the ratios of user defined comparisons, the significance of those comparisons based on the p-value or P-value, and the GO annotations for each protein group. Additionally, the user can specific any

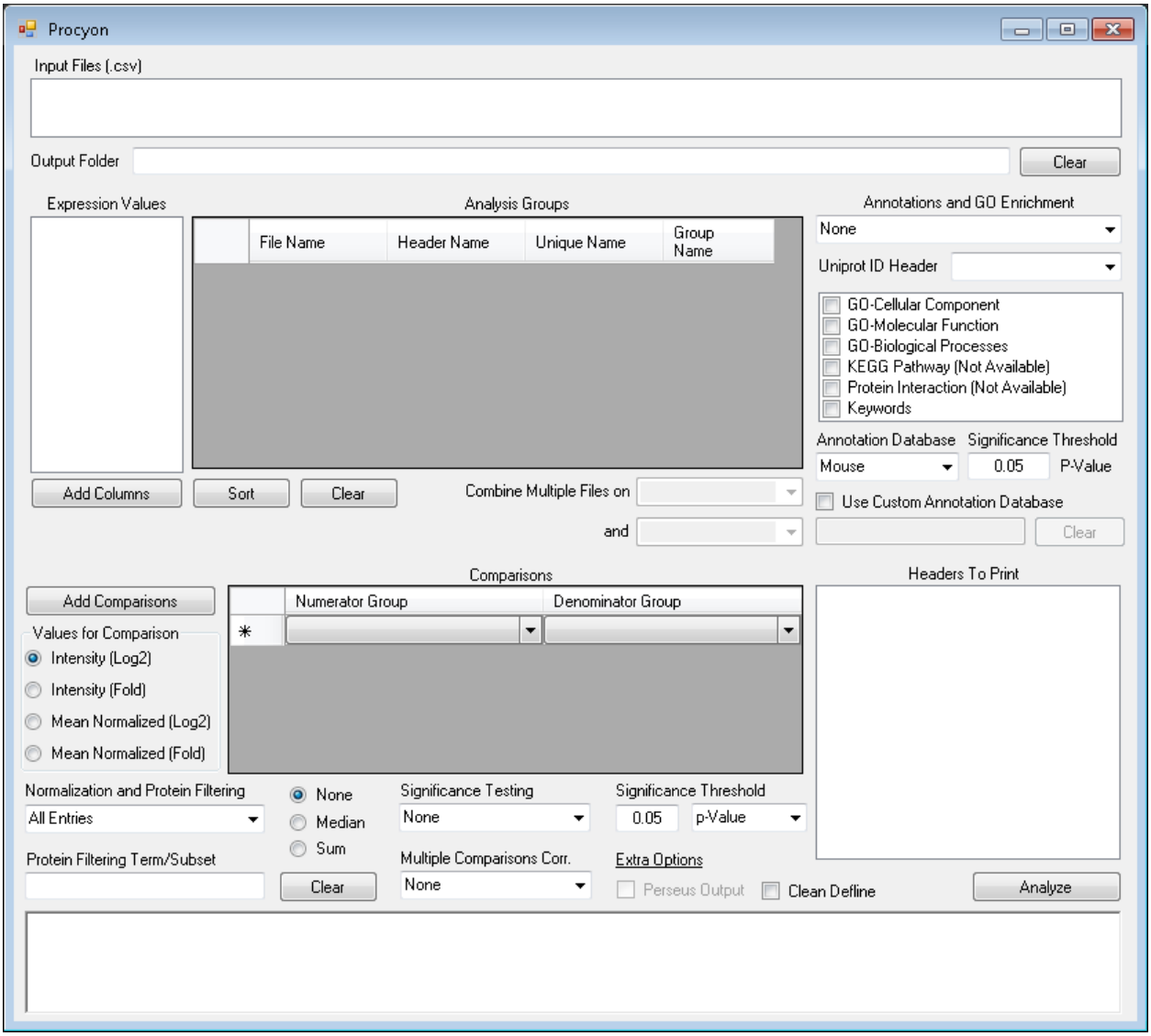


Figure 10. Procyon for analysis of quantitative proteomic data.

column in the original input file that they wish to be in the output file creating a custom file that can easily be distributed to collaborators.

Acknowledgements

C.M.R was funded by an NSF Graduate Research Fellowship and NIH Traineeship (T32GM008505). A.E.M. gratefully acknowledges support from an NIH-funded Genomic Sciences Training Program (5T32HG002760).

References

1. Kendall SD, Adam SJ, & Counter CM. Genetically Engineered Human Cancer Models Utilizing Mammalian Transgene Expression. *Cell Cycle*, 2006, 5(10):1074-1079.

2. Becher OJ & Holland EC. Genetically Engineered Models Have Advantages over Xenografts for Preclinical Studies. *Cancer Research*, 2006, 66(7):3355-3359.

3. Frese KK & Tuveson DA. Maximizing mouse cancer models. *Nat Rev Cancer*, 2007, 7(9):654-658.

4. Hebert Alexander S, Dittenhafer-Reed Kristin E, Yu W, Bailey Derek J, Selen Ebru S, Boersma Melissa D, Carson Joshua J, Tonelli M, Balloon Allison J, Higbee Alan J, Westphall Michael S, Pagliarini David J, Prolla Tomas A, Assadi-Porter F, Roy S, Denu John M, & Coon Joshua J. Calorie Restriction and SIRT3 Trigger Global Reprogramming of the Mitochondrial Protein Acetylome. *Molecular Cell*, 2013, 49(1):186-199.

5. Hebert AS, Richards AL, Bailey DJ, Ulbrich A, Coughlin EE, Westphall MS, & Coon JJ. The One Hour Yeast Proteome. *Molecular & Cellular Proteomics*, 2014, 13(1):339-347.

6. Senko MW, Remes PM, Canterbury JD, Mathur R, Song Q, Eliuk SM, Mullen C, Earley L, Hardman M, Blethrow JD, Bui H, Specht A, Lange O, Denisov E, Makarov A, Horning S, & Zabrouskov V. Novel Parallelized Quadrupole/Linear Ion Trap/Orbitrap Tribrid Mass Spectrometer Improving Proteome Coverage and Peptide Identification Rates. *Analytical Chemistry*, 2013, 85(24):11710-11714.

7. Brumbaugh J, Rose CM, Phanstiel DH, Thomson JA, & Coon JJ. Proteomics and pluripotency. *Critical Reviews in Biochemistry and Molecular Biology*, 2011, 46(6):493-506.

8. Merrill AE & Coon JJ. Quantifying proteomes and their post-translational modifications by stable isotope label-based mass spectrometry. *Current Opinion in Chemical Biology*, 2013, 17(5):779-786.

9. Weisser H, Nahnsen S, Grossmann J, Nilse L, Quandt A, Brauer H, Sturm M, Kenar E, Kohlbacher O, Aebersold R, & Malmström L. An Automated Pipeline for High-Throughput Label-Free Quantitative Proteomics. *Journal of Proteome Research*, 2013, 12(4):1628-1644.

10. Cox J, Hein MY, Luber CA, Paron I, Nagaraj N, & Mann M. MaxLFQ allows accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction. *Molecular & Cellular Proteomics*, 2014.

11. Rosenberger G, Ludwig C, Röst HL, Aebersold R, & Malmström L. aLFQ: An R-package for estimating absolute protein quantities from label-free LC-MS/MS proteomics data. *Bioinformatics*, 2014.

12. Thompson A, Schäfer J, Kuhn K, Kienle S, Schwarz J, Schmidt G, Neumann T, & Hamon C. Tandem Mass Tags: A Novel Quantification Strategy for Comparative Analysis of Complex Protein Mixtures by MS/MS. *Analytical Chemistry*, 2003, 75(8):1895-1904.

13. Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, Khainovski N, Pillai S, Dey S, Daniels S, Purkayastha S, Juhasz P, Martin S, Bartlet-Jones M, He F, Jacobson A, & Pappin DJ. Multiplexed Protein Quantitation in Saccharomyces cerevisiae Using Amine-reactive Isobaric Tagging Reagents. *Molecular & Cellular Proteomics*, 2004, 3(12):1154-1169.

14. McAlister GC, Huttlin EL, Haas W, Ting L, Jedrychowski MP, Rogers JC, Kuhn K, Pike I, Grothe RA, Blethrow JD, & Gygi SP. Increasing the Multiplexing Capacity of TMTs Using Reporter Ion Isotopologues with Isobaric Masses. *Analytical Chemistry*, 2012, 84(17):7469-7478.

15. Werner T, Becher I, Sweetman G, Doce C, Savitski MM, & Bantscheff M. High-Resolution Enabled TMT 8-plexing. *Analytical Chemistry*, 2012, 84(16):7188-7194.

16. Werner T, Sweetman G, Savitski MF, Mathieson T, Bantscheff M, & Savitski MM. Ion Coalescence of Neutron Encoded TMT 10-Plex Reporter Ions. *Analytical Chemistry*, 2014, 86(7):3594-3601.

17. Rose CM, Venkateshwaran M, Volkening JD, Grimsrud PA, Maeda J, Bailey DJ, Park K, Howes-Podoll M, den Os D, Yeun LH, Westphall MS, Sussman MR, Ané J-M, & Coon JJ. Rapid Phosphoproteomic and Transcriptomic Changes in the Rhizobia-legume Symbiosis. *Molecular & Cellular Proteomics*, 2012, 11(9):724-744.

18. Honarpour N, Rose CM, Brumbaugh J, Anderson J, Graham RLJ, Sweredoski MJ, Hess S, Coon JJ, & Deshaies RJ. F-box Protein FBXL16 Binds PP2A-B55α and Regulates Differentiation of Embryonic Stem Cells along the FLK1+ Lineage. *Molecular & Cellular Proteomics*, 2014, 13(3):780-791.

19. Ow SY, Salim M, Noirel J, Evans C, Rehman I, & Wright PC. iTRAQ Underestimation in Simple and Complex Mixtures: "The Good, the Bad and the Ugly". *Journal of Proteome Research*, 2009, 8(11):5347-5355.

20. Wenger CD, Lee MV, Hebert AS, McAlister GC, Phanstiel DH, Westphall MS, & Coon JJ. Gas-phase purification enables accurate, multiplexed proteome quantification with isobaric tagging. *Nat Meth*, 2011, 8(11):933-935.

21. Ting L, Rad R, Gygi SP, & Haas W. MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. *Nat Meth*, 2011, 8(11):937-940.

22. McAlister GC, Nusinow DP, Jedrychowski MP, Wühr M, Huttlin EL, Erickson BK, Rad R, Haas W, & Gygi SP. MultiNotch MS3 Enables Accurate, Sensitive, and Multiplexed Detection of Differential Expression across Cancer Cell Line Proteomes. *Analytical Chemistry*, 2014, 86(14):7150-7158.

23. Ong S-E, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, & Mann M. Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. *Molecular & Cellular Proteomics*, 2002, 1(5):376-386.

24. Zhu H, Pan S, Gu S, Bradbury EM, & Chen X. Amino acid residue specific stable isotope labeling for quantitative proteomics. *Rapid Communications in Mass Spectrometry*, 2002, 16(22):2115-2123.

25. Lee MV, Topper SE, Hubler SL, Hose J, Wenger CD, Coon JJ, & Gasch AP. A dynamic model of proteome changes reveals new roles for transcript alteration in yeast. *Molecular Systems Biology*, 2011, 7(1):n/a-n/a.

26. Merrill AE, Hebert AS, MacGilvray ME, Rose CM, Bailey DJ, Bradley JC, Wood WW, ElMasri M, Westphall MS, Gasch AP, & Coon JJ. NeuCode labels for relative protein quantification. *Molecular & Cellular Proteomics*, 2014.

27. Hebert AS, Merrill AE, Bailey DJ, Still AJ, Westphall MS, Strieter ER, Pagliarini DJ, & Coon JJ. Neutron-encoded mass signatures for multiplexed proteome quantification. *Nat Meth*, 2013, 10(4):332-334.

28. Rose CM, Merrill AE, Bailey DJ, Hebert AS, Westphall MS, & Coon JJ. Neutron Encoded Labeling for Peptide Identification. *Analytical Chemistry*, 2013, 85(10):5129-5137.

29. Richards AL, Vincent CE, Guthals A, Rose CM, Westphall MS, Bandeira N, & Coon JJ. Neutron-encoded Signatures Enable Product Ion Annotation From Tandem Mass Spectra. *Molecular & Cellular Proteomics*, 2013, 12(12):3812-3823.

30. Rhoads TW, Rose CM, Bailey DJ, Riley NM, Molden RC, Nestler AJ, Merrill AE, Smith LM, Hebert AS, Westphall MS, Pagliarini DJ, Garcia BA, & Coon JJ. Neutron-Encoded Mass Signatures for Quantitative Top-Down Proteomics. *Analytical Chemistry*, 2014, 86(5):2314-2319.

31. Wu CC, MacCoss MJ, Howell KE, Matthews DE, & Yates JR. Metabolic Labeling of Mammalian Organisms with Stable Isotopes for Quantitative Proteomic Analysis. *Analytical Chemistry*, 2004, 76(17):4951-4959.

32. McClatchy DB, Dong M-Q, Wu CC, Venable JD, & Yates JR. 15N Metabolic Labeling of Mammalian Tissue with Slow Protein Turnover. *Journal of Proteome Research*, 2007, 6(5):2005-2010.

33. McClatchy DB, Liao L, Park SK, Venable JD, & Yates JR. Quantification of the synaptosomal proteome of the rat cerebellum during post-natal development. *Genome Research*, 2007, 17(9):1378-1388.

34. Geiger T, Velic A, Macek B, Lundberg E, Kampf C, Nagaraj N, Uhlen M, Cox J, & Mann M. Initial Quantitative Proteomic Map of 28 Mouse Tissues Using the SILAC Mouse. *Molecular & Cellular Proteomics*, 2013, 12(6):1709-1722.

35. Krüger M, Moser M, Ussar S, Thievessen I, Luber CA, Forner F, Schmidt S, Zanivan S, Fässler R, & Mann M. SILAC Mouse for Quantitative Proteomics Uncovers Kindlin-3 as an Essential Factor for Red Blood Cell Function. *Cell*, 2008, 134(2):353-364.

36. Robles MS, Cox J, & Mann M. In-Vivo Quantitative Proteomics Reveals a Key Contribution of Post-Transcriptional Mechanisms to the Circadian Regulation of Liver Metabolism. *PLoS Genet*, 2014, 10(1):e1004047.

37. Bailey DJ, Rose CM, McAlister GC, Brumbaugh J, Yu P, Wenger CD, Westphall MS, Thomson JA, & Coon JJ. Instant spectral assignment for advanced decision tree-driven mass spectrometry. *Proceedings of the National Academy of Sciences*, 2012, 109(22):8411-8416.

38. Testa JR, Cheung M, Pei J, Below JE, Tan Y, Sementino E, Cox NJ, Dogan AU, Pass HI, Trusa S, Hesdorffer M, Nasu M, Powers A, Rivera Z, Comertpay S, Tanji M, Gaudino G, Yang H, & Carbone M. Germline BAP1 mutations predispose to malignant mesothelioma. *Nat Genet*, 2011, 43(10):1022-1025.

39. Carbone M, Yang H, Pass HI, Krausz T, Testa JR, & Gaudino G. BAP1 and cancer. *Nat Rev Cancer*, 2013, 13(3):153-159.

40. Ventii KH, Devi NS, Friedrich KL, Chernova TA, Tighiouart M, Van Meir EG, & Wilkinson KD. BRCA1-Associated Protein-1 Is a Tumor Suppressor that Requires Deubiquitinating Activity and Nuclear Localization. *Cancer Research*, 2008, 68(17):6953-6962.

41. Dey A, Seshasayee D, Noubade R, French DM, Liu J, Chaurushiya MS, Kirkpatrick DS, Pham VC, Lill JR, Bakalarski CE, Wu J, Phu L, Katavolos P, LaFave LM, Abdel-Wahab O, Modrusan Z, Seshagiri S, Dong K, Lin Z, Balazs M, Suriben R, Newton K, Hymowitz S, Garcia-Manero G, Martin F, Levine RL, & Dixit VM. Loss of the Tumor Suppressor BAP1 Causes Myeloid Transformation. *Science*, 2012, 337(6101):1541-1546.

42. Chao EC & Henry RR. SGLT2 inhibition — a novel strategy for diabetes treatment. *Nat Rev Drug Discov*, 2010, 9(7):551-559.

43. Nagy A. Cre recombinase: The universal reagent for genome tailoring. *genesis*, 2000, 26(2):99-109.

44. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, & Sherlock G. Gene Ontology: tool for the unification of biology. *Nat Genet*, 2000, 25(1):25-29.

45. Bruce JE, Anderson GA, Lin C-Y, Gorshkov M, Rockwood AL, & Smith RD. A novel high-performance Fourier transform ion cyclotron resonance cell for improved biopolymer characterization. *Journal of Mass Spectrometry*, 2000, 35(1):85-94.

46. Syka JEP, Marto JA, Bai DL, Horning S, Senko MW, Schwartz JC, Ueberheide B, Garcia B, Busby S, Muratore T, Shabanowitz J, & Hunt DF. Novel Linear Quadrupole Ion Trap/FT Mass Spectrometer: Performance Characterization and Use in the Comparative Analysis of Histone H3 Post-translational Modifications. *Journal of Proteome Research*, 2004, 3(3):621-626.

47. Makarov A, Denisov E, & Lange O. Performance Evaluation of a High-field Orbitrap Mass Analyzer. *Journal of the American Society for Mass Spectrometry*, 2009, 20(8):1391-1396.

48. Michalski A, Damoc E, Hauschild J-P, Lange O, Wieghaus A, Makarov A, Nagaraj N, Cox J, Mann M, & Horning S. Mass Spectrometry-based Proteomics Using Q Exactive, a High-performance Benchtop Quadrupole Orbitrap Mass Spectrometer. *Molecular & Cellular Proteomics*, 2011, 10(9).

49. Kaiser N, Quinn J, Blakney G, Hendrickson C, & Marshall A. A Novel 9.4 Tesla FTICR Mass Spectrometer with Improved Sensitivity, Mass Resolution, and Mass Range. *Journal of the American Society for Mass Spectrometry*, 2011, 22(8):1343-1351.

50. Denisov E, Damoc E, Lange O, & Makarov A. Orbitrap mass spectrometry with resolving powers above 1,000,000. *International Journal of Mass Spectrometry*, 2012, 325–327(0):80-85.

51. Michalski A, Damoc E, Lange O, Denisov E, Nolting D, Müller M, Viner R, Schwartz J, Remes P, Belford M, Dunyach J-J, Cox J, Horning S, Mann M, & Makarov A. Ultra High Resolution Linear Ion Trap Orbitrap Mass Spectrometer (Orbitrap Elite) Facilitates Top Down LC MS/MS and Versatile Peptide Fragmentation Modes. *Molecular & Cellular Proteomics*, 2012, 11(3).

52. Spellman DS, Deinhardt K, Darie CC, Chao MV, & Neubert TA. Stable Isotopic Labeling by Amino Acids in Cultured Primary Neurons: Application to Brain-derived Neurotrophic Factor-dependent Phosphotyrosine-associated Signaling. *Molecular & Cellular Proteomics*, 2008, 7(6):1067-1076.

53. Fröhlich F, Christiano R, & Walther TC. Native SILAC: Metabolic Labeling of Proteins in Prototroph Microorganisms Based on Lysine Synthesis Regulation. *Molecular & Cellular Proteomics*, 2013, 12(7):1995-2005.

54. Bhatnagar S, Oler AT, Rabaglia ME, Stapleton DS, Schueler KL, Truchan NA, Worzella SL, Stoehr JP, Clee SM, Yandell BS, Keller MP, Thurmond DC, & Attie AD. Positional Cloning of a Type 2 Diabetes Quantitative Trait Locus; <italic>Tomosyn-2</italic>, a Negative Regulator of Insulin Secretion. *PLoS Genet*, 2011, 7(10):e1002323.

55. Wenger CD, Phanstiel DH, Lee MV, Bailey DJ, & Coon JJ. COMPASS: A suite of pre- and post-search proteomics software tools for OMSSA. *Proteomics*, 2011, 11(6):1064-1074.

56. Elias JE & Gygi SP. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Meth*, 2007, 4(3):207-214.

57. Consortium TU. Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic Acids Research*, 2012, 40(D1):D71-D75.

58. Geer LY, Markey SP, Kowalak JA, Wagner L, Xu M, Maynard DM, Yang X, Shi W, & Bryant SH. Open Mass Spectrometry Search Algorithm. *Journal of Proteome Research*, 2004, 3(5):958-964.

Chapter 10

Joint accumulation multiplexing and elution order prediction enables parallel reaction monitoring of more than 1,000 peptides

CMR designed research, performed experiments, analyzed data, and wrote the paper.

This chapter is in preparation for submission.

Abstract

Parallel reaction monitoring enables the simultaneous analysis of all fragment ions from a precursor peptide by utilizing high resolution mass analyzers, which afford great specificity. High mass accuracy can be exploited to enable the analysis of multiple (2 to 10) precursors in one analysis, a method we refer to as joint accumulation multiplexing (JAM). Here we describe the implementation of joint accumulation multiplexing for use with parallel reaction monitoring (JAM-PRM) to enable targeted analysis of a large number of peptides. We combined JAM-PRM with a previously described online elution order (EO) prediction algorithm, which corrects for retention time shifts that might occur run to run, to enable robust targeted analysis. Furthermore, by utilizing triplex NeuCode SILAC, we increased our multiplexing capabilities by three fold. All together these analyses were implemented for the targeted analysis of 510 of the most intense peptides, of which ~94% were quantified with excellent accuracy. We then analyzed 311 peptides from the lowest abundance proteins and found that we quantified only 87% with adequate accuracy. By implementing custom dwell times we increased our coverage of the 311 low abundance peptides to 97% while regaining the excellent accuracy we had achieved in the prior analysis. Lastly, we pushed the limits of this method by targeting 1,450 of the most intense peptides, finding that we quantified 94% with adequate accuracy.

Introduction

Selected reaction monitoring (SRM), targets specific proteins and fragment ions to achieve highly accurate quantitation and afford the ability to determine absolute protein amount.[1,2] However, this targeted approach currently has much lower throughput than other quantitation methods, such as SILAC[3,4] or isobaric labeling methods.[5-7] In a typical SRM experiment isotopically labeled peptide standards are synthesized and mixed with an experimental sample at a known concentration.[8] Great care is taken to design assays which analyze intense fragment ions and produce retention times that do not shift from analysis to analysis. Mass spectrometers such as triple quadrupole mass spectrometers are used to isolate the peptide of interest with the first quadrupole, fragment in the second quadrapole, followed by selection of the fragment ion with the third quadrupole.[8] The analysis consists of oscillating between the standard peptide and the experimental peptide, a process that reduced the duty cycle and limits the number of peptides that can be analyzed in a given analysis.

SRM methods have given way to multiple reaction monitoring (MRM),[9] a related technique that typically selects and analyzes three transitions for each peptide in one mass analysis, gaining sensitivity while still limiting the amount of time that is devoted to mass analysis. MRM experiments have been the focus for many groups attempting to increase the number of peptides that are targeted in one analysis. Carr and co-workers recently detailed the importance of chromatography, combining analytical chromatography columns with longer LC gradients (180 min) to enable the targeted analysis of 400 peptide partners (800 peptide species).[10] We surmised, the same level of targeting could be achieved in a shorter nLC-MS/MS analysis, such as 90 min. To accomplish this we aimed to exploit a relatively new technique called parallel reaction monitoring (PRM).

Recently, high resolution mass spectrometers have been used to perform PRM, where peptides are fragmented and all of the fragment ions are analyzed.[11-13] The high resolution mass analysis slows duty cycle, but sensitivity can be regained by simultaneous analysis of all fragment ions. PRM enables the simultaneous analysis of all fragment ions from a precursor peptide by utilizing high resolution mass

analyzers, which afford great specificity. High mass accuracy can be exploited to enable the analysis of multiple (2 to 10) precursors in one analysis, a method we refer to as joint accumulation multiplexing.

Here we describe the implementation of joint accumulation multiplexing for use with parallel reaction monitoring (JAM-PRM) to enable the targeted analysis of a large number of peptides. We combined JAM-PRM with a previously described online elution order (EO) prediction algorithm,[14] which corrects for retention time shifts that might occur run to run, to enable robust targeted analysis. Furthermore, by utilizing triplex NeuCode SILAC, we increased our multiplexing capabilities by three fold. All together these analyses were implemented for the targeted analysis of 510 of the most intense peptides, of which ~94% were quantified with excellent accuracy. We then analyzed 311 peptides from the lowest abundance proteins and found that we quantified only 87% with adequate, but not great, accuracy. By implementing custom dwell times we increased our coverage of the 311 low abundance peptides to 97% while regaining the excellent accuracy we had achieved in the prior analysis. Lastly, we pushed the limits of this method by targeting 1,450 of the most intense peptides, finding that we quantified 94% with adequate accuracy.

Results

Joint accumulation multiplexing and elution order prediction implemented on an Orbitrap Fusion

Selected reaction monitoring (SRM) the workhorse of many targeted experiments, as well as multiple reaction monitoring (MRM) are typically implemented on a low resolution mass spectrometer (e.g., triple quad). The specificity of the quadrupole mass selection events compensates for the low resolution of the mass analysis performed in the third quadrupole or ion trap mass analyzer. High resolution (HR) mass analysis trades speed for mass accuracy to enable parallel reaction monitoring (PRM), where all fragment ions are analyzed together in a HR mass analyzer (e.g., Orbitrap). While HR analysis is longer (48-500 ms) as compared to a quadrupole or ion trap (10 ms), the ability to quantify a larger number of transitions enables quantitative analysis similar to that of a triple quadrupole.[11] One clear advantage of PRM is the ability to multiplex the analysis of multiple precursors, referred to from here on as joint accumulation multiplexing (JAM). A JAM scan sequence involves the injection and fragmentation of multiple precursors

(2-10 total) after which all of the fragment ions are analyzed simultaneously in the Orbitrap. The ability to perform JAM analysis is implemented on the newest generation mass spectrometer (Orbitrap Fusion). Combining JAM and PRM (JAM-PRM) into one targeted scan sequence enables a drastic increase in the number of peptides that can be targeted in a routine analysis. To further increase the number of peptides that are targeted nLC-MS/MS analysis reproducible retention time scheduling with narrow window (1-2 min) is required.

To enable reproducible scheduling with narrow targeted windows, regardless of shifts in LC performance, we implemented a previously described elution order (EO) prediction algorithm. Utilizing a set of training runs (3-5 runs) the elution order of identified peptides is calculated by an in house program *Pacer* and saved to a table that can be loaded onto the instrument firmware. For these experiments, a sample grown in light lysine is used for training runs and the *m/z* value is altered to account for the mass shift imparted by lysine NeuCode isotopologues. A medium resolution (15,000) survey scan is used to calculate the EO and populate the list of targets for the current JAM-PRM scan. The target list is generated by *Pacer*, a program that was specifically written to construct JAM-PRM target lists, and loaded into the method editor as normal, but with elution order values as the start and stop times. A more detailed description of *Pacer* is located in the methods. As demonstrated below the EO order algorithm has enabled the targeted analysis of over 1,000 peptides.

Targeted analysis of the 510 most intense peptides

To determine if the combination of JAM-PRM and EO prediction could increase the number of targets analyzed in an nLC-MS/MS experiment we constructed a target list of 510 of the most intense peptides from an analysis of LysC digested yeast (**Figure 1**). *Pacer* was used to ensure that no more than 20 targets were available at any EO, with the simultaneous analysis of 5 precursors (JAM = 5) and an EO window of ± 1.45. This list was used in a triplicate analysis to determine the reproducibility of JAM-PRM (**Figure 1 and Figure 2**). To increase the multiplexing of this analysis, a NeuCode triplex experiment was implemented with samples mixed in a ratio of 1:1:1. A HR PRM scan (240,000) was used to ensure that

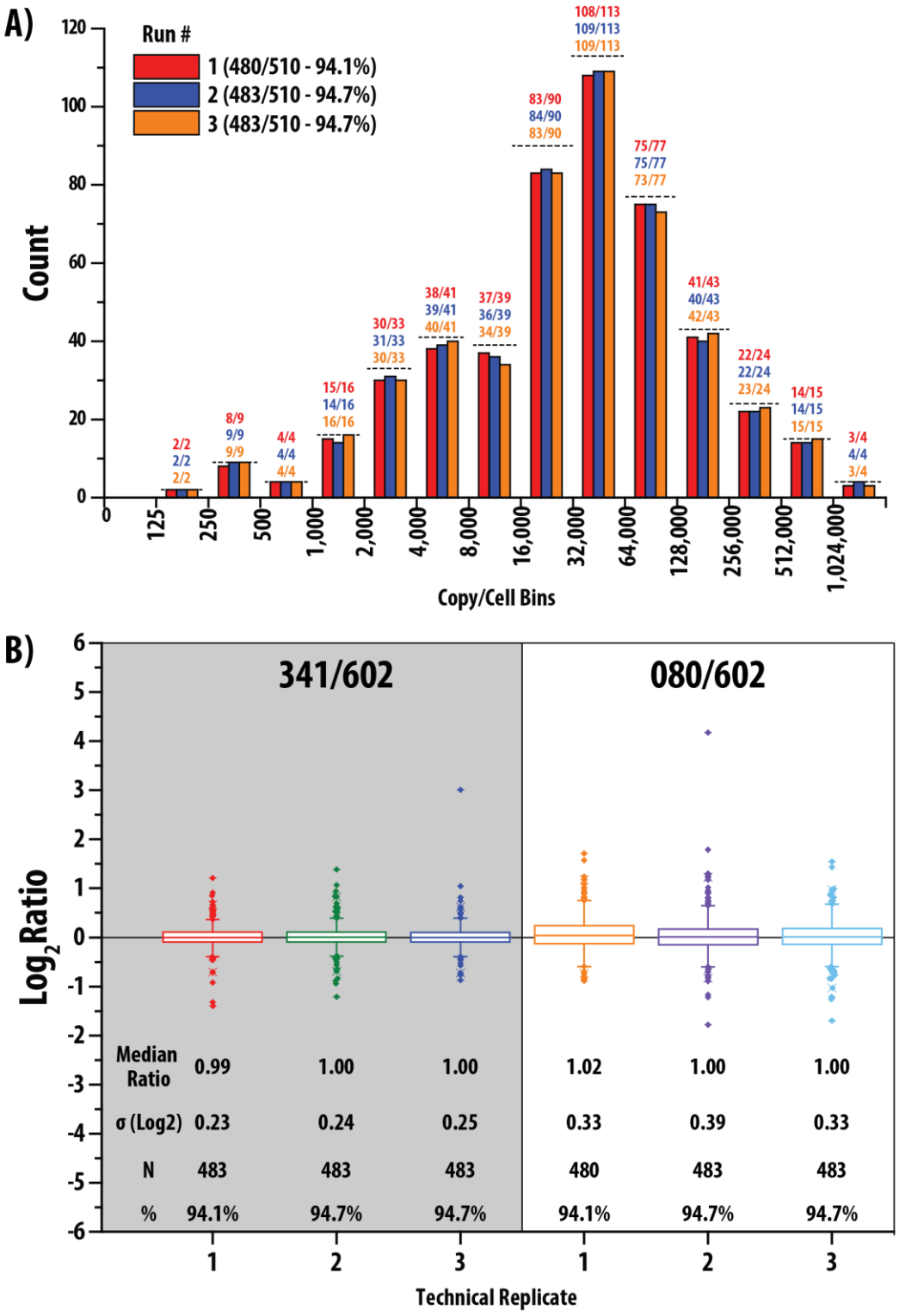


Figure 1. JAM-PRM analysis of 510 Targets.

A) Copy per cell distribution of the 510 most intense targets used for these analyses. B) Quantitative accuracy of three technical replicates of the targeted analysis.

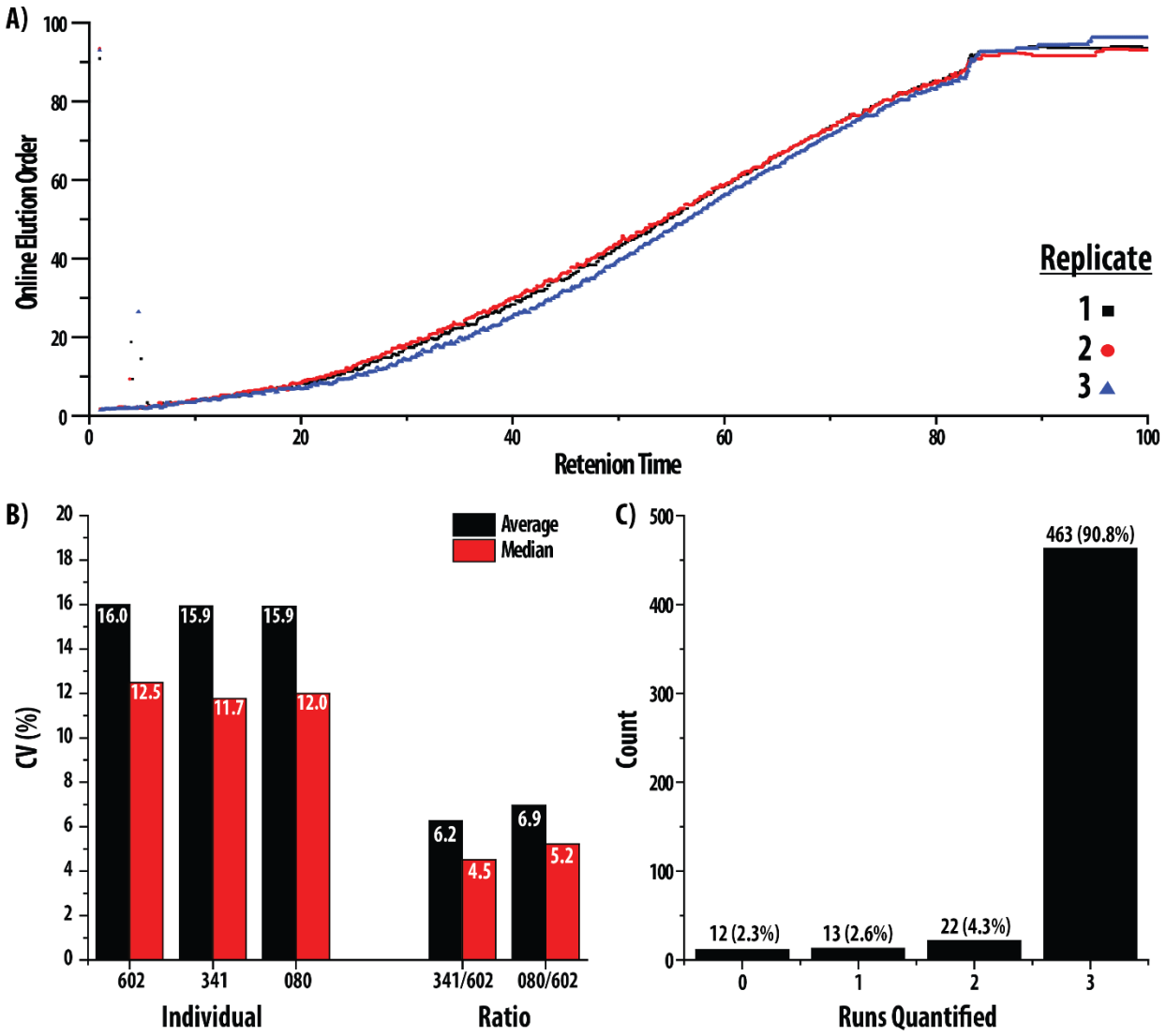


Figure 2. Analysis of technical replicate analysis of 510 targets.

A) The online elution order prediction for the three technical analyses of 510 targets B) The coefficient of variance for the three technical analyses of 510 targets C) The distribution of the number of analyses in which each peptide was quantified.

the fragment ions of NeuCode partners would be resolved. Impressively, ~94% of the targets were identified in each of the three analyses – representing ~1440 different peptide species, including peptides from low copy per cell proteins (**Figure 1A**). Additionally, the quantitative accuracy of the NeuCode measurements is excellent with the correct median ratio and low standard deviation (0.23 to 0.39, **Figure 1B**).

Further analysis of the triplicate runs details very promising results (**Figure 2**). Here, the EO prediction algorithm was vital to the replicate analysis, as the peptides appeared to shift between the 2nd and 3rd nLC-MS/MS experiments (**Figure 2A**). Despite this retention time shift, the same number of peptides were quantified in these runs emphasizing the importance of the EO algorithm. To determine if the intensity of fragment ions is consistent across technical replicates the average and median coefficient of variance (CV) was plotted for the intensity of each NeuCode channel or ratio of the NeuCode partners (**Figure 2B**). At this initial stage, the median values for individual (~12%) and ratio (~5%) measurements are very encouraging. Lastly, the percentage of peptides identified in all three experiments (~91%), is promising, but further improvements to the EO algorithm will hopefully raise that percentage to 95%.

Targeted analysis of the 311 low copy per cell peptides

To ensure that the results demonstrated with the 510 most intense peptides would translate to peptides belonging to low copy per cell proteins, a list of 311 peptides was created from the same set of library runs as above - ensuring that no more than 20 peptides were targeted at any time, with a JAM of 5, and an EO window of ± 2.5. Due to the low abundance of these peptides, we made additional alterations to the instrument code to enable the implementation of custom dwell times. Additionally, because the 240,000 resolution scan takes ~500 ms, we altered the code to limit the overall injection time of precursors by grouping peptides together based on their custom dwell time. To determine if these alterations improved the performance of the targeted method we analyzed the list of 311 peptides with each approach (**Figure 3**). Interestingly, **Figure 3B** demonstrates a lower percentage of quantified peptides for the 311 targets (87%) with higher standard deviation (0.55 and 0.63) than the analysis of the 510 peptides above (94% and 0.23 to 0.39, respectively). This is most likely due to the lower abundance of these peptides as compared to

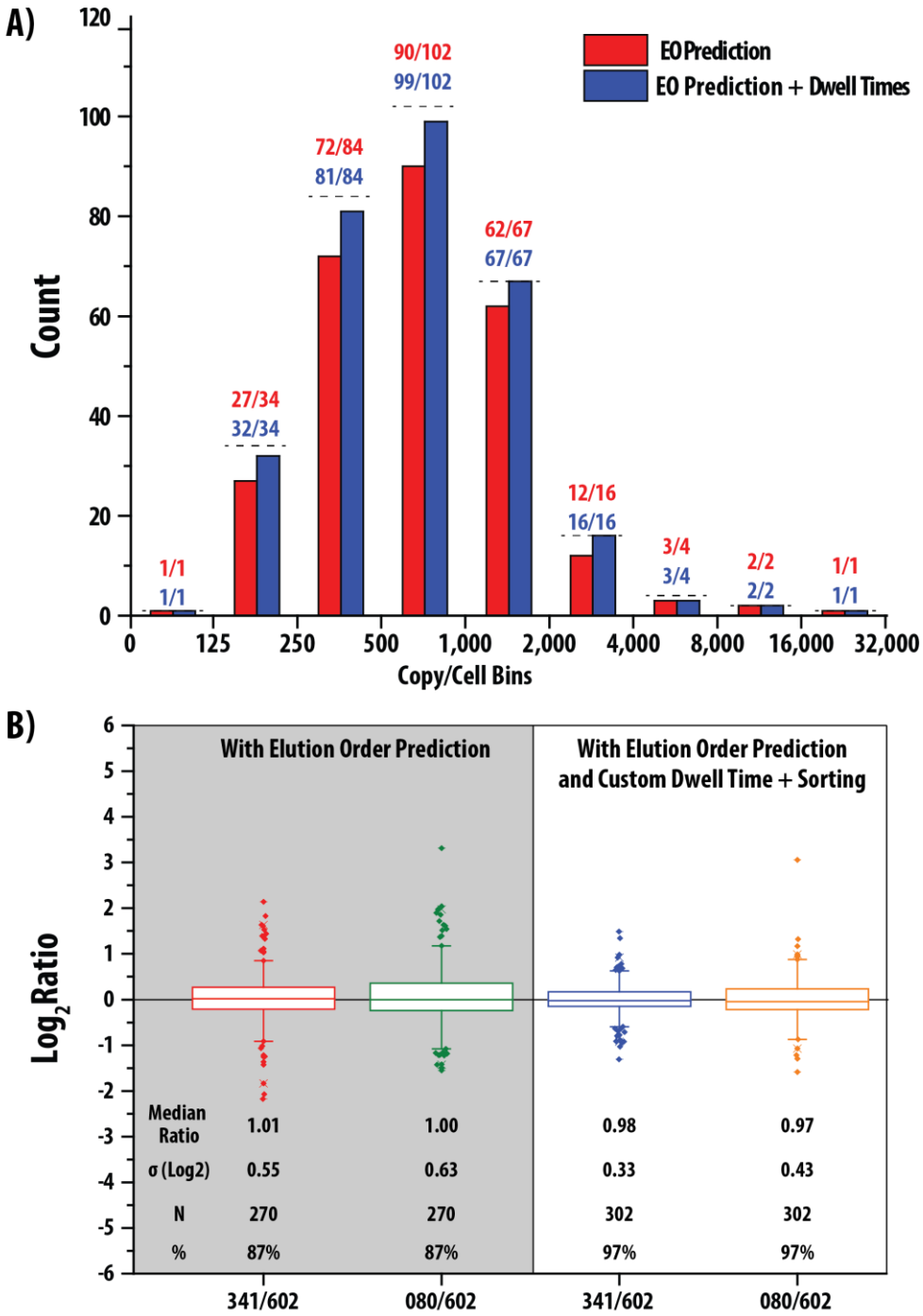


Figure 3. JAM-PRM analysis of 311 Targets.

A) Copy per cell distribution of the 311 lowest copy per cell targets used for these analyses using either EO prediction or EO prediction plus custom dwell times and dwell time sorting B) Quantitative accuracy of three technical replicates of the targeted analysis.

the 510 list (**Figure 3A**). To account for the low intensity of these peptides we implemented custom dwell times with a sorting algorithm to limit overall injection time. These changes led to an increase in the number of quantified peptides to 97% while decreasing the standard deviation to 0.33 and 0.43 for the NeuCode partner ratios (**Figure 3B**). These results underscore the importance of appropriate dwell times, especially for peptides that are lower abundance, to ensure that the quantitative accuracy is sufficient. These improvements to the method will be implemented in future experiments.

Targeted analysis of 1450 of the most intense targets

Lastly, to demonstrate the potential power of this technique a target list was made from a LysC digest of yeast to ensure that no more than 40 targets were considered at any time, with a JAM of 10 and an EO window of ± 1.0 - yielding 1,450 targets (**Figure 4**). These preliminary results utilized the EO algorithm, but not the custom dwell time method. Encouragingly, utilizing an automated quantitation program *PrimeQuant*, 1,367 targets were quantified, representing ~ 4,100 peptide species with the triplex NeuCode isotopologues (**Figure 4**). The accuracy of these measurements could be improved as the standard deviation of 0.64 and 0.72 for the two NeuCode ratios is higher than desired (**Figure 4**). Improvements to *PrimeQuant* will ideally limit the need to manually curate the quantitative data, as manual inspection of this data would most likely improve the quantitative accuracy. Experiments are currently being performed to reproduce these results in a triplicate analysis to demonstrate a routine targeted analysis of over 1,000 peptides.

Discussion

We describe the implementation of joint accumulation multiplexed parallel reaction monitoring (JAM-PRM) combined with online elution order (EO) prediction to demonstrate targeting of over 1,000 peptides. We first introduced the concept of JAM-PRM and the importance of online EO prediction to enable the targeting of a large number of targets. Using a list of 510 intense targets we demonstrated that JAM-PRM quantifies ~94% of the targets with excellent accuracy across triplicate analyses. A targeted

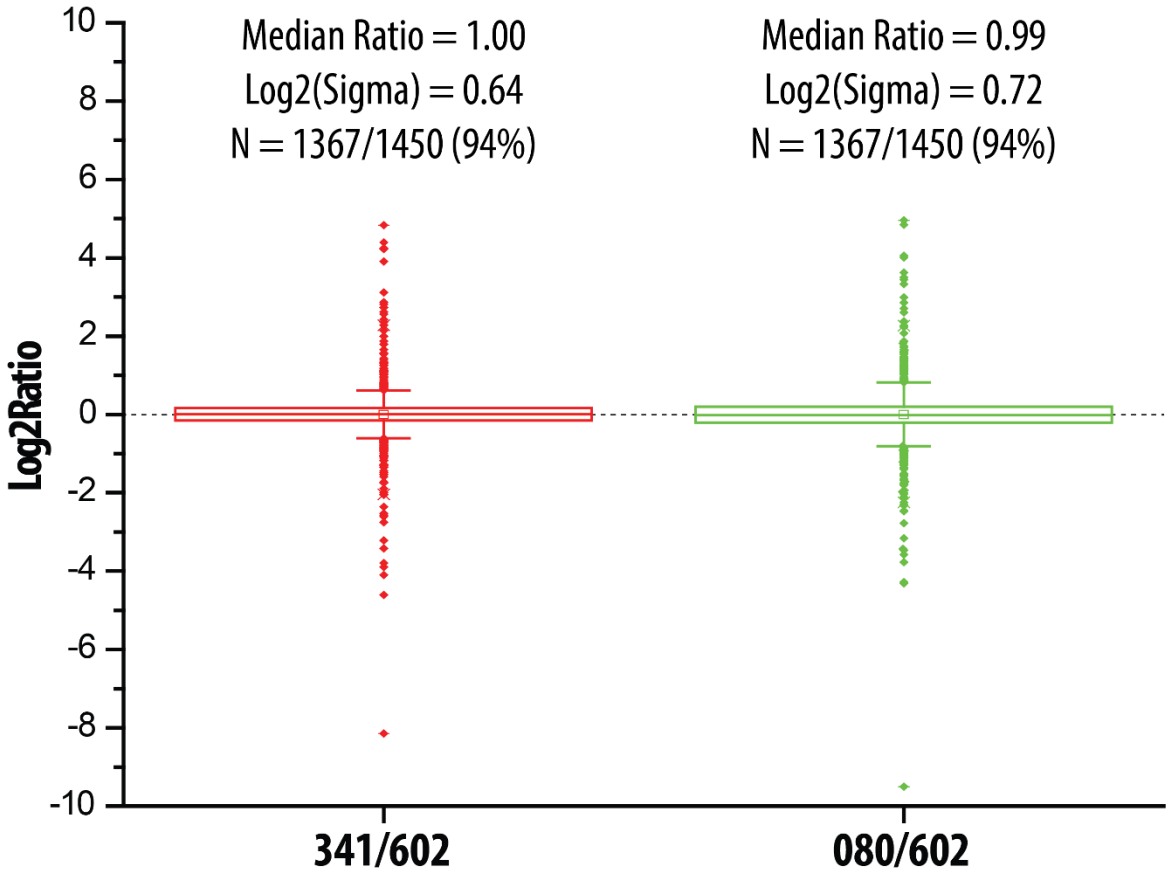


Figure 4. JAM-PRM analysis of 1450 Targets.
Targeted analysis of the 1450 most abundant peptides using elution order prediction.

analysis of 311 peptides from low copy per cell proteins demonstrated that the percentage of targets quantified and quantitative accuracy is diminished using just the EO algorithm. Alterations to the instrument code enabled the use of custom dwell times without drastically increasing the precursor injection time, recovering the percentage of quantified targets (97%) and the quantitative accuracy (standard deviations of 0.33 and 0.43). These results underscore the importance of custom dwell times in the analysis low abundance peptides. Lastly, we constructed a target list of 1,450 peptides and quantified 1,367 with adequate accuracy, demonstrating the potential for this method to analyze over 1,000 peptide targets in one nLC-MS/MS analysis.

The results presented here represent preliminary experiments of the implementation of JAM-PRM and future experiments will focus on the targeted analysis of peptides across various knockout strains of yeast as they undergo to a diauxic shift. This experiment will underscore the ability of JAM-PRM to routinely analyze peptides from various types of samples. One of the largest challenges will continue to be the reproducibility of chromatography, the main driver behind the implementation of the EO algorithm. From these experiments it appears that linear gradients afford the highest reproducibility for the EO algorithm, a potentially limiting factor in some laboratories. Additionally, reproducible chromatography in the library runs is crucial in the construction of an optimized target list – a step that will need to be emphasized if this methods is to be adopted. The methods described here are not limited to NeuCode analysis and could be implemented with traditional standard peptides or even traditional SILAC reagents, opening new possibilities for targeted assays.

Experimental procedures

Sample Preparation

For lysine NeuCode SILAC, *Saccharomyces cerevisiae* strain BY4741 Lys1Δ was grown in defined, synthetic-complete (SC, Sunrise Science) drop out media supplemented with either "*light*" lysine (+0 Da), "K_{602}" $^{13}C_6/^{15}N_2$ lysine (+8.0142 Da, Cambridge Isotopes), "K_{341}" $^{13}C_3/^{2}H_4/^{15}N_1$ lysine (+8.0322 Da, Cambridge Isotopes), or "K_{080}" $^{2}H_8$ (+8.0502 Da, Cambridge Isotopes). Cells were allowed to propagate

for a minimum of 10 doublings to ensure complete lysine incorporation. Upon reaching mid-log phase, the cells were harvested by centrifugation at 3,000 ×g for 3 minutes and washed three times with chilled ddH_2O. Cell pellets were re-suspended in 5mL lysis buffer (50mM Tris pH8, 8M urea, 75mM sodium chloride, 100mM sodium butyrate, 1mM sodium orthovanadate, protease and phosphatase inhibitor tablet), and total protein was extracted by glass bead milling (Retsch).

Protein concentration of yeast lysate was measured by BCA (Pierce). For NeuCode analysis yeast proteins from respective cultures were mixed in a 1:1:1 ratio before reduction by 5 mM DTT, alkylation with 14 mM iodoacetamide, and capping by an additional 5 mM DTT. Prior to digestion, the sample was diluted to 2 mM Urea using 50 mM Tris and 3 mM $CaCl_2$. Digestion was carried out by adding LysC (Roche Applied Science, Indianapolis, IN) at a 1:100 enzyme to substrate ratio and incubating overnight at room temperature. Peptides were then acidified with trifluoroacetic acid (TFA) to quench the reaction and de-salted using C-18 solid phase extraction (SPE) columns (Waters, Milford, MA) before analysis.

Library nLC-MS/MS experiments

Yeast samples grown on light lysine were used to construct the peptide library. Online reverse-phase chromatography was performed using a nanoAcquity UPLC (Waters, Milford, MA). Peptides were eluted over an analytical column (75 µm ID) heated to 60°C and packed with 30 cm of 1.7 µm 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 electrospray ionization and analyzed on an Orbitrap Fusion mass spectrometer (Thermo Scientific). A survey scan was performed in the Orbitrap at 60,000 resolving power to identify precursors to sample for data-dependent, top speed HCD MS/MS ion trap analysis (rapid scan analysis. 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 $5x10^5$ and $5x10^3$, respectively. Dynamic exclusion was set to 20 s for -10 ppm and +10 ppm around the selected precursor.

Pacer

Target peptide lists were constructed using an in house programs *Pacer* (**Figure 5**). *Pacer* takes peptide outputs from *ProteinHoarder* and target list outputs from *Prime* to create a JAM-PRM target list, with the ability to perform EO prediction. This is done by first finding the peak apex for each identified peptide and calculating the EO for each library run. The lists of peptides are then compiled from each run and the average precursor intensity and EO for each peptide is saved. Utilizing the target lists input by the user the list of peptides identified is filtered further and used for target list construction. The user specifies the number of peptides that should be considered at any point, as well as the number of precursors that will be simultaneously analyzed (JAM). Additional features allow users to specify the number of peptides per copy per cell bin, standard deviation of the EO, EO window, number of library runs in which the peptide was identified, start and stop EO, and the number of y ions necessary for consideration. *Pacer* also enables the user to determine how the target list will be filled in, allowing the list to be constructed starting with the least or most intense, lowest or highest copy per cell, or highest intensity then lowest copy per cell. *Pacer* will order the peptide targets by these attributes and then starting with the user specified parameter will being to add targets to the list. Using EO increments of 0.01, each target is added until the maximum number of targets is reached for that bin. This is done for each peptide target to ensure that the maximum number of peptides are targeted in each run. *Pacer* displays the copy per cell distribution, number of peptides targeted at each EO bin, and the approximate injection time if custom dwell times are used. Additionally, *Pacer* enables users to make target lists based on gene ontology (GO) annotations. By selecting a GO annotation, such as mitochondrion, *Pacer* will create a target list to maximize the number of peptide that are targeted in a JAM-PRM analysis. *Pacer* outputs ".csv" files that can be loaded directly into the method editor of an Orbitrap Fusion. Additionally, *Pacer* outputs background ".sav" files that can be used for EO prediction and target ".sav" files that can be used to implement custom dwell times for the peptide targets.

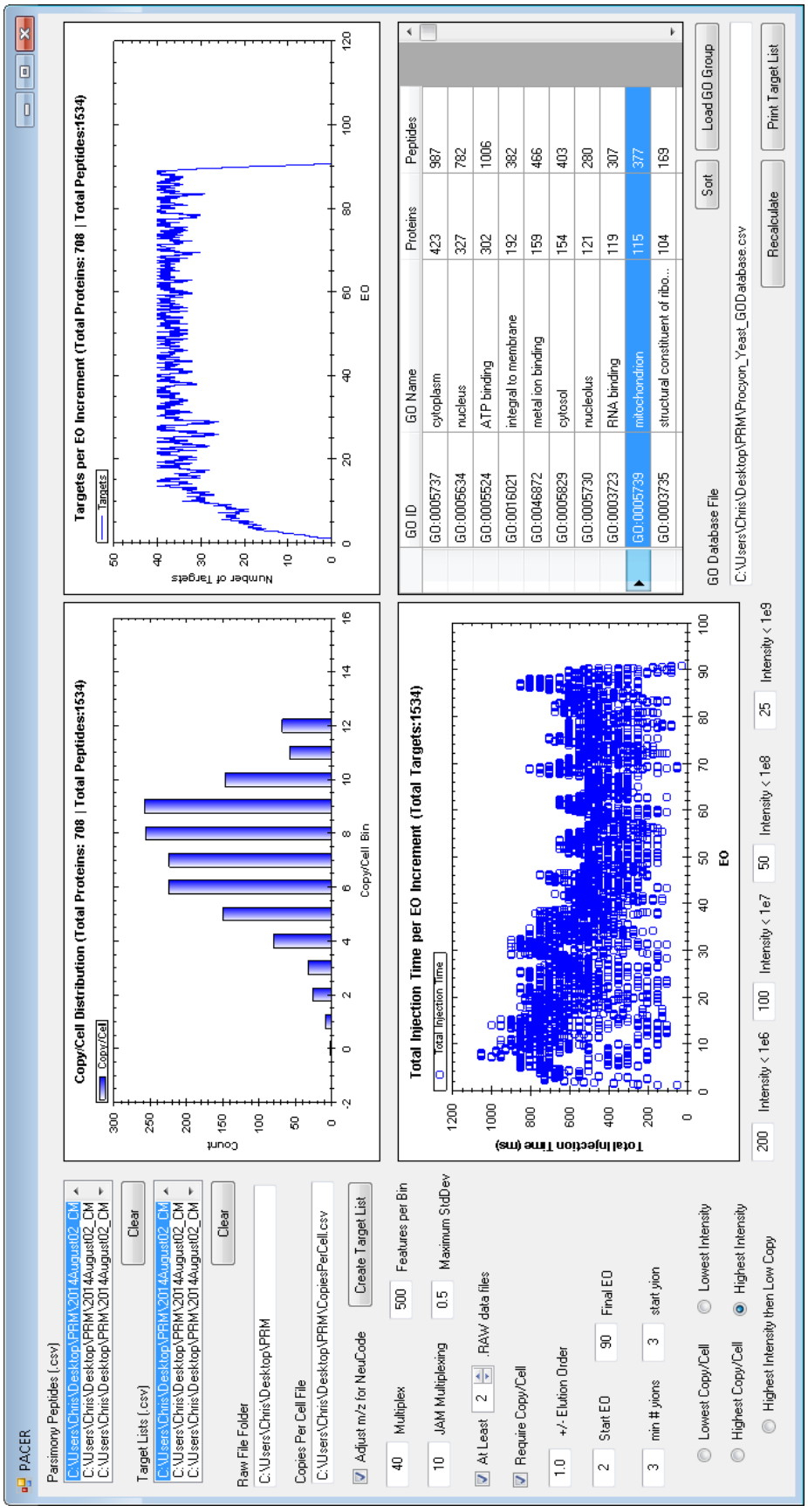


Figure 5. PACER for the construction of JAM-PRM targeted methods

Instrument code and targeted analysis

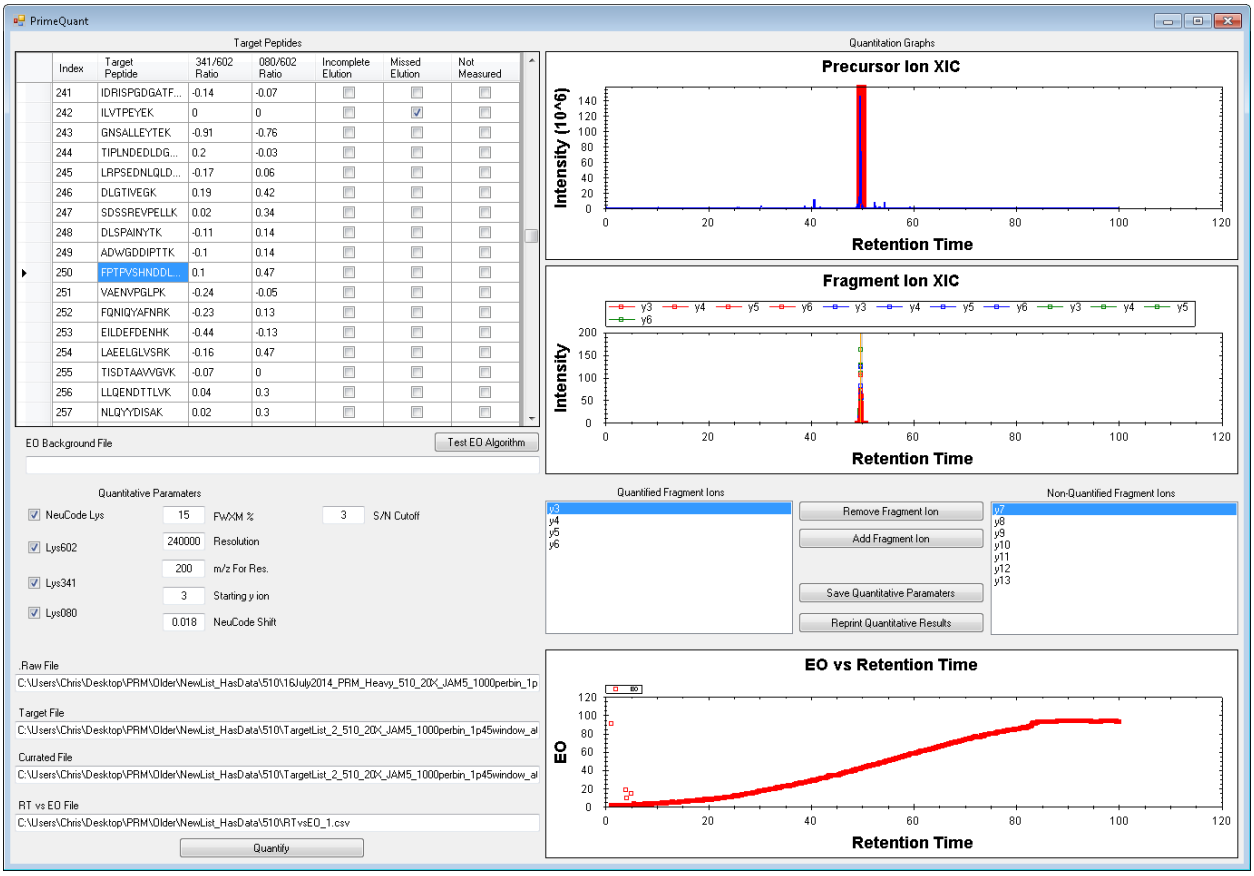
Instrument control code was modified to enable EO prediction, the use of custom dwell times, and peptide dwell time clustering; the code was written in the programing language "Lua". The EO prediction algorithm was similar that previously published; however, instead of returning a range of EO values to consider peptide targets an average EO was returned. This average EO was then converted into a "second" value that was used to filter the targeted list using the canonical instrument code. To enable proper filtering of the list, the imported target list was furnished with EO start and stop values in place of minute start and stop values. The background *m/z* and EO values were loaded into memory and queried from while the instrument was running. Dwell time sorting was performed within the instrument code and was accomplished by determining the number of JAM scans necessary to analyze the current targets, then sorting targets into those lists in a sequence manner starting with the lowest abundance peptides. Considerations were made to ensure that isolation windows of the peptides were not overlapping. This departs from the normal peptide sorting which is done solely on *m/z*. Target lists containing peptide identities and custom dwell times were also loaded onto the instrument and queried during the nLC-MS/MS experiment.

Target analysis was performed by constructing a method with two experiments. The first experiment comprises a 15,000 resolution Orbitrap scan followed by a top 1 ion trap HCD MS/MS scan. This first experiment is necessary to perform the EO prediction. The second experiment was set up as a targeted MS^2 experiment with the number of multiplexed ions set to 5 or 10. For these analyses the MS^2 resolution was set to 240,000 with a maximum injection time of 500 ms, AGC target of $5e^5$, isolation window of 2, and HCD NCE of 30. Target lists exported from *Pacer* were loaded into the method editor directly. For these analyses a liner gradient was used and all nLC-MS/MS experiment were performed over a 90 minute period.

PrimeQuant

To analyze the targets data we used an in house program *PrimeQuant*, which is specifically written to analyze JAM-PRM data collected using EO prediction (**Figure 6A**). *PrimeQuant* takes the target list output from *Pacer*, online RT vs. EO calculation, and .RAW data file to quantify both light or NeuCode labeled peptides. *PrimeQuant* enables the user to simultaneously view both the precursor and fragment ion extracted ion chromatographs (XIC) to help confidently identify the correct elution profile (**Figure 6B**). Additionally, *PrimeQuant* enables the user to curate the quantified peptides and classify low quality measurements in three categories: partial elution profile, missed elution profile, and not quantified. The user can also use the graphical user interface to change the retention time bounds and fragment ions used for quantitation. These features enable the user to filter fragment ions that contain interfering signals. The updated retention time bounds and quantified fragment ions can be saved, such that they can be re-loaded at a future point in time.

A)



B)

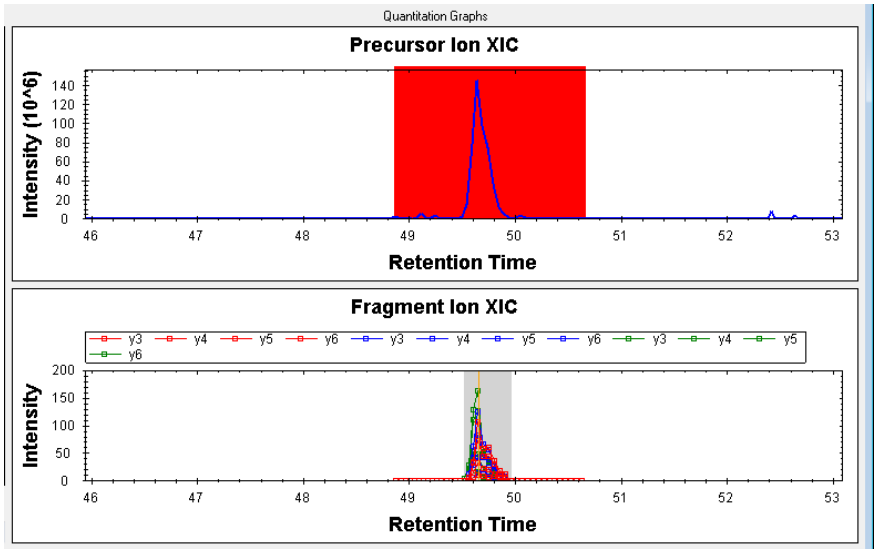


Figure 6. PrimeQuant for the analysis of JAM-PRM data

A) The graphical user interface that allows the user to curate of the targeted data. B) Zoom in of the precursor peptide FPTPVSHNDDLYGK.

References

1. Gerber SA, Rush J, Stemman O, Kirschner MW, & Gygi SP. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc Natl Acad Sci U S A*, 2003, 100(12):6940-6945.

2. Gerber SA, Kettenbach AN, Rush J, & Gygi SP. The absolute quantification strategy: application to phosphorylation profiling of human separase serine 1126. *Methods Mol Biol*, 2007, 359:71-86.

3. Jiang H & English AM. Quantitative analysis of the yeast proteome by incorporation of isotopically labeled leucine. *J Proteome Res*, 2002, 1(4):345-350.

4. Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, & Mann M. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics*, 2002, 1(5):376-386.

5. Thompson A, Schafer J, Kuhn K, Kienle S, Schwarz J, Schmidt G, Neumann T, Johnstone R, Mohammed AK, & Hamon C. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem*, 2003, 75(8):1895-1904.

6. Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, Khainovski N, Pillai S, Dey S, Daniels S, Purkayastha S, Juhasz P, Martin S, Bartlet-Jones M, He F, Jacobson A, & Pappin DJ. Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics*, 2004, 3(12):1154-1169.

7. Choe L, D'Ascenzo M, Relkin NR, Pappin D, Ross P, Williamson B, Guertin S, Pribil P, & Lee KH. 8-plex quantitation of changes in cerebrospinal fluid protein expression in subjects undergoing intravenous immunoglobulin treatment for Alzheimer's disease. *Proteomics*, 2007, 7(20):3651-3660.

8. Lange VPPDBAR. Selected reaction monitoring for quantitative proteomics: a tutorial. *Molecular Systems Biology*, 2008, 4(1).

9. Wolf-Yadlin A, Hautaniemi S, Lauffenburger DA, & White FM. Multiple reaction monitoring for robust quantitative proteomic analysis of cellular signaling networks. *Proceedings of the National Academy of Sciences*, 2007, 104(14):5860-5865.

10. Burgess MW, Keshishian H, Mani DR, Gillette MA, & Carr SA. Simplified and Efficient Quantification of Low-abundance Proteins at Very High Multiplex via Targeted Mass Spectrometry. *Molecular & Cellular Proteomics*, 2014, 13(4):1137-1149.

11. Peterson AC, Russell JD, Bailey DJ, Westphall MS, & Coon JJ. Parallel Reaction Monitoring for High Resolution and High Mass Accuracy Quantitative, Targeted Proteomics. *Molecular & Cellular Proteomics*, 2012, 11(11):1475-1488.

12. Gallien S, Duriez E, Crone C, Kellmann M, Moehring T, & Domon B. Targeted Proteomic Quantification on Quadrupole-Orbitrap Mass Spectrometer. *Molecular & Cellular Proteomics*, 2012.

13. Gallien S, Bourmaud A, Kim SY, & Domon B. Technical considerations for large-scale parallel reaction monitoring analysis. *Journal of Proteomics*, 2014, 100(0):147-159.

14. Bailey DJ, McDevitt MT, Westphall MS, Pagliarini DJ, & Coon JJ. Intelligent Data Acquisition Blends Targeted and Discovery Methods. *Journal of Proteome Research*, 2014, 13(4):2152-2161.

Chapter 11

Conclusions and Future Directions

Conclusions

Here I have described improvements to peptide and protein identification and quantification by enhancing the ability of ETD to identify peptides and applying the NeuCode technology to analyze samples from cell culture and mammalian samples. Electron transfer dissociation (ETD) will continue to offer researchers an alternative fragmentation method that enables the analysis of labile PTMs and whole protein samples. Time limitations of the ETD scan may prohibit ETD from being the sole fragmentation method that allows the identification of thousands of peptides in one analysis, but it continues to provide means of identifying subsets of peptides that might not be identified by collisional-based methods. NETD, the negative analogue of ETD, is a promising method to analyze peptide anions and fragment them using reagent cations. Recently, our lab has achieved the identification of 5,000 peptide anions in one nLC-MS/MS experiment, signaling a bright future for ETD in the negative mode.

Neutron encoded labels have offered an exciting new quantitative technique to the field and created new excitement for ultra-high resolution mass spectrometry. Here, I presented the use of NeuCode labels to identify peptides and quantify peptides and proteins. Parallel reaction monitoring using NeuCode labels is an intriguing technique that could optimistically enable multiplex quantitation of 3,000 peptide species (1,000 unique sequences) in one targeted experiment. Currently, NeuCode reagents include deuterium, which causes retention time shifts of peptides and necessitates capture of the full elution of both peptides. A multiplexed NeuCode label without deuterium would enable a more robust analysis as quantitation could be performed on data that does where the complete elution profile of peptides was not captured. Lastly, I presented NeuCode labeling of mammalian samples, offering a shorter labeling period and multiplexed analysis of samples. This technique is promising given the popularity of SILAM analysis within the pharmaceutical industry. The future of this technique will be predicated on the availability of instrumentation to enable 4-plex, or potentially 6-plex, analysis of these tissues on a routine basis.

Future Directions of Mass Spectrometry

From the time I first entered graduate school (July 2010) mass spectrometers have made astounding leaps with regards to their performance, improving from roughly 4 MS/MS scans/sec (Hz) to 20 Hz with the newest generation mass spectrometers. Improvements to the mass spectrometers underscored the importance to constantly re-evaluate the methods that we use to analyze biological molecules. The increase in mass spectrometer scan speeds put extra importance on the separation that is performed immediately prior to analysis, making optimal chromatography a vital aspect to mass spectrometry measurements – perhaps more so than ever before. Given this it is clear that if mass spectrometers are going to continue to improve with regards to scan speed, then the chromatography conditions will also need to evolve in order to ensure there will be new peptides to sample with each scan cycle.

The speed of the mass spectrometer is also enabling the sampling of more molecules than ever before. With the newest generation mass spectrometer (Orbitrap Fusion) we can achieve 100,000 MS/MS spectra in a standard 90 minute nLC-MS/MS experiment. Many of these peptides are from low level proteins, but some are low abundance modified forms of peptides. At the American Society for Mass Spectrometry meeting this year one talk centered on discovering the identity of the roughly 50% of species that are sampled, give quality spectra, but are not identified. It has long been thought that these species might arise from unidentified splice events or peptide modifications that are not part of a routine analysis. This does explain some cases, but many of these species are simply adducts with various ions that form during sample preparation or fragments that arise from injection during analysis. Peptide adduct species are interesting because they presumably form as early as cell or tissue lysis, often the first step of proteomic analysis. This suggests that from the moment we begin to analyze the samples we are limiting the number of peptide and proteins we might identify. It is reasonable then, that future workflows might be optimized to limit the formation of these adducts and ensure that each species entering the mass spectrometer can be identified.

The speed of mass spectrometers has enabled collection of data for thousands of proteins in a short period of time; for example, we described the identification of 4,000 yeast proteins in one hour of analysis time. Additionally, quantitative proteomics methods are becoming routinely used and offer the ability to analyze up to 10 samples in one analysis. The ability to generate large amounts of quantitative data, complete with replicates and deep coverage of the proteome, will hopefully make proteomics analysis more attractive to biological researchers. As mass spectrometry becomes more routine in the analysis of biological systems it will be very important to each experiment to strive for two goals: 1) Ensure there is a clear testable hypothesis that can be answered by the mass spectrometry data, and 2) Ensure the quantitative approach is the best fit for the biological experiment design. There is currently a push in the mass spectrometry field to create a quantitative method that can or will be used for all experiments; however, with the multitude of options available it stands to reason that all methods currently employed will have a place in future experiments. This will put onus on researchers to choose the correct quantitative method to best fit the testing of the biological hypothesis (e.g., multiple replicates or conditions), rather than simply creating biological hypotheses that fit the preferred quantitative method.

Lastly, newer generations of mass spectrometers have enabled researchers to envision grander experiments that have not yet been accomplished within the proteomic field. For example, although it might takes many months to years, it is now feasible to analyze thousands of samples or conditions. Two of the most interesting developments in the field currently is the analysis of clinical samples that have been stored from thousands of cancer patients and the analysis of thousands of protein pull downs aimed at elucidating the human interactome. These are experiments that could only be designed with the advent of newer instrumentation and underscore the importance of looking ahead past where current technology can take us to envision the types of experiments that could be possible in the future.