10-30-2020

Investigation of proteins and their modifications using high-resolution mass spectrometry

Justin Roberto
University of Windsor

Investigation of proteins and their modification using high-resolution mass spectrometry

By

Justin Roberto

A Dissertation

Submitted to the Faculty of Graduate Studies

through the Department of Chemistry and Biochemistry

in Partial Fulfillment of the Requirements for

the Degree of Doctor of Philosophy

at the University of Windsor

Windsor, Ontario, Canada

Investigation of proteins and their modifications using high-resolution mass spectrometry

by

Justin Roberto

APPROVED BY:

_____

Z. Zhang, External Examiner
Purdue University

_____

W. Crosby
Department of Biological Sciences

_____

S. Pandey
Department of Chemistry and Biochemistry

_____

S. Ananvoranich
Department of Chemistry and Biochemistry

_____

P. Vacratsis, Advisor
Department of Chemistry and Biochemistry

September 3rd, 2020

DECLARATION OF CO-AUTHORSHIP / PREVIOUS PUBLICATION

I hereby declare that this thesis incorporates material that is result of joint research, as follows:

Chapter 2 of the thesis was co-authored with Katie Sykes, under the supervision of Dr. Vacratsis. The key ideas, experiments, and figures were performed by me. The contribution of the co-author was pilot studies identifying novel phosphorylation sites of TAZ. Chapter 3 of this thesis was authored with Kathy Poulin from the University of Ottawa under the co-supervision of Dr. Park, University of Ottawa, and Dr. Vacratsis, University of Windsor. Kathy was responsible for isolation of exosomes, which I prepared for quantitative mass spectrometry analysis. I analyzed and processed data, creating all figures relating to those measurements. Kathy performed validation experiments from candidate proteins found in our MS identifications. For Chapter 4, in a collaboration with Dr. Don Karl Roberto, a previous doctoral student in Dr. Crosby's group, provided Arabidopsis thaliana samples. I was responsible for sample measurement and assembling figures for the manuscript. Dr. Seth Munholland generated a TAIR10 database removing duplicates within the database, which I subsequently used while processing our MS data. I was responsible for sample clean-up, involved with method development, data processing, and writing and preparation of results and figures.

I am aware of the University of Windsor Senate Policy on Authorship and I certify that I have properly acknowledged the contribution of other researchers to my thesis, and have obtained written permission from each of the co-authors to include the above materials in my thesis.

I certify that, with the above qualification, this thesis, and the research to which it refers, is the product of my own work.

I. Previous Publication

This thesis includes 2 original papers that have been previously submitted for publication in peer reviewed journals, as follows:

Thesis Chapter	Publication title/full citation	Publication status*
Chapter 2	Roberto, J., Sykes, K., and Vacratsis, P. Characterization of phosphopeptide positional isomers on the transcriptional co-activator TAZ. *Biochemistry*.	Submitted
Chapter 3	Roberto, J., Poulin, K. L., Parks, R., and Vacratsis, P. Label-free quantitative proteomic analysis of extracellular vesicles derived from muscular atrophy cells. *Journal of Proteome Research*.	Submitted

I certify that I have obtained a written permission from the copyright owners to include the above published materials in my thesis. I certify that the above material describes work completed during my registration as a graduate student at the University of Windsor.

II. General

I declare that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques,

quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owners to include such materials in my thesis.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee and the Graduate Studies office, and that this thesis has not been submitted for a higher degree to any other University or Institution.

ABSTRACT

Advances in mass spectrometry (MS) has allowed for the deep analysis of various proteomes, providing identifications of proteins and their modifications. The true power in modern-day proteomics is the application of MS techniques to address various biological questions, propelling disease research and biochemical understanding of organisms. We have utilized high-resolution mass spectrometry to investigate biological questions leading to a greater knowledge of cellular biology.

The transcriptional co-activator with PDZ-binding motif (TAZ), is regulated by reversible phosphorylation. However, sequence analysis suggests many potential uncharacterized sites of TAZ phosphorylation, specifically in regions in close proximity to a critical phosphorylation site making site assignment challenging. Using both targeted and untargeted approaches, we identified novel TAZ phosphorylation sites, using a reaction monitoring scheme to resolve positional phosphoisomers, and determined the biological consequence of a novel site, serine 93, on TAZ localization.

Spinal muscular atrophy (SMA) is a motor neuron disease affecting 1 in 10,000 individuals. SMA has been shown to involve the release of extracellular vesicles (EVs), which have been used as a source of biomarkers for disease. We examined the use of EVs as a source of SMA biomarkers. We isolated and quantified >650 proteins from SMA-derived vesicles finding potential biomarkers, one of which was confirmed in patients, suggesting these vesicles coupled with our methods are suitable for SMA biomarker discovery.

In the model plant species *Arabidopsis thaliana* gene expression is heavily regulated through post-translational ubiquitination, however a gap between the

number of ubiquitinated substrates identified and genes encoding the ubiquitin machinery exists, suggesting many unidentified modifications exist. The main strategy for studying ubiquitomes across species uses diglycine enrichment followed by MS analysis. We developed a DIA-based MS method coupled with novel sample preparation methods to overcome plant-specific challenges and increase the repository of the Arabidopsis ubiquitome, identifying 160 proteins with over 400 ubiquitination sites.

The prevalent Charcot-Marie-Tooth disease can be caused by mutations in the lipid phosphatase MTMR2, a protein critical for regulating endosomal dynamics. MTMR2 is regulated by phosphorylation at serine 58. However, the phosphatase and the alterations in protein-protein interactions occurring with this modification have not been thoroughly investigated. To isolate MTMR2 interacting proteins, we utilized *in vivo* labeling fusing BirA biotin ligase to MTMR2, followed by MS analysis, identifying a putative interactor, TSSC1. We also provide evidence that MTMR2 itself may be subjected to phosphorylation-dependent degradation.

This work utilizes high-resolution MS techniques to link protein regulation and function in a variety of biological and cellular contexts. The techniques presented here can be applied to address the gaps of knowledge in various proteomes and are amenable to user-specific modifications. The techniques here provide a framework for determining disease biomarkers for neurological diseases from EVs, investigating proteome-wide changes through protein modifications, and ultimately link high-resolution analytical mass spectrometry techniques and data to address critical biological events in a robust fashion.

DEDICATION

To those who made this work possible

ACKNOWLEDGEMENTS

Firstly, many thanks to Dr. Vacratsis for inviting me into the lab years ago and continually teaching, supporting, and providing unimaginable opportunities. To Dr. Crosby, who served as a committee member and continually afforded me opportunities to study *Arabidopsis* and provide invaluable feedback. To Dr. Ananvoranich for her guidance and feedback throughout all the years dating back to my early scientific work during my undergraduate studies. To Dr. Pandey, for being part of my committee and critiquing our work.

Over the years I have been fortunate to work with amazing scientists and good people in the Vacratsis lab, who will have a lasting impression. To Dr. Christopher Bonham for his display of thoughtful and unrelenting science. Dr. Aaron Steevensz, who used his vast research history to hold my hand through navigating proteomic studies, Dr. Besa Xhabija, for her encouragement and willingness to help throughout the years, Dr. Qiudi Geng, for her training in bacterial work and continual support throughout the years and to Jasmin Nari, for initially training me in the lab. To Jas for all the interesting conversations and help in cloning. To Dr. Robert Gombar, for his aid in data processing and friendship through the years. A very special thanks to Ashley DaDalt, the person who put up with me the most and made a very memorable lab experience. To Dr. Janeen Auld, honorary member of the Vacratsis lab, whose patience, attention to detail, and kind attitude helped propel this work and my knowledge in the MS field.

To my brother Dr. DK Roberto who was an immense help and constant throughout my life in and out of the lab. It was a pleasure working alongside you.

Thank you for continually answering my questions and supporting me. To the other Crosby lab members, Dr. Espanta Jalili, Dr. Moe Dezfulian, and Dr. Seth, for their help with equipment and expertise in the molecular biology of Arabidopsis.

To Dr. Robin Park and all of his members who allowed for our work on SMA and ALS to be possible, especially Kathy Poulin and Charlotte Manser, whose work on exosome isolation and organization was a great help. To Dr. Bonifacino for providing us with TSSC1 constructs. To Dr. Raught for providing us with the BioID vectors, and once again to Dr. Crosby's group for helping us manipulate/work with the vectors.

To my friends and coworkers remaining in Essex who have helped in several ways; Angela, Dave, Scott, Ronaldo, Baturh, Caleb, Chris, Paul, and Michelle. As well as to all my colleagues in the department of Chemistry & Biochemistry. The Boffa, Kochinsky, Ananvoranich, Mutus, Marquardt, Pandey, Taylor, and Tong labs.

Thanks to the 2nd floor Core researchers for the kind atmosphere. To Cody Caba for all the insightful discussions. To my friends in the Marquardt group, Mitch, Mike, Brett, Nicholas, and Stuart. The work I was able to perform would not have been possible without the help of Dr. Lee, Beth, Marlene, Cathy, Jayne, Joe, Bob, Jerry, Norah, Anna, and Ingrid. Thank you to anyone I may have missed in these acknowledgements. Lastly, to my family, especially Lucas, Marco, kuya, and my parents for all of their support and guidance.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS/SYMBOLS

ABA	abscisic acid
ABC	ammonium bicarbonate
AC	alternating current
ACN	acetonitrile
ADC	analog-to-digital converter
amp	ampicillin
amu	atomic mass units
AP	adaptor protein
AP-MS	affinity purification mass spectrometry
AQUA	absolute quantification
ASO	antisense oligonucleotide
Aux/IAA	auxin/indole 3-acetic acid
BioID	proximity-dependent biotin identification
BMDC	bone marrow derived dendritic cell
BPI	base peak intensity
CalyA	calyculin A
CC	coiled-coil
CDC25	cell division cycle 25
CDK	cyclin-dependent kinase
CID	collision-induced dissociation
CMT	Charcot-Marie-Tooth disease
CNS	central nervous system
CSC	collisional-cross section
DAG	diacylglycerol
DC	direct current
DDA	data-dependent acquisition
DENN	differentially expressed in neoplastic vs normal cells
DIA	data-independent acquisition
DLG	discs-large homolog 1
DUSP	dual-specificity phosphatase
DT	drift time
EARP	endosome-associated recycling protein
EE	early endosome

EEA1	early endosomal antigen 1
EGF	epidermal growth factor
EGFR	epidermal growth factor receptor
EIPR	EARP and GARP complex interacting protein 1
EMT	epithelial-to-mesenchymal transition
ERK	extracellular-regulated kinase
ERK1/2	extracellular-regulated kinase 1/2
ESCRT	endosomal sorting complex required for transport
ESE	exonic splicer enhancer
ESI	electrospray ionization
EtBr	ethidium bromide
ETD	electron transmission dissociation
EV	extracellular vesicles
eV	electron volts
FA	formic acid
FDR	false discovery rate
FIG4	factor-induced gene 4
FYVE	fab1 YOTB, Vac1, and EEA1
GARP	Golgi-associated retrograde protein complex
GPCR	G-protein coupled receptor
GPPT	gas-phase protein transfer
GTP	guanosine tri-phosphate
HD	high-definition
HDMS	high-definition mass spectrometry
HE	high energy
Hi3	high three method of label-free quantitation
HILIC	hydrophilic interaction chromatography
hnRNPA2B1	heterogenous ribonucleoprotein A2B1
HOPS	homotypic fusion and protein-sorting
HPLC	high performance liquid chromatography
Hrs	hepatocyte growth factor-regulated tyrosine kinase substrate
Hsc70	heat shock cognate protein 70
ICAM	intracellular adhesion molecule
ICAT	isotope coded affinity rags
IEC	ion-exchange chromatography

IGFBP3	insulin-like growth factor binding protein 3
ILV	intraluminal vesicles
IM-MS	ion mobility mass spectrometry
IMAC	immobilized metal affinity chromatography
IMS	ion mobility separation
INPP4	inositol 4-phosphatase
IP	immunoprecipitation
IP3	inositol 3-phosphate
iTRAQ	isobaric tags for relative and absolute quantitation
kana	kanamycin
keV	kiloelectron volts
LAMP	lysosome-associated membrane protein
LATS	large tumor suppressor kinase
LB	Luria broth
LBPA	lysobisphosphatidic acid
LC-MS/MS	liquid chromatography tandem mass spectrometry
LE	low energy
LFA	lymphocyte function-associated antigen
LPA	lysophosphatidic acid
LS	lucky survivor
m/z	mass to charge ratio
MAGUK	membrane associated guanylate kinase
MALDI	matrix-assisted laser desorption ionization
MAPK	mitogen activated protein kinase
MEK	mitogen-activated protein kinase kinase
miRNA	microRNA
MRM	multiple reaction monitoring
MS	mass spectrometry
MSC	mesenchymal stem cells
MS^{E}/MS^{ALL}	mass spectrometry of everything
MTM	myotubularin
MTMR2	myotubularin-related protein 2
Murr1	mouse USaf1-rs1 region 1
MVB	multivesicular bodies
NDRG1	N-myc downstream-regulated gene-1

Nedd4	Neural precursor cell expressed, developmentally downregulated 4
NFL	neurofilament light chain
NFT3	neurotrophic factor 3
NGS	next-generation sequencing
Nrg1	neuregulin 1
oa	orthogonal acceleration
oa-TOF	orthogonal acceleration-time-of-flight
PAX	paired box
PD	Parkinson's disease
PDZ	PSD-95, Discs-large, ZO-1 binding
PEI	polyethyleneimine
PH-GRAM	pleckstrin homology/glucosyltransferases, Rab-like GTPase activators, and myotubularins
PI3K	phosphatidylinositol-3 kinase
PI3P	phosphatidylinositol-3 phosphate
PGRC	Plant growth regulatory compounds
PKC	protein kinase C
PLC	phospholipase C
PLGS	ProteinLynxGlobalServer
PLST3	Plastin-3
PNS	peripheral nervous system
PP1	protein phosphatase 1
PP2A	protein phosphatase 2A
PPARγ	peroxisome proliferator-activated receptor gamma
PPM	Mg2+- or MN2+-dependent phosphatase
PSD	post-source decay
PSP	PSPs
PTM	posttranslational modification
PTP	protein tyrosine phosphatase
PX	phox homology
Q1	quadrupole 1
Rab	Ras-like protein from rat brain
RF	radio frequency
RME-8	receptor mediated endocytosis-8
RNAseq	RNA sequencing
RP	reversed-phase

RT	retention time
Runx2	Runt-related transcription factor 2
S1P	sphingosine-1 phosphate
Sbf	set-binding factor
SC	Schwann cell
SCF	SKP, CUL1, RBX
SCX	strong cation exchange
SDM	site directed mutagenesis
SDS-PAGE	sodium-dodecyl sulphate-polyacrylamide gel electrophoresis
SET	Suvar3-9, enhancer-of-zeste, thrithorax
SHIP2	SH2-containing inositol 5'-phosphatase
SID	SET interaction domain
SILAC	stable isotope labeling by amino acids in cell culture
SMA	spinal muscular atrophy
SMN	survival motor neuron
SNARE	soluble N-ethylmaleimide sensitive factor attachment protein receptor
snRNP	small ribonucleoprotein
SNX	sorting nexin
SRIG	stacked ring ion guide
SRM	select reaction monitoring
SUMO	small ubiquitin related modifier
SWATH	sequential window acquisition of all theoretical fragment-ion spectra
TAD	transactivation domain
TAE	tris-base, acetic acid, EDTA
TAZ	transcriptional co-activator with PDZ binding motif
TBC-2	Tre-2/Bub2/Cdc16
TCP1	T-complex protein 1
TDA	target-decoy approach
TE Buffer	tris-EDTA
TEAD	transcriptional enhanced associate domain
TEM	tetraspanin-enriched microdomains
TFA	trifluoroacetic acid
TGN	trans-Golgi network
Th	Thompsons (step size)
TIC	total ion chromatogram

TIR1	transport inhibitor response 1
TLR	toll-like receptor
TMT	tandem mass tags
TOF	time-of-flight
TRiC	TCP-ring complex
TSSC1	tumour suppressing subtransferable fragment candidate gene 1
TW	traveling wave
UPLC	ultrahigh-performance liquid chromatography
UPS	ubiquitin proteasome system
Vps	vacuolar protein sorting
WEF	wideband enhancement fragmentation
WW	WW domain
YAP	Yes-associated protein

CHAPTER 1

MASS SPECTROMETRY LITERATURE REVIEW

MASS SPECTROMETRY BACKGROUND

The ability to measure the mass of an analyte and determine its composition has served as an important tool in multiple scientific disciplines. This accurate measurement is primarily achieved through the use of an instrument known as a mass spectrometer. The technique of mass spectrometry (MS) has allowed for the sequencing of proteins, quantitation of subtle changes to proteins and proteomes, as well as the identification of the molecular composition of small molecules. The experiments described in this dissertation heavily relied on MS as an analytical tool to identify proteins and the various post-translational modifications to which the proteins were subjected.

MS strategies for protein identification

MS has been used to study various proteomes providing small molecule analysis, protein identification,[1] protein modifications,[2] quantitation,[3] imaging,[4] and structure elucidation.[5] To enable these measurements, many methodologies have been created. In the field of proteomics, defined studies encompassing complex peaks of proteins, their modifications and the techniques used to analyze them, MS has become an invaluable tool to conduct proteomics research. Two main strategies exist when conducting proteomics workflows: bottom-up and top-down analyses. Top-down proteomics analyzes intact proteins (i.e., without proteolysis),[6] whereas bottom-up approaches utilize proteolytic processing prior to infusion into the mass spectrometer. Bottom-up approaches were utilized in the experiments described in this dissertation (**Figure 1.1**). Bottom-up methods utilize proteolysis by endopeptidases such as trypsin.[7] Digestion of all proteins in a sample generates peptides through cleavage of the peptide bond linking adjacent amino acids.

These peptides can then be analyzed and further processed for fragmentation into their constituent amino acids.

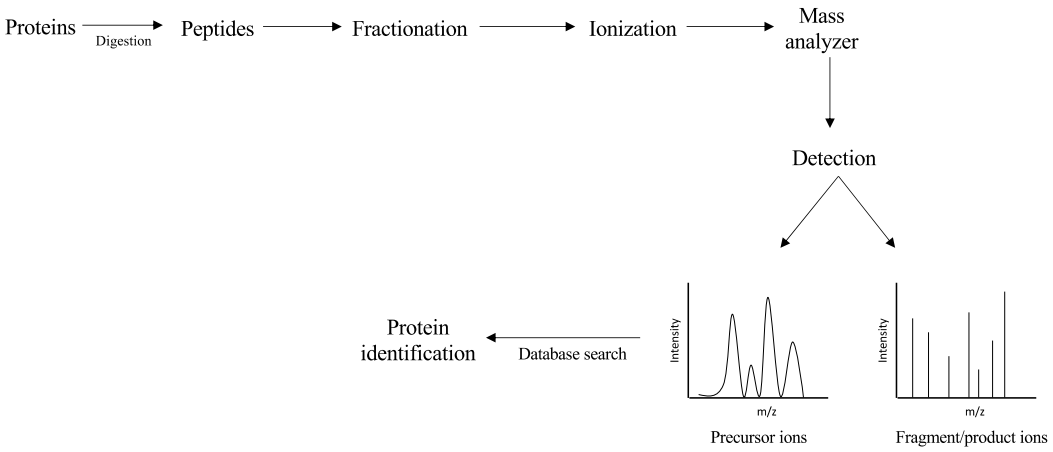


Figure 1.1. General bottom-up proteomics workflow.

Proteins isolated from an ionization source are enzymatically digested into peptides and are typically fractionated, injected into the instrument where peptides are typically fractionated prior to ionization and data measurements. Spectra generated are deconvoluted allowing for peptide and protein identification.

Ionization sources

With the advent of large-scale proteomic studies focusing on both the identification and quantitation of proteins, a wide variety of MS instruments have been developed. The core architecture of every mass spectrometer contains an ionization source, mass analyzer, and a detector. Finding an appropriate ionization source for generating ions from proteins was a hurdle in early proteomics studies. Proteomics endeavours require "soft ionization" techniques which preserve the composition of analytes in a sample. A commonly utilized form is electrospray ionization (ESI), where a sample solution flows through an electrospray chamber through a stainless-steel needle held at a few kilovolts, shaping the distribution of potential and directs the flow of the bath gas. The needle tip charges the surface of liquid as it flows, allowing charges to disperse through Coulomb forces, forming a fine spray of charged droplets.[8] The droplets migrate towards a glass capillary with electric fields. While en route to the end of the capillary, the increase in temperature causes shrinkage of the droplets, making the droplets reach the Rayleigh limit, where the surface charge density increases until the point where the Columbic repulsion equals the surface tension.[8] This phenomenon leads to the "Coulomb explosion," where droplets break apart forming smaller droplets that also evaporate. This explosion continues until the smaller droplets become small enough, have a curvature radius, and carry a surface charge density that leads to ions droplet desorption into gas phase ions. ESI generally forms multiply charged ions, making it a critical technique to study larger proteins and peptides.

The second-most common ionization method utilized for proteomics is matrix-assisted laser desorption ionization (MALDI). MALDI requires liquid peptide samples to form co-crystals a UV-sensitive matrix such as alpha-cyanno-4-hydroxycinnamic acid.

This UV-absorbing matrix aids in protecting the analytes from the laser used to ionize the solid co-crystals. Once co-crystals are allowed to form on spots on target plates, a laser (typically nitrogen) with a wavelength of 337 nm is used.[9] It is postulated that the matrix facilitates soft ionization by allowing an appropriate amount of energy to affect the analyte, as well as allowing chemical reactions to occur between a mixture of excited-state and ground-state molecules from both the matrix and analyte.[10] There is debate as regards the formation of peptide ions through MALDI. There are two prevailing thoughts: gas-phase proton transfer (GPPT),[11,12] and the lucky survivor (LS) hypotheses.[13,14] In GPPT, charged matrix ions are desorbed into the gas phase while analyte molecules are desorbed in an uncharged state. The proton from the matrix ion is then transferred to the neutral analyte, forming a charged analyte ion that can be detected by MS.[14] Under the LS hypothesis, it is believed that different forms of the analyte already exist prior to desorption with a neutral net charge by the presence of counterions. Here, the singly charged peptide or analyte ion will survive the desorption process while other preformed ions will have undergone proton transfer or neutral loss.[14] Both theories share the idea that analyte protonation requires a proton from the matrix.[14]

Mass analyzers

Analyzing the created ions requires navigation through the mass spectrometer's vacuum, guided by various arrangements of electric or magnetic fields generated by mass analyzers. How individual ions react to these forces is based on their mass to charge ratio (*m/z*). Proteomics endeavours commonly utilize time-of-flight (TOF), orbitraps, quadrupoles, and ion trap mass analyzers. TOF mass analyzers are very versatile and amenable to different MS configurations. TOF mass analyzers accurately measure mass by

allowing ions coming from the ion source (mainly generated by MALDI or ESI) to enter the TOF tube and are accelerated through the tube by an applied potential (**Figure 1.2A**). The ions travel down the flight tube according to their *m/z*, with smaller ions reaching the detector first. By rearranging the formula for kinetic energy, the amount of time required for an ion to travel from the ion source to detector, or the flight time is described by $t = L\left(\frac{m}{2E}\right)^{\frac{1}{2}}$, where *m* is the ion mass, *E* is kinetic energy, and *L* is the distance. Many TOF analyzers have an ion mirror at the end of the flight tube, increasing the flight path by reflecting ions back through the flight tube to the detector, correcting for small differences in energy among ions (**Figure 1.2B**).[15] This orthogonal acceleration TOF (oa-TOF) of ions has increased mass accuracy. Since TOF has fairly high mass resolution and has no theoretical limit to large masses, it is the mass analyzer of choice for many hybrid MS instruments.[16] The instrument used in these studies utilizes a TOF mass analyzer with multiple ion mirrors.

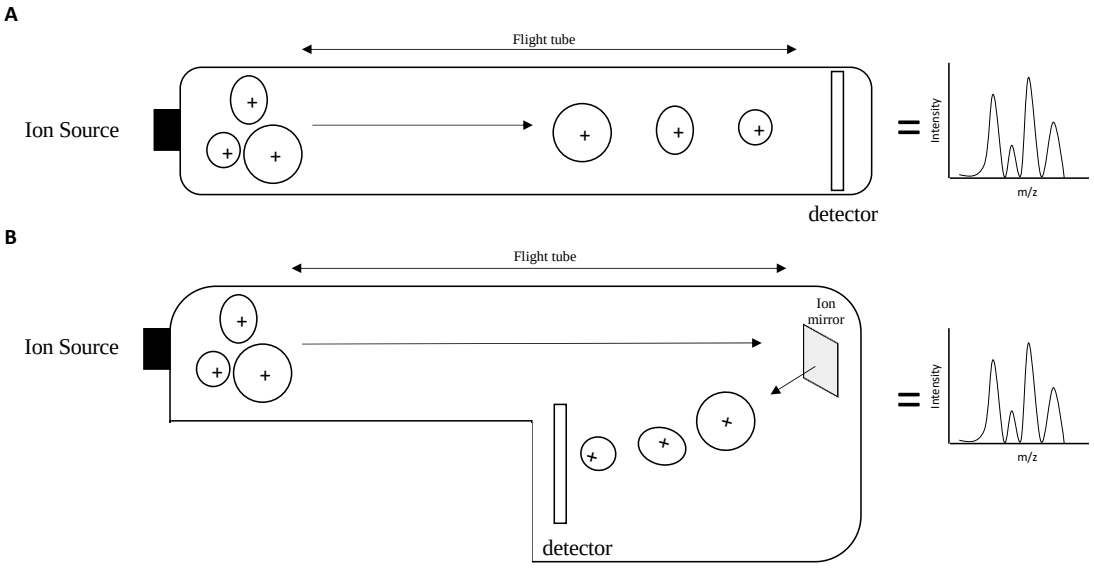


Figure 1.2. Time-of-flight mass analyzer.

Ions created at the source are accelerated through a flight tube separating ions according to their *m/z*. (**A**) Ions accelerate to the detector being separated according to their mass to charge ratio, with ions exhibiting a lower *m/z* travelling faster through the drift tube. Spectra are then generated relating intensity to *m/z*, allowing for detection of peptides/fragment ions of interest. (**B**) Ions generated are reflected by ion mirrors altering the path of the ions prior to reaching the detector. The Ion Mirror reflects ions increasing the ions flight path allowing for better separation of the ions. Once the ions reach the detector, their *m/z* and ion intensity is calculated.

Quadrupole mass analyzer

The quadrupole has been extensively used as a mass analyzer, ion separator, and collision cell, allowing analysis of complex mixtures of ions.[3] The analyzer is composed of four parallel electric rods. Two of the rods opposite of each other are held at a constant direct current (DC) while the remaining two possess an alternating radio frequency (RF) or alternating current (AC) found perpendicular to the rods held at constant DC.[17] This environment forms an electric field, requiring low kinetic energy to allow for ions to pass inside the path or ion beam created by the four electrodes. Ions that are able to traverse the path created by the electric field during voltage alterations are able to pass through the analyzer and be detected, while those that cannot may strike an electrode excluding them from analysis (**Figure 1.3**). Quadrupoles are suitable for the guiding of ions produced by most ionization sources but has principally been used with ESI and MALDI. Ions entering the quadrupole are attracted to the opposite charges of the electrodes. Positively-charged ions entering the quadrupole are stabilized by the positive charges from the DC electrodes, allowing ions with high mass to travel through correctly; however, a negative DC potential occurs at certain voltages, causing positively-charged ions to deflect away from the normal trajectory if the AC does not help guide them, excluding them from analysis as shown in the equations below. RF potentials are able to rescue low mass *m/z* whereas DC maintains higher *m/z*, thereby serving as a mass filter.[17] Therefore, at a certain AC/DC voltage, only a range of *m/z* will be able to travel through the analyzer.

The quadrupole can also act as a collision cell, producing fragment ions from precursors, aiding in peptide sequencing and chemical structure elucidation.[18] Mass spectrometers utilizing the quadrupole as a collision cell have the typical configuration of a triple quadrupole MS ($Q_1q_2Q_3$), where ions are transmitted through three independent quadrupoles, with the first Q1 acting as a mass filter, q2 behaving as a collision cell, and Q3 acting as a mass analyzer. When operating as a collision cell, q2 has an applied RF voltage and contains an inert collision gas, nitrogen or argon.[19] The kinetic energy within the precursor ions and collision with the neutral gas causes fragment ion formation.

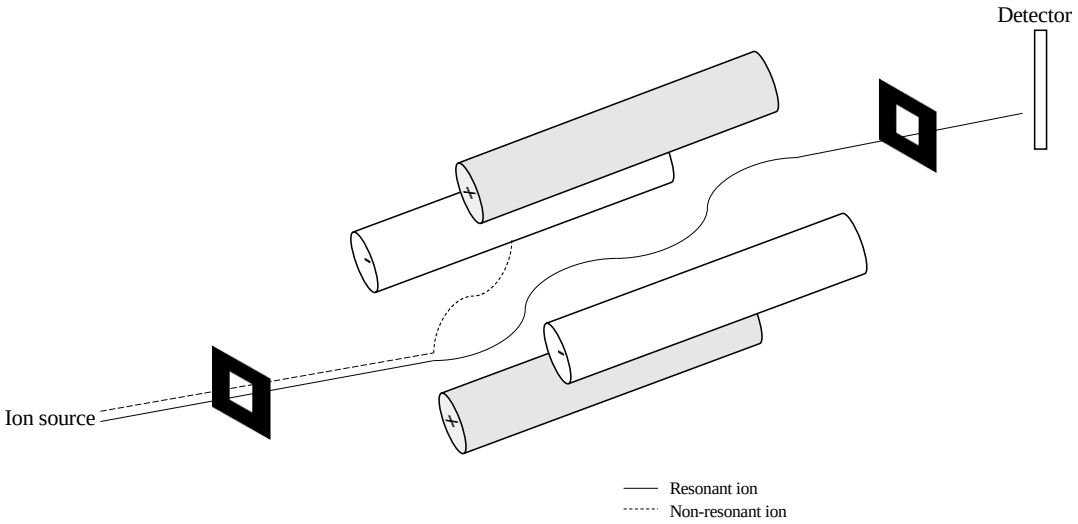


Figure 1.3. The quadrupole mass analyzer.

Ions made at the source are accelerated into the quadrupole mass analyzer composed of electrodes programmed at constant DC and alternating RF set to opposite phases. Ions following the trajectory can reach the detector (solid line), those off the path can collide with an electrode and become discharged, thus failing to reach the detector (dotted line). Ions able to navigate through the quadrupoles can exist and reach the detector where their signals are translated to spectra.

Ion trap mass analyzer

Ion traps describe a group of mass analyzers that utilize electric or magnetic fields to capture ions. A multitude of arrangements of the ion trap exist — linear, segmented ring, and orbital — with the electrostatic trap (known as the "orbitrap") being the most prevalently-used in modern-day proteomics.[20] Most ion traps operate similarly to quadrupoles, with a constant DC voltage while oscillating the RF voltage. Orbitraps differ by having an inner spindle-like electrode and an outer barrel-like electrode.[21] Ions generated from the source are injected into the space between the central and outer electrodes along a tangent, orbiting around the inner spindle, where ions are electrostatically attracted to the electrode. The tangential velocity of ions entering into the orbitrap provide centrifugal force keeping ions from colliding with the inner electrode as the ions experience electrostatic fields producing spiral patterns.[21] Fourier transformation is then used to calculate the oscillation frequencies for ions producing *m/z* values. The first commercial orbitrap was released in 2005 by Thermo Fisher Scientific and has since been modified and coupled with other ion traps and mass analyzers allowing for deeper proteome and lipidome coverage.[22]

Detectors

Ions leaving mass analyzers are detected by coming into contact to a detector. These detectors work to amplify a low signal predominately by using secondary electron multipliers or Faraday detectors. Electron multipliers or microchannel plate detectors release a large number of electrons when hit by an ion and are commonly used with quadrupole-dependent MS analyses.[23] Meanwhile, Faraday cup detectors amplify ions that collide with the metal conductive cup without multiplying electrons, providing a direct

charge detection,[23] and are typically coupled to MALDI ionization prior to analysis.[24] Following these measurements, spectra are acquired and can be interpreted.

Peptide separation

Front-end chromatography techniques have been utilized to decrease the complexity of analytes that enter the MS at one time, and further help separate and purify samples of contaminants that are present during sample preparation. Two prevalent forms of chromatography are strong cation exchange (SCX) and reversed-phase high performance liquid chromatography (RP-HPLC),[25] which act to separate analytes that are hydrophilic or positively charged, respectively. The most widely used fractionation technique is RP-HPLC, however, many have used the two approaches in conjunction with one another.[26] RP-HPLC utilizes columns packed with sorbents that retain more hydrophobic analytes, such as peptides, while removing salts. The mobile phase in RP-HPLC is acetonitrile (ACN) and water typically supplemented with an ion-pairing agent like trifluoroacetic acid (TFA) or formic acid (FA).[27,26] Peptides are retained on the sorbet until the concentration of ACN increases enough allowing peptides to elute off of the column, which is measured as a retention time. RP-HPLC is compatible with ionization techniques used in the MS, allowing differentially retained peptides to enter the ionization source of the mass spectrometer.

MS1 Mass fingerprint

When utilizing protein digests, the ions detected produce a spectrum of percent intensity as a function of *m/z* providing peptide mass fingerprints of proteins.[28] While early work heavily relied on this information for protein identification, these fingerprints are unable to distinguish between homologous peptides arising from conserved domains and

motifs comprising proteins. In order to derive sequence information, these molecular or parent ions, undergo further fragmentation via tandem MS (MS/MS).

MS/MS and peptide fragmentation

In order to generate amino acid sequence information during bottom-up proteomics screens, secondary fragment or product ions need to be generated from precursor ions. Various methods have been developed allowing for the creation of fragment ions from peptides during analysis such as electron transfer dissociation, post-source decay, collision-induced dissociation and higher energy collision dissociation. One of the simplest methods of precursor ion fragmentation is through post-source decay (PSD), where the metastable ions generated at the ion source have sufficient internal energy allowing for spontaneous fragmentation.[29] Here, metastable ions coming from the source are able to decompose when entering the TOF mass analyzer flight tube.[29] Often, the internal energy of a molecular ion is insufficient for fragmentation, thus complicating objectives such as protein or peptide sequencing. To overcome this challenge, many methods "activate" ions to increase their internal energy, thus allowing for fragmentation to occur. The most prevalent fragmentation employed is collision-induced dissociation (CID).[30]

CID provide electrons with kinetic energy, where small ions collide with neutral gas, transferring the relative kinetic energy of the collision partners to internal energy of the ion.[31] In order for fragmentation to occur during the experiment, an amount of energy exceeding a threshold must be delivered. Ions needing low-energy collisions (several hundred daltons) need require 1–110 eV of collision energy, while high-energy collisions are in the keV range.[29] Certain instruments operate using either high- or low-energy collision energies, but all instruments have collision cells found between the ion source

and the mass analyzer. Collision cells are filled with neutral gasses, usually nitrogen (N_2) or argon (Ar_2) and are held at a low pressure to promote collision between ions and the neutral gas.[29] High-energy (HE) CID increases the amount of bond cleavage in comparison to low-energy (LE) CID. HE CID of peptides can also generate side-chain cleavages, immonium ions, and internal fragments.[32]

The ability of some MS instruments to cycle through low- and high-energy collisions has aided in precursor and product ion scanning, respectively, utilized in triple quadrupoles and qTOF systems.[18]

CID of charged precursors form fragment ions - ideally *y*- and *b*-ions formed by dissociation at the peptide bond (the N-C amide bond). The collection of *y*- and *b*-ions aids in the sequence determination of peptides within a sample (**Figure 1.4**). The collection of these ions allows for peptide sequencing, with *b*-ions containing the N-terminal amino acids, and *y*-ions containing the C-terminal amino acids.[33]

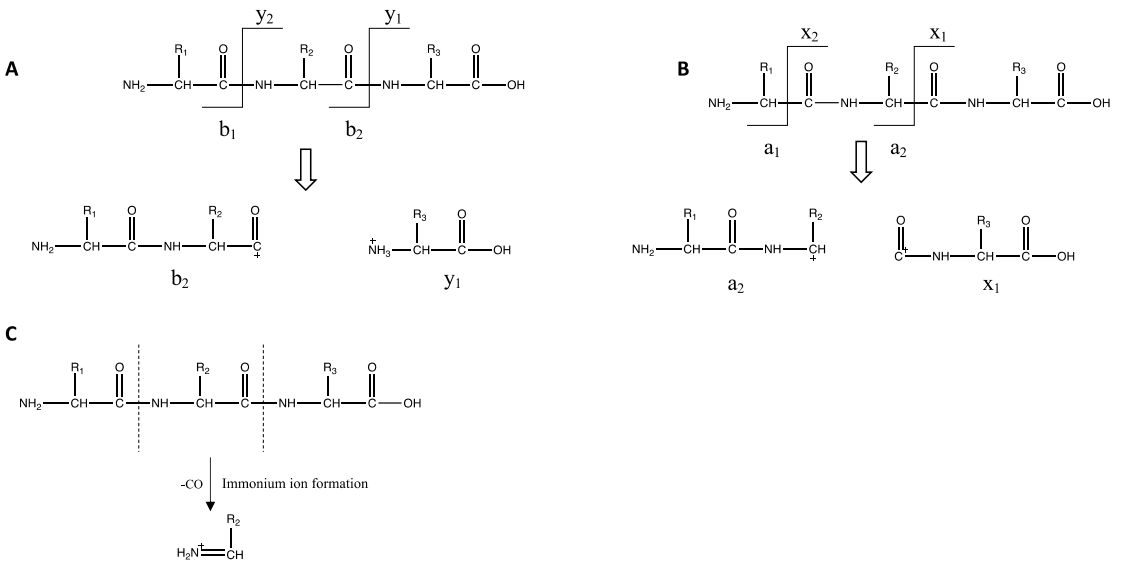


Figure 1.4 Peptide fragmentation.

Intact peptides undergo collision-induced dissociation allowing for peptide sequencing. (**A**) Ideally *b* ions and *y* ions are formed. *b* ions possess amino acids belonging to the N-terminal portion of the peptide while *y* ions possess C-terminal peptide sequences. Fragmentation of peptides not about the peptide bond. Peptides can dissociate into more than *b* or y ions depending on where bond breakage occurs. Similar to *b* and *y* ions, are (**B**) *a* and *x* ions, however fragmentation is not about the peptide bond.

(**C**) Immonium ions are internal fragments formed *a*-or *y*-type ions possessing the amino acid side chain amino group.

(Targeted and untargeted approaches) Data acquisition methods

Advancements in proteomics technologies have allowed for more sensitive and higher resolution analysis of complex samples. As such, alternative types of data acquisition methods were developed in attempts to perform identification. The first method relied on data-dependent acquisition (DDA) in which the top ions — based on spectra obtained during MS1 scans — are selected for further fragmentation (MS/MS) allowing for analysis of their product ions.[2] This technique has aided in the analyzing of diverse sets of proteomes and is robust enough to perform quantitation and identify PTMs.[2,34] DDA has allowed for the specific monitoring of precursor ions and their fragment ions, with select reaction monitoring (SRM) being developed for quantitation.[35] SRM is typically performed using triple quadrupole instruments, where the first and third quadrupole act as mass filters, selecting for predetermined *m/z* values belonging to a certain peptide ion and fragment created from this peptide. The fragment ions are created using q2 as a collision cell, allowing the transition from precursor to fragment ions. These transitions are monitored over a time interval, creating chromatograms recording retention time and signal intensity for specific transitions.[36] SRM can be used to monitor multiple product ion formation from a precursor or monitoring transitions from multiple precursors to fragments, known as multiple reaction monitoring (MRM).[37] The use of quadrupoles as mass filters allows the specific isolation of known ions, allowing researchers to focus on selected phenomena such as PTMs.[38] A drawback of these techniques is the requirement for *a priori* knowledge of fragment ions that are reproducibly detected from each selected precursor in order to monitor transitions for analysis. Furthermore, the most abundant ions are biased for further fragmentation, leading to a reduced or under-sampling of the less abundant ions within a

complex mixture. To address these shortcomings of DDA, data-independent acquisition (DIA) techniques have been developed.

In DIA studies, analyte detection requires no knowledge of precursor ions to monitor the generated fragments. Early work in this field focussed on analyzing all precursors and fragment ions generated within small *m/z* ranges or isolation windows. However, these unbiased approaches have inherent difficulties in linking the correct precursor to fragment ion(s). A significant advancement in DIA data analysis was made by the introduction of sequential window acquisition of all theoretical fragment-ion spectra (SWATH MS), referring to the isolation windows acquired for a given precursor mass range across a chromatography run.[39] In this technique, all time-resolved analytes within a certain *m/z* range precursor range and retention time window are detected.[39] A challenge with DIA is the complexity of the data obtained since there is little association between precursor and fragment ions in a situation where all ions are fragmented. Early methods of interpreting DIA spectra utilized DDA-like database searching,[40] where search algorithms select the list of possible fragment ion matches using known precursor masses or by searching pseudo MS2 spectra created from coelution of intact precursors and fragment ions peaks.[41] Data can be extracted using targeted data extraction strategies where fragment ion maps are scanned for peptides of interest, thus helping deconvolute data and aiding in fragment ion sequence assignment.[39]

In 2006 Plumb and colleagues described a new LC/MS data acquisition approach, known as MS^E (MS of everything) or MS^{ALL}.[42] For these analyses, the collision cell is configured to alternate between low- and high-collision energy, thereby generating precursor and fragment ion data, respectively. Used on $Q_1q_2Q_3$ and qTOF systems,[43] this

technology differs from the previously described SWATH MS in the energy ramp, where there is a simultaneous high and low energy cycling in MS^E. Although both methods utilize Q1 to allow ions to pass. A key difference between the two techniques is the *m/z* range or window that is examined, with most SWATH measurements occurring between 20- and 25-Da (~100 millisec) windows across the whole *m/z* range, allowing for a smaller mass range to reach the collision cell and generate their respective fragment ions.[44] The isolation windows used in SWATH are analyzed in more narrow mass ranges in comparison to MS^E, where advanced front-end liquid chromatography decreases sample complexity for any instant, allowing measurement of all ions generated, or the full *m/z* range is constantly being measured,[45] not sequential mass windows. Data analysis for MS^E heavily relies on correlating chromatographic separation and product ion detection to deconvolute the data acquired.[46] The use of ion mobility gas phase separation has been proposed to aid in relating the fragmentation of precursor ions with their product ions, thereby aiding in the deconvolution of data analysis,[47] although this approach, requires a higher sample concentration input.

Ion mobility

Ion mobility allows for better separation and transmission of coeluting peptides and their fragment ions by utilizing the shape and charge of ions in the collision cells of hybrid mass spectrometers, such as the qTOF.[48] Ion mobility MS (IM-MS) achieves separation based on ions colliding with a neutral gas or drift gas like nitrogen or helium under an electric field. This approach retards ion mobility based on various criteria such as the ion's *m/z* and the collision cross section (CCS),[49] resulting in a unique drift time (DT) for each ion (**Figure 1.5**). The shape, polarity, and CCS of an ion under investigation describes its

DT.[50] DT calculations can be made by measuring the migration of ions in a drift tube[51] or travelling wave in an RF ion guide.[51,52] The traveling wave IM (TWIM) method uses stacked ring ion guides arranged orthogonally to the transmission axis with RF voltages applied to adjacent electrodes.[52] Ions traverse these waves differentially, with ions exhibiting high mobility slipping behind the waves and are able to pass through IMS devices more rapidly than low mobility ions. In both devices, ions with a more compact shape are able to traverse relatively quickly in comparison to bulkier ions. IM-MS has helped improve peptide identification, quantitation,[53,54,55] and PTM investigation.[22,56,57] This added layer of separation allows for high-definition MS (HDMS) measurements.

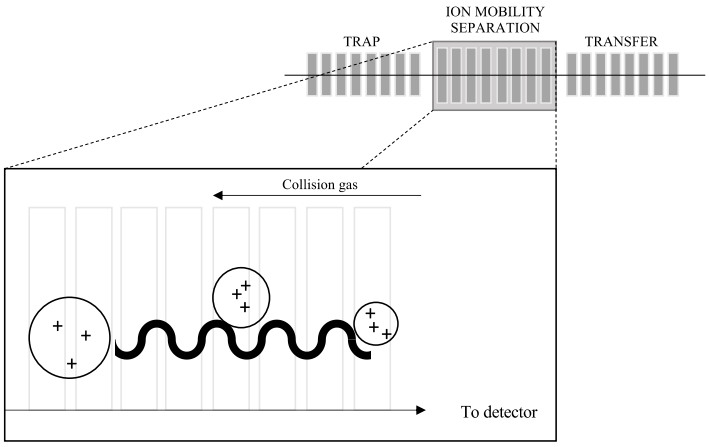


Figure 1.5. Ion mobility separation or Triwave ion guide.

For ion mobility mass spectrometry (IM-MS) using the Synapt G2-Si, the IMS cell composed of RF ion guides allowing for the separation of ions through traveling wave. The Trap and transfer T-waves be operated as collision cells or ion guides. A collision gas, typically nitrogen or helium is used in the IMS to make contact with the ions in the ion packet. The IMS cell (box), isobaric ions with different shapes exhibit different mobilities on the travelling wave generated by the RF ion guide, thus allowing for separation of the isobaric ions, ultimately separating ions of identical chemical composition and providing high-resolution ion data.

Database searching

Data obtained during complex proteomic measurements require deconvolution to provide meaningful data, thus highlighting the critical role that database searching plays in proteomics workflows. Early proteomics endeavours required manual *de novo* sequencing, matching fragment ions to a series *in silico* digests of the hypothetical protein/peptides under investigation. These workflows are amenable in investigating lowly complex samples; however, with increased demands of protein identification and the advent of more complex instrumentation, computational algorithms to deconvolute data were required. Various software and programs employed different strategies in attempts to identify proteins from complex spectra employing strategies including multistep database searches,[58] spectral matching,[59] sequence tagging, and spectral alignment.

When performing tandem MS analysis, generated spectra provide amino acid sequences which can be searched against protein databases available online, such as UniProt or NCBI. Such database searches serve to match tandem MS spectra to precursor ions. Comparative scoring of the ions measured during MS/MS selects which precursors and fragment ions belong together.[59] The genesis of spectral libraries served as a tool for researchers to upload their findings and compare them to previously identified proteins. The peptide sequence tag approach takes advantage of the fragment and product ions formed during MS analysis.[60] This method scans for unique sets of amino acids specific for a peptide as identifiers of the peptide of interest, allowing its identification as a component of the larger proteome.[61] A partial sequence as well as the N-terminal prior to and C-terminal masses following that sequence are necessary.[60] Spectral alignment has been utilized to study sites of PTMs where tandem MS are aligned with protein sequences

with PTMs residing at specific amino acid residues.[62] Since amino acids in peptides modified with a PTM involve a specific mass shift, spectral alignment identifies these mass shifts and highlights regions that do not match the predicted modifications thereby decreasing false positive identification.

Following database searches, different parameters are measured.[63] In order to describe the quality of data obtained for each protein. One such parameter is the false discovery rate (FDR), used to ensure protein identifications obtained from MS experiments are acceptable within a given range. This becomes a challenge when identifying low-intensity ions in a spectrum.[64] To ensure false positives are not being identified, the target-decoy approach (TDA) was created,[64] in which a decoy database is created and used to analyze the data collected. Any target matches using the decoy database approximates the numbers of false positives that can occur, and are used to estimate the FDR.[65] Decoy databases can be created by reversing or switching the amino-carboxyl orientation of a protein or shuffling protein sequences, where amino acids of each target protein are randomly rearranged.[66,67] These methods allow decoy and target databases to have similar composition — having similar length proteins, amino acid distributions, proteins, and peptides — but minimizing the number of common peptide sequences.[64,68] Both the target and the decoy databases are searched, where PSMs coming from the target database should reflect higher scores, whereas lower scores likely are likely attributed to decoy matches. By choosing a specific threshold, the number of false discoveries is elucidated. Database searching has provided protein identifications, but more work is needed improve database mining and curation, including the construction of custom databases that allows the examination of sample-specific databases.[69]

QUANTITATIVE MASS SPECTROMETRY

Modern day proteomics allows for both protein identification as well as measurement of protein abundance. This added measurement deepens proteomics research by providing a biological connection allowing for more detailed and informative proteomic data, allowing for the quantitation of protein abundance changes in disease states or following specific treatments. However, most data generated via MS analysis is not appropriate for quantitation. As such, numerous methodologies for quantitation have been devised ranging from expression of mass tags, using differentially-labeled isotopes, and addition of reference peptides. These methods have allowed for relative and absolute quantitation of proteins in a mixture, where relative quantitation measures abundance changes between different conditions,[70] whereas absolute quantitation provides a concentration of protein within the sample. Use of these different methods has afforded MS analysis to be a powerful discovery and analytical tool for a multitude of workflows.

Labeling strategies for protein quantitation

The first quantitation methods utilized chemically labeling proteins in complex mixtures. There are various forms of labeling available using both *in vivo* and *in vitro* methods. A prominent *in vitro* chemical labeling technique utilizes isobaric coded affinity tags (ICAT), where chemical tags are synthesized which interact with proteins isolated from two sources allowing for comparison between the two groups. ICAT experiments use tags that possess an isotopically coded linker, a specific chemical reactivity, and an affinity tag.[71] Here, proteins and the ICAT reagent interact via chemical reactivity. Typically with a thiol-specific reactive group of ICAT and a reduced cysteine in the proteins. The isotopically coded linker can have either heavy or light isotopes, allowing for labeling of

two different cellular populations. Following labeling, the proteins from each population are combined and enzymatically digested. The ICAT modified peptides are then enriched using an affinity tag, after which samples are injected into the mass spectrometer and subjected to LC-MS/MS analysis. Peptide sequences are recorded and quantified by measuring the relative signal intensities coming from labeled peptides found in both populations, differing by the mass found on the isotopic linker (i.e., heavy or light), ultimately allowing for a direct comparison of tagged peptides found in both populations.[72] ICAT technology has spawned numerous other labeling approaches, such as isobaric tags for relative and absolute quantitation (iTRAQ) and tandem mass tags (TMT).[73,74] iTRAQ labeling modifies primary amino groups with iTRAQ reagent. Isotopes are then detected following collision-induced dissociation where unique mass signatures remain for each condition. Similarily, TMT technology labels the N-termini of peptides via a reactive group and can facilitate labelling of many conditions simultaneously. This technique has been used to measure PTMs including ubiquitination in complex systems.[75] A second form of isotope labeling for quantitation using deuterated amino acids is the *in vivo* labeling technique SILAC (stable isotope labeling by amino acids in cell culture). Rather than incorporating a tag as in ICAT or iTRAQ, SILAC analysis requires growing cells under two different conditions using deuterated amino acids such as leucine, methionine, or lysine in the growth media.[76] This allows for proteins to be synthesized with heavy or light amino acids, facilitating the comparison between different cellular conditions.[76] Following growth in different isotopic media, proteins are combined and digested, thus reducing handling variation and allowing for simultaneous quantitation via MS analysis. SILAC theoretically allows for a wider proteome coverage, as the essential amino acids are

isotopically labeled, removing the bias of a probe having to bind to cysteine residues like in ICAT or iTRAQ, as well as allowing direct comparison between samples.

Label-free quantitation

The ability to perform protein quantitation without chemical tags avoids biases of differential tag binding by select targets. Further, the different amino acid compositions of proteins may lead to different isotope incorporation when using the SILAC approach.[77] A further difficulty with SILAC is that some studies are not amenable to culturing cells or measuring PTMs making quantitation challenging.[73] To alleviate the issues of PTM quantitation and detection, reaction monitoring approaches (like SRM) were developed.[78] To make the system more quantitative, absolute quantification (AQUA) was developed.[35] Here, prior knowledge of the peptides within a sample is required to allow for the production of synthetic peptides that are isotopically labeled. The synthetic peptide acts as a standard that can be spiked into the sample under analysis, allowing for calculation of the analyte of interest through comparison of the signal intensities from the known standard and the peptides/proteins of interest. This method has been utilized to quantitate PTMs such as phosphorylation and ubiquitination.[35] DIA has also been used to perform relative protein quantitation. Silva, et al., have developed the Hi3 method, where a reference set of peptides is spiked into the sample and the most abundant peptides from the reference are compared to the top three peptides for a given analyte in the sample.[55] This method has aided in the label-free quantitation of *E. coli* lysates to other model organisms.[34,55,79] The Hi3 method has also been coupled to high-definition mass spectrometry (HDMS) analysis utilizing the DIA method of MS^E coupled with ion mobility separation. Here, precursor ions formed from a complex mixture are further separated in an ion mobility cell, allowing

for an added layer of separation and accurate assignment of precursor or product ions.[80] Following addition of Hi3 peptides into the mixture, quantitative HDMS^E is required for a deep range of proteome quantitation.

POST-TRANSLATIONAL MODIFICATION SCANNING

For many years the definition of the "proteome" would simply describe all the proteins found within an organism; however, the definition is well-suited to expand and include the modifications that occur on these proteins, which often dictate their structure, activity, and function. Since these modifications, such as phosphorylation, ubiquitination, acetylation, and glycosylation are often critical to protein function, many techniques have been employed to isolate modified proteins followed by MS analysis, or to perform PTM identification as part of the MS analysis.

PTM searches present their own challenges as the complexity of the proteome exponentially increases, and they are often in low stoichiometry compared to their unmodified forms. Two approaches are commonly used: the first involves restricted modification searches where there is knowledge of the PTMs present, allowing computational programs to calculate possible combinations of modified theoretical spectra. In contrast, unrestricted modification searches attempt to identify PTMs in the absence of assumptions.[1] Phosphorylation is one prevalent PTM, which involves the catalytic addition of a phosphate group, by enzymes called kinases, onto serine, threonine, or tyrosine residues found within proteins, adding 79.99 Da to the modified protein, as well as on other macromolecules such as lipids and DNA. As phosphorylation is one of the most abundant cellularly-occurring PTMs, many methods — both on-line and off-line of the MS — have been developed towards detecting these changes. A second PTM that adds 79.98 Da to tyrosine residues is sulfation,[81] further highlighting the importance of the availability of modification-specific detection techniques to distinguish PTMs present.

Off-line phosphorylation enrichment strategies prior to MS analysis: Immobilized metal affinity chromatography (IMAC) enriches for phosphopeptides, taking advantage of the affinity between the negatively-charged phosphate groups with positively-charged metal ions like Al^{3+} and Ti^{4+} and can be used in combination with chromatographic separation like hydrophilic interaction chromatography (HILIC) and ion-exchange chromatography (IEC).[82,83] A beneficial ability of these chromatography techniques is on-line coupling to the mass spectrometer, reducing sample handling.

Phosphorylation detection methods utilizing MS hardware and software include electron capture dissociation, electron transfer dissociation, and neutral loss.[84] Phosphoric acid (H_3PO_4), ~98 Da, is lost from phosphoserine/threonine residues upon CID based on beta-elimination or E2 eliminations.[85] Another phenomena routinely seen is the loss of 80 Da in the form of meta-phosphoric acid (HPO_3) from phosphoserine/threonine residues.[85] When identifying precursor and fragment ions, these losses suggest a peptide is phosphorylated and, depending on the composition of the peptide, a specific phosphorylation site can be reaffirmed through mutagenesis studies. As many peptides possess multiple phosphorylation sites, identifying the specific sites modified under specific conditions is paramount. These neutral loss scanning approaches on their own do not provide sufficient enough information for site mapping which typically requires a targeted reaction monitoring as previously discussed.

REVERSIBLE PHOPSHORYLATION OF PROTEINS AND LIPIDS

Phosphorylation of proteins and lipids has proven to be critical in regulating cellular homeostasis. As such, enzymes regulating these processes have been of significant interest to the scientific community, one such class being the heavily-studied being through reversible phosphorylation catalyzed by the protein tyrosine phosphatase (PTP) superfamily of enzymes capable of dephosphorylating a plethora of targets. A subfamily within the large PTP superfamily is the dual-specificity phosphatases (DSPs).[86] The myotubularins (MTMs) comprise an additional subset of proteins within this subfamily that dephosphorylate phospholipids proteins are expressed 15 proteins are expressed in humans. The Myotubularin-related Protein 2 (MTMR2) is a member which regulates the phosphoinositide phosphatidylinositol-3 phosphate (PI3P) levels and myelination in the peripheral nervous system (PNS).

Phosphoinositides - substrates of lipid phosphatases

Seven different forms of phosphoinositides exist in cells, generated by specific kinases and phosphatases that are activated by specific signals altering phosphorylation of phosphatidylinositol (PI) at the 3,4, or 5 positions of the inositol head group (**Figure 1.6**). PI can form phosphatidylinositol 3-phosphate (PI3P) through phosphorylation at the D3 position of PI primarily by class III phosphatidylinostiol 3-kinases (PI3K), Vps34.[94] PI(3)P can then be dephosphorylated by active members of the myotubularin family of phosphatases, regenerating PI, or further phosphorylated to form different phosphoinositides. Phosphorylation by PI(3)P 5-kinase (PI(3)P5K) produces phosphatidylinositol 3,5-bisphosphate ($PI(3,5)P_2$), which can further be processed to form $PI(3,4,5)P_3$, discussed below. Alternatively, $PI(3,5)P_2$ can be dephosphorylated by

$PI(3,5)P_2$ 5-phosphatases, such as FIG4, a gene mutated in CMT subtypes.[95] Phosphatidylinositol 3,4-bisphosphate ($PI(3,4)P_2$) can also be formed by phosphorylation of PI(3)P by PI(3)P 4-kinases (PI(3P)4K); enzymes catalyzing the reverse reaction have not been reported thus far. $PI(3,4)P_2$ is found in low amounts generated through type I phosphatidylinositol 3-kinases (PI3Ks), phosphorylating PI(4)P at the D3-position. Production of $PI(3,4)P_2$ can also occur through dephosphorylation of PIP_3 by the 5-phophatase SH2-domain containing inositol polyphosphate 5-phosphatase (SHIP), dephosphorylating PIP_3 at the 5-positon of the inositol head group at endocytic vesicles and lamellipodia.[96]

PI can form phosphatidylinositol 4-phosphate (PI4P) through the action of phosphatidylinositol 4-kinase (PI4K), which can be reversed by PI 4-phosphatases, such as inositol polyphosphate 4-phosphatase (INPP4). PI4P can be further phosphorylated to form $PI(3,4)P_2$ and $PI(3,4,5,)P_3$ through the sequential action of PI3K and $PI(3,4)P_2$ 5-kinase ($PI(3,4)P_25K$), respectively. Phosphatidylinositol 4,5-bisphosphate ($PI(4,5)P_2$) is the most abundant phosphoinositide in cells and is synthesized primarily through phosphorylation of PI by PI4-kinases (PI4K), generating PI4P, and is then further phosphorylated by PI(4)P 5-kinase (PI(4)P5K), creating $PI(4,5)P_2$.[90,97] Biologically, $PI(4,5)P_2$ is thought to be involved in endosome, lysosome, autolysosome, Golgi, endoplasmic reticulum, and nuclear processes and promotes intraluminal sorting of epidermal growth factor receptor (EGFR), underscoring the significance of this phosphoinositide in endolysosomal trafficking.[98] $PI(4,5)P_2$ is a substrate for phospholipase C (PLC) producing the second messengers diacylglycerol (DAG) and inositol triphosphate (IP_3) as a response to G protein-coupled receptor or tyrosine kinase receptor activation.

Phosphatidylinostiol 5-phosphate (PI5P), one of the more recently identified phosphoinositides, localizing to the endoplasmic reticulum, Golgi apparatus, nucleus, and plasma membrane, where it facilitates pathogen invasion to hosts and as a platform for T cell signaling.[99,100,101] PI(5P) can be formed through phosphorylation of PI by phosphatidylinositol 5-kinases (PI5K), such as PIKfyve,[101] the dephosphorylation of $PI(4,5)P_2$ by inositol 4-phosphate phosphatase (INPP4) or dephosphorylation of $PI(3,5)P_2$ through myotubularin phosphatases such as MTM1 or MTMR2 as previously mentioned.[102]

Phosphatidylinostiol 3,4,5-triphosphate ($PI(3,4,5)P_3$ or PIP_3) is one of the most important lipids as second messengers in the cell, serving as a platform for the binding of a multitude of proteins to the plasma membrane allowing for different cellular events to occur, such as phagocytosis, pinocytosis, regulated exocytosis, and cytoskeletal organization by coupling signaling to actin polymerization.[103] Cleavage of PIP_3 by phospholipases forms inositol 1,4,5-trisphosphate (IP_3) in turn causing the release of intracellular calcium.[103] Pathways of PIP_3 production that have been validated include phosphorylation of PI(3,4)P2 by $PI(3,4)P_2$ 5-kinases, or through phosphorylation of $PI(4,5)P_2$ by PI3K. PIP_3 can be dephosphorylated at the D3 position by the phosphatase and tensin homolog (PTEN), a lipid phosphatase resembling the MTMs.[103]

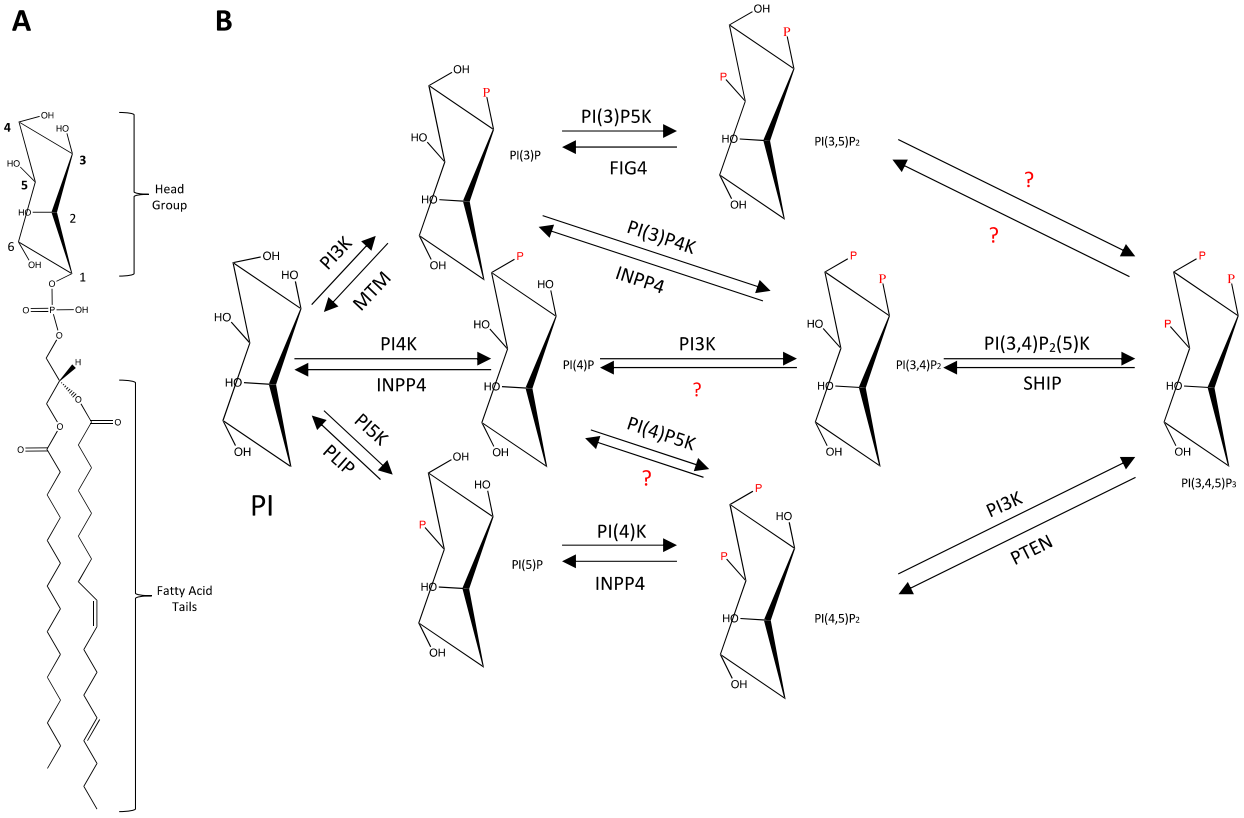


Figure 1.6. Phosphoinositide structure and phosphoisoform synthesis.

(**A**) Structure of phosphatidylinositol, being composed of two fatty acid tails and an inositol group with the potential to be phosphorylated at positions 3, 4, and 5.

(**B**) Generation of the 7 phosphoforms of phosphoinositides through the differential phosphorylation of phosphatidylinositol (PI). PI3K, PI4K, or PI5K generating PI(3)P, PI(4)P, or PI(5)P, respectively. PI3P can be reversed to PI through the action of MTMs, or can be further phosphorylated by PI3P5K, generating $PI(3,5)P_2$, with the reverse reaction being catalyzed by $PI(3,5)P_2$ phosphatases like FIG4. PI(4)P can be converted to PI through dephosphorylation via INPP4; otherwise, PI(4)P can be dual phosphorylated through phosphorylation by PI3K, generating $PI(3,4)P_2$, and can be further phosphorylated by $PI(3,4)P_25K$, generating PIP_3. PIP_3 can be converted to $PI(3,4)P_2$ via PTEN,

dephosphorylating the 3-position of PIP_3. PI(5)P can be phosphorylated by PI(4)K to generate $PI(4,5)P_2$ which serves as a substrate for PI3K to synthesize PIP_3. P = phosphate group; "?" = unknown kinase or phosphatases; PLIP = phospholipid-inositol phosphatase; MTM = myotubularin; inositol 4-phosphate phosphatase = INPP4; SHIP = SH2-domain containing inositol polyphosphate 5-phosphatase.

PTP catalytic mechanism

MTMR2, like all active members of the PTP superfamily of enzymes, possess a CX_5R motif within the 'P-loop' in the catalytic domain.[87] The active site of PTPs is found within a cleft of varying depths, with MTMR2's found to be similar size to PTP1B, though wider to compensate for the headgroup of phosphoinositides.[88,89] Phosphoinositides are a group of lipids that are a component of cellular membranes located on the cytoplasmic leaflet which regulate the trafficking of proteins and other macromolecules in cells.[90,91] They are composed of a glycerol backbone esterified with two fatty acid chains and a phosphate group bound to a polar head group. The head group, polyol *myo*-inositol, is subject to various phosphorylation events at the D3, D4, and D5 positions by lipid kinases (**Figure 1.6A**). MTMR2's arginine residue, Arg^{423}, within the P-loop, helps stabilize the negatively charged phosphate moiety. While PTPs employ the 'WPD' loop, in which an aspartic acid residue (Asp) acts as a general acid/base during dephosphorylation, MTMR2 uses an Asp in close proximity to the catalytic Cys (Asp^{422}), which is critical for MTMR2 activity.[88,92] The active site cysteine, Cys^{417}, acts as a nucleophile, and the phosphotyrosine substrate is coordinated by main-chain amides of the PTP loop and Arg^{423}. Typically, the Asp in the WPD loop moves towards the substrate where it acts as a general acid to donate protons to the phenolic oxygen of the substrate, while the catalytic Cys attacks the phosphorous atom of the substrate, forming a thiophosphate intermediate (**Figure 1.7**).[93] MTMR2 possess an asparagine residue (Asn, Asn^{355}) at the expected position of the WPD loop, which is unable to participate as a general acid/base. MTMR2 uses Asp^{422} found within the CX_5R motif to perform this general acid/base activity.[88,89] The intermediate dissociates when an activated water molecule enters aided by an Asp in the WPD and an

adjacent Gln residue, forming a dephosphorylated substrate, free phosphate, and regenerated PTP thiolate ion.[93] As lipids and proteins found at the cell membrane allow for rapid cellular response and structural integrity, the activity of kinases and phosphatases—which generate different phosphoinositides—is heavily examined to understand their role in the biogenesis and maturation of vesicular structures known as endosomes, where MTMR2 has been shown to remove the 3-phosphate from PI3P and PI(3,5)P_2, altering the endosomal landscape. As such regulation of these lipids through reversible phosphorylation is critical for cellular homeostasis.

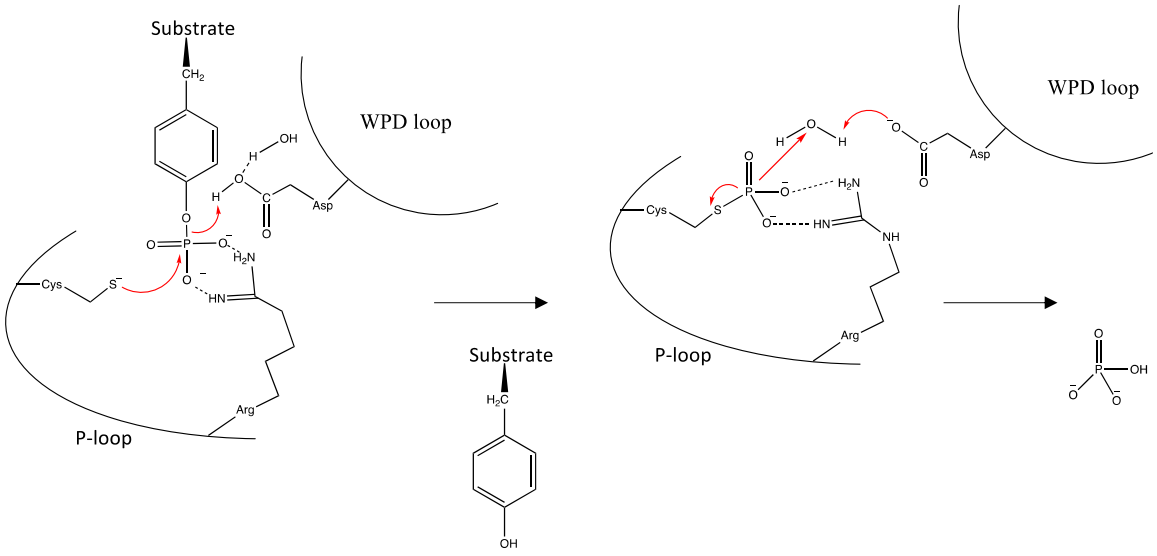


Figure 1.7. Protein tyrosine phosphatase (PTP) catalytic mechanism.

A substrate containing a phosphorylated tyrosine residue enters the P-loop of the PTP where it is stabilized by an Arg residue in the P-loop. The Asp of the WPD loop acts as a general acid while a Cys in the P-loop performs a nucleophilic attack of the phosphate on the phosphoprotein, forming a phospho-substrate intermediate. An ordered water molecule regenerates Asp in the WPD loop and allows for phosphate to leave the PTP active site, thus regenerating the thiolate ion on the phosphatase.

Composition and regulation of serine/threonine phosphatases

Over 98% of cellular protein phosphorylation occurs on serine or threonine residues.[104] As such, the proteins that are able to catalyze the dephosphorylation of these residues are of immense importance. These proteins, the serine/threonine phosphatases (PSPs), are structurally distinct from PTPs in that they are divided into three families possessing different catalytic subunits. The two largest families are the phosphoprotein phosphatases (PPPs) and the Mg^{2+}/Mn^{2+}-dependent protein phosphatases (PPMs). The two most prevalent PPPs are protein phosphatase 1 and 2A (PP1 and PP2A, respectively). PP1 and PP2A exist as holoenzymes, with PP1 having a catalytic and a regulatory subunit, whereas PP2A has these same subunits plus an additional scaffolding subunit. A unique member of the PPP family is PP2B (also known as calcineurin or PP3) which has a catalytic and regulatory Ca^{2+}-binding subunit and is critical for aspects of immune response in humans.[105] The combinatorial diversity afforded by subunit combinations compensate for the disparity between the number of serine/threonine kinases compared to phosphatase genes.[104] The PPM class of PSPs possess a pair of active site divalent metal ions (Mg^{2+} or Mn^{2+}) which are thought to promote dephosphorylation by activating a metal-bound water, similar to the PPPs, although it is only composed of one subunit, an example being PP2C. Knowing the identity of the kinase-phosphatase pair that regulates a given phosphorylation site helps in understanding the biological contexts in which these events occur. One method for determining the phosphatase identity is to treat cells with phosphatase inhibitors and screen for increases in phosphorylation using site-specific antibodies. Polyketide inhibitors, such as okadaic acid and calyculin A, are typically used, which are capable of inhibiting PP1, PP2A, and PP4.[106,107] However, PP2B is insensitive to okadaic acid and

calyculin A, but *is* sensitive to inhibitors like cyclosporine a or tacrolimus (FK506).[108] These inhibitors bind proteins called immunophilins, which in turn bind PP2B, thus inhibiting PP2B activity.[109] PPM inhibitors are difficult to design owing to the different isoforms of phosphatases present. A specific inhibitor for PP2C is sanguinarine, a plant alkaloid, which likely operates as a competitive inhibitor of phosphosubstrates.[110] A downfall of these approaches is the lack of specificity for each phosphatase, as seen with okadaic acid and calyculin A targeting overlapping substrates and potential alterations in kinase activity, influencing phosphorylation. As such, a need exists to develop methods allowing for the elucidation of specific phosphatase-substrate interactions.

UBIQUITINATION AT A GLANCE

A second abundant PTM that is highly conserved in eukaryotes is ubiquitination, where the small molecule ubiquitin (Ub) is conjugated to primary amines found on proteins, mainly on lysine residues.[111] The consequence of ubiquitination varies from regulating protein half-life, to allowing a myriad of biological process to occur.[112,113] By way of review, ubiquitination occurs through the sequential action of three enzymes: an E1 ubiquitin activating enzyme; an E2 ubiquitin-conjugating enzyme; and an E3 ubiquitin ligase, ultimately adding the small 76 amino acid-long protein ubiquitin (Ub) onto substrates. Mechanistically, the E1 enzyme binds Mg^{2+}-ATP and Ub, allowing for the adenylation of the Ub C-terminal carboxyl group thereby activating the protein. The activated Ub molecule subsequently forms a thioester linkage with a cysteinyl residue on the E1 and is transferred to a cysteinyl residue on an E2 conjugating enzyme. An E3 enzyme then effects the ubiquitination of specific substrates - either through transfer of Ub to the substrate directly by the E2, or via the Ub molecule forming a thioester bond with the E3 prior to protein ubiquitination (**Figure 1.8B**).

Ubiquitination occurs on primary amines on substrates, predominantly on the ϵ-amino group of lysine, with isopeptide bond formation between substrate lysine residues and the C-terminal glycine of ubiquitin, forming a ubiquitinated protein (**Figure 1.9**). This ubiquitin moiety that is conjugated to the substrate can exist in several forms. Monoubiquitination occurs when a single molecule of Ub is attached to a substrate and is typically involved in regulating processes such as endocytosis and DNA repair (**Figure 1.8C**).[122] In polyubiquitination, Ub monomers form linkages with one another forming chains on substrates (**Figure 1.8E**).[123] Ub molecules can link with one another forming

chains via isopeptide bonds between one of the 7 lysine (Lys, K) residues of Ub (K6, K11, K27, K29, K33, K48, and K63) leading to the formation of an isopeptide bond between the C-terminal glycine of another Ub molecule.

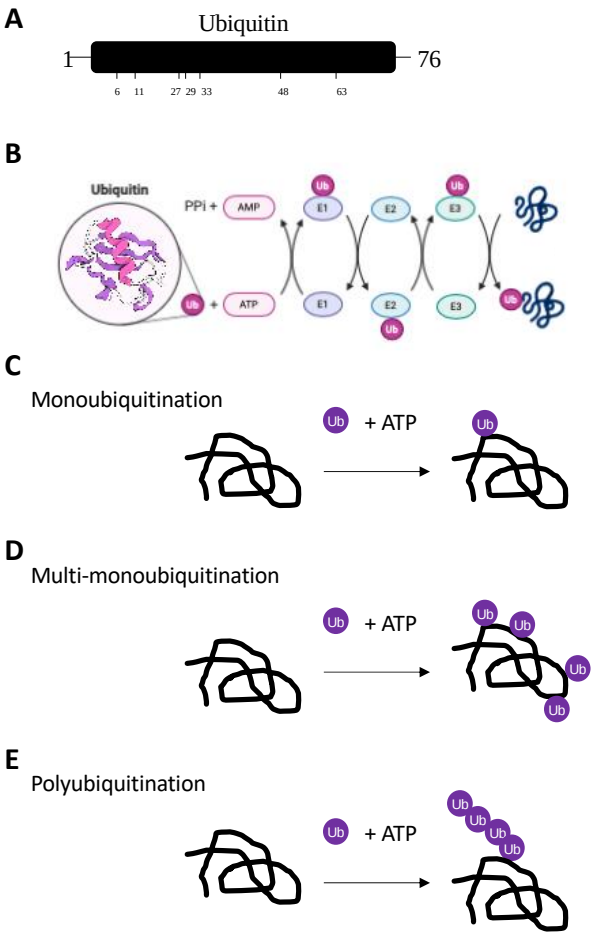


Figure 1.8. Ubiquitination cascade.

(**A**) Schematic of ubiquitin (Ub) primary sequence, highlighting the lysine residues found within Ub, indicated by their amino acid residue number.

(**B**) Ubiquitination is the highly conserved process of covalently linking a Ub molecule to a substrate through the action of an E1 Ub activating enzyme and ATP, forming a thioester bond with ubiquitin. The activated Ub is then transferred to an E2 Ub-conjugating enzyme, before interaction with an E3 Ub ligase, which catalyzes the addition of Ub to the protein substrate. Substrates can be modified in different ways by mono/poly/-or multiple mono/poly-Ub conjugations leading to different cellular effects. Made using Biorender (https://biorender.com)

(**C**) Monoubiquitination describes the modification of a protein with a single Ub molecule at a site.

(**D**) Multi-monoubiquitination or multiubiquitination describes the process whereby a substrate protein is modified with a single/multiple Ub molecule(s) at multiple sites, causing the substrate to be conjugated with multiple ubiquitin molecules.

(**E**) Polyubiquitination occurs when branches of ubiquitin are formed on a single site on the substrate. The seven Ub lysine residues allows for the formation of Ub "chains" through various linkages.

The most abundant polyubiquitination linkage class is represented by K48-linked Ub chains, which is typically functionally associated with the degradation of ubiquitin-tagged substrates via the 26S proteasome,[124] K63 which is responsible for proteasome-independent roles,[125] K11-linked ubiquitin, and K27, K29, and K33 in lower levels. The complexity of various branching combinations and different lysine residues that can be modified provide far adversity of cellular regulatory functions and is referred to as the "ubiquitin code,"[126] and may help explain why ubiquitination is such a prominent post-translational modification, especially for the regulation of eukaryotic metabolism and development. Ubiquitination in the model plant *Arabidopsis thaliana* has been closely examined and will be discussed later.

Consequences of protein ubiquitination

Ubiquitination has been shown to regulate a myriad of processes, most commonly associated with protein quality control through degradation. Here, ubiquitin serves as a signal that can be recognized leading to substrate degradation via the 26S proteasome.[127] The 26S proteasome is a multi-subunit complex composed of two 19S regulatory particles, and a 20S core particle, and is referred to as the ubiquitin-proteasome system (UPS) when facilitating degradation of ubiquitinated proteins. The 19S particle recognizes ubiquitinated substrates, facilitates in their unfolding, and removes the Ub moiety through its deubiquitinase (DUB) activity. The deubiquitinated protein enters the core particle where it is degraded via the trypsin-like, chymotrypsin-like, and caspase-like activities of the core particle.[128] This type of regulation has been shown to be associated other modifications such as phosphorylation, where phosphorylation motifs known as "phosphodegrons" serve as a signal to recruit E3 Ub ligases, leading to substrate ubiquitination and degradation.

This type of regulation is observed for many proteins from all organisms, notably the transcriptional co-activator TAZ, and cyclins, important for cell cycle progression.[129,130] Proteasome-independent routes of regulation also exist. Rather than utilizing a large protein complex for degradation, autophagy functions through the synthesis of membrane vesicles called autophagosomes, which engulf cellular material and eventually fuse with the lysosome, causing recycling or degradation of the engulfed material. Some proteins involved in ubiquitination, such as the E3 Ub ligase TRAF6 that participates in the modification of a variety of autophagy initiating genes such as Beclin-1. TRAF6 is also involved in the production of PI3P by facilitating the interaction of two autophagy-interacting kinases leading to recruitment of proteins responsible for membrane extension, autophagosome formation and engulfment of ubiquitinated proteins. [131,132]

Non-degradative functions of Ub have also reported, emphasizing the widespread role of Ub in eukaryotic cellular function. Typically, K63-linked Ub modifications do not lead to protein degradation but influence other functions such as endocytosis, DNA repair, and signaling.[133] For these activities to take place, Ub-interaction motifs (UIMs), Ub associated domains (UBAs), or Ub-binding domains (UBDs) found on other proteins allow for their interaction with the ubiquitinated protein. UIMs are prevalent in proteins involved in endocytosis, typically involving interaction with Ile44 through hydrophobic interactions with the conserved 'LALAL' sequence.[134] Many components in the process of endocytosis are regulated through this type of nonproteolytic ubiquitination, such as epidermal growth factor receptor (EGFR), a receptor critical for cellular signaling that can be quickly internalized following binding its ligand. Here, the E3 ligase Cbl catalyzes the ubiquitination of EGFR, allowing the receptor to interact with the endosomal sorting

complex required for transport (ESCRT) machinery.[135] Proteins comprising ESCRT like hepatocyte growth factor-regulated tyrosine kinase substrate (Hrs) contain UIMs, allowing the tagged EGFR to contact and complete the endocytic process.[136] UBA domains are also able to bind mono- or polyUb, aiding in both degradative and nondegradative functions. Studies on the UBA of the human nucleotide excision repair protein UV excision repair protein RAD23 homolog A (HHR23A), have shown that UBA makes large hydrophobic contacts with Ub, allowing for interaction.[137] The UBA of HHR23A has been shown to also possess demethylase activity, allowing HHR23A to participate in DNA repair.[138] Since Ub modifications are known to regulate a diverse range of cellular and developmental processes, capturing a broader understanding of the Ub landscape of a given proteome is important.

Preserving ubiquitination

In order to monitor the Ub modification of substrates, the modification itself must be preserved. The process of deubiquitination, or the removal of Ub from substrate proteins, is performed by deubiquitinases (DUBs). DUBs are able to process Ub precursors, recycle Ub molecules during ubiquitination, cleave polyUb chains, and remove Ub from modified substrates. The human genome encodes for ~100 DUBs, divided into families based on their catalytic activity, mainly cysteine proteases or zinc metalloproteases.[139] Some of the most prevalent and widely studied DUBs are the Ub specific protease (USPs), catalyzing the deubiquitination of several targets. The human genome encodes for ~60 USP family members,[140] whereas the Arabidopsis genome encodes for 27, which appear to be highly conserved in plants.[141] Mechanistically, USPs catalyze deubiquitination through a catalytic cysteine, which performs a nucleophilic attack on the carbonyl carbon of Ub

glycine 76, ultimately leading to a cysteine-Ub intermediate and deubiquitination of the modified substrate.[142] When investigating an organism's ubiquitome, many groups have employed DUB inhibitors to prevent loss of ubiquitination, allowing for the identification of different Ub chain topologies and modification sites. A number of inhibitors have been used, with a majority being small molecules such as 2,6-diaminopyridine-3,5-bis-thiocyanate (PR-619) is a reversible broad-spectrum DUB inhibitor that has been used in large scale ubiquitome studies, showing specificity to DUBs exhibiting cysteine protease activity.[143,144,145] Mechanistically, these inhibitors can target the DUB-Ub intermediate formed, or near the active site of the DUB inhibiting the catalytic cysteine.[143] The study of DUBs and their inhibitors could be strengthened through the elucidation of the crystal structures of these interactions, which may aid in the production of more specific inhibitors with reduced off-target effect. The DUB inhibitor PR-619 has been shown to also stimulate DNA-topoisomerase complex formation, redistributing topoisomerase in the nucleus.[144]

A second and prominent form of experimental Ub preservation—either independently or in combination with DUB inhibition—involves inhibition of the 26S proteasome. MG132 is the most prominent inhibitor used, and is composed of the amino acids carbobenzoxy-Leu-Leu-Leucinal. This peptide aldehyde is a reversible inhibitor able to inhibit the chymotrypsin-like activity of the 26S proteasome by forming adducts with threonine residues in the 20S core particle, thereby mimicking the transition state intermediate made during proteolytic cleavage.[146] Like DUB inhibition, off-target effects caused by proteasome inhibitor treatment have been reported such as Jun kinase activation, apoptosis induction, and DNA damage.[147,148,149] As both DUB and proteasome inhibitors

stimulate off-target effects, users must be aware of the potential alteration of the ubiquitome/proteome of the organism under investigation.

Strategies for studying protein ubiquitination

Since ubiquitination is a prevalent post-translational modification that regulates various cellular processes, many groups have worked to elucidate the ubiquitomes of a variety of model organisms. Prior to the development of mass spectrometry (MS), substrate-specific ubiquitination was monitored utilizing prior knowledge or evidence for ubiquitination. Ubiquitination can be detected through antibodies that recognize different forms of Ub, as well as by observing shifts in molecular weights or smears in immunoblots: a monoubiquitinated protein will migrate at the mass of the substrate plus 8 kDa, or will exhibit a distribution of high-molecular weight variants when polyubiquitinated or multi-monoubiquitinated.[150] Improvements to MS methodologies have been used to undetake the expansion of many organism ubiquitomes, including Arabidopsis, relying on the use of tagged variants of Ub,[117,118,120] expression of Ub binding/interacting domains,[151,152,153] and the use of antibodies directed against ubiquitinated residues,[154] for the enrichment of ubiquitinated substrates followed by MS analyses for protein identification. While for providing protein identification, protocols were laborious and required manipulation of the organism's proteome through knockdowns or overexpression. Interestingly, when Ub-conjugated proteins are digested with trypsin, which cleaves C-terminal to lysine or arginine residues, peptides possessing a diglycine remnant are generated.[119] This diglycine (Gly-Gly) dipeptide is composed of the last two amino acids of ubiquitin covalently bound to the modified lysine residue, resulting in a 114.0429 Da mass shift after proteolysis (**Figure 1.9A**). This mass shift is readily detectable through MS, allowing for the

identification of specific Ub sites on a given peptide. A second remnant following trypsin digestion on ubiquitinated lysine residues has also been reported, leaving a terminal four amino acids of Ub (Leu-Arg-Gly-Gly) attached to lysine, and is observed as a 383.2381 Da mass shift (**Figure 1.9B**).[155]

The development of anti-diglycine antibodies has propelled the field forward when combined with traditional affinity-purification MS workflows.[119] The commonly available anti-diglycine antibodies have been raised against ubiquitinated histone clone GX41,[119] or alternatively one raised against a panel of diglycine-modified peptides (Cell Signaling Technology®). Workflows utilizing either of these antibodies require trypsin digestion prior to enrichment, allowing for the production of the diglycine remnant on modified residues. Following digestion, peptides are enriched using the Ub-remnant antibodies and submitted for MS analysis. These diglycine scanning approaches have shown great utility by coupling enrichment with fractionation and quantitation, leading to the identification of tens of thousands of ubiquitination sites from various cell sources.[75,114,115,116] These antibodies and similar approaches have also been utilized in various plant species such as Arabidopsis, *Triticum aestivum* (wheat), *Camellia sinensis* (tea leaves), and *Oryza sativa* (rice). However, most of the well-cited studies have resulting in the identification of fewer than ~1500 sites per scan, with Arabidopsis investigations revealing about 200 sites.[156,154,157,158] To date, a majority of these investigations have utilized DDA MS methods to expand the catalogue of ubiquitination, where improvements to sample preparation and MS acquisition will impact the quality and quantity of new information revealed.

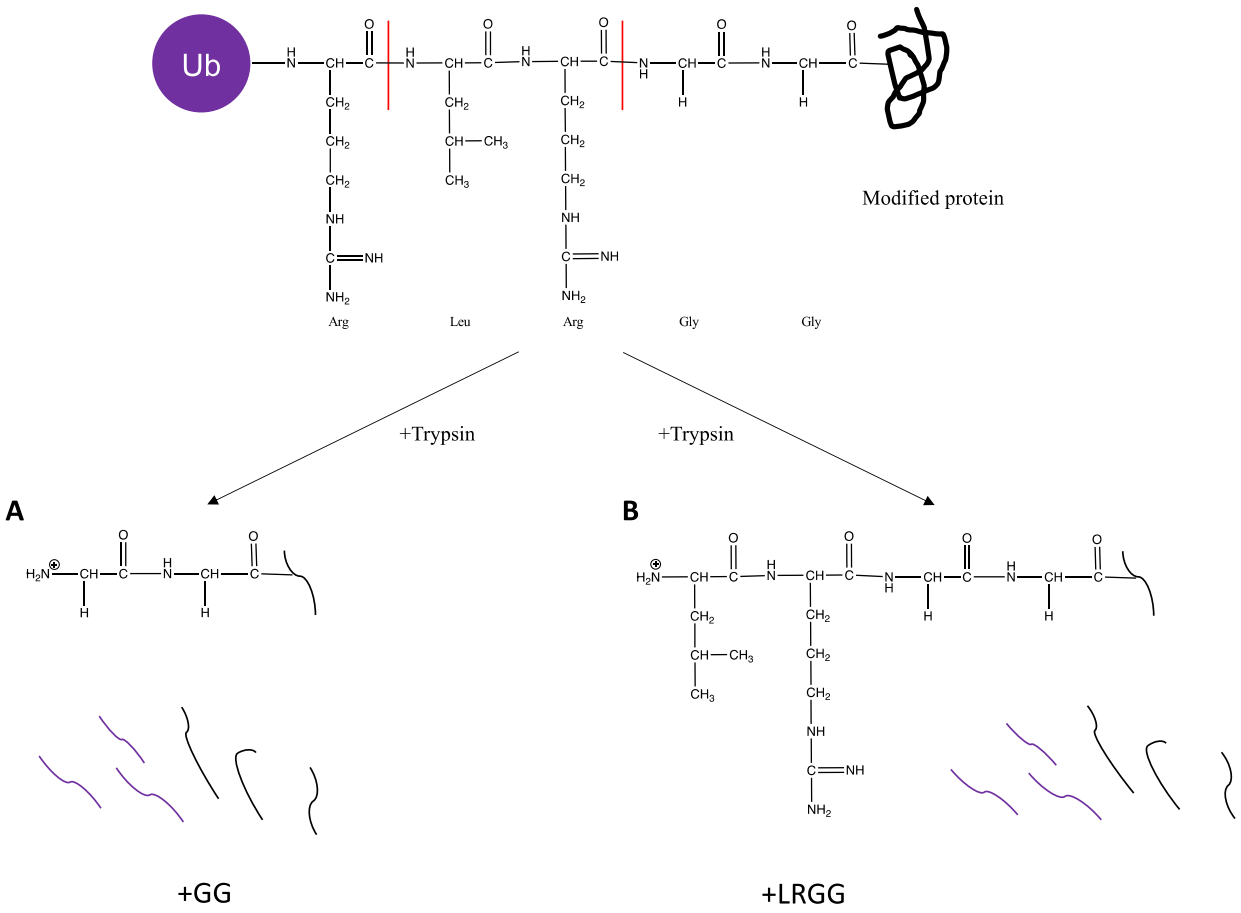


Figure 1.9. Diglycine remnant.

Ubiquitinated proteins are linked through an isopeptide bond between a lysine residue on the protein and the C-terminus of ubiquitin.

(**A**) Following trypsin digestion, a diglycine remnant can remain on the modified peptide, adding 114.04 Da to the peptide.

(**B**) Following trypsin digestion, a missed cleavage can occur rather than leaving a diglycine remnant, the last 4 amino acids of ubiquitin remain on the modified peptide, LRGG, adding a mass of 383.2281 Da. Ub = ubiquitin; red lines indicate trypsin digestion sites.

INSTRUMENTATION: THE SYNAPT G2-Si

The primary mass spectrometer utilized in this study was the Waters Synapt G2-Si HDMS^E with travelling wave technology. The instrument possesses a front-end nano-UPLC chromatography system for reverse phase separation when operating the instrument with an ESI ionization source. Ions generated can then be separated using the Stepwave™ ion guide (**Figure 1.10**), comprised of stacked ring ion guides (SRIG), a series of washer-like ring electrodes with alternating phases RF only voltages arranged in sequence allowing for the removal of neutrals and confining ions in a selected direction.[52] Much of the power of this instrument comes from this ion mobility cell, divided into three regions of ion guides; the trap, the IMS, and the transfer. The trap ion guide accumulates ions from the previous mobility separation releasing ion packets into the IMS ion guide for mobility separation (**Figure 1.10**). Both nitrogen and helium are utilized in IMS, with ions being more mobile in helium than nitrogen;[159] however, this increased mobility influences the height of the T-wave amplitude leading to decreased resolution, therefore nitrogen gas in the IMS acts as the collision gas with a chamber filled with helium before the main IMS to balance the pressure caused by nitrogen.[159] Ion packets then arrive at the transfer guide, allowing for mobility separated ions to be directed to the oa-TOF for mass analysis.[52] Ion mobility separation in the Synapt utilizes a sinusoidal-like wave, with the potential being produced by four plates set to a high potential followed by a set of four plates set to a low potential.[160]

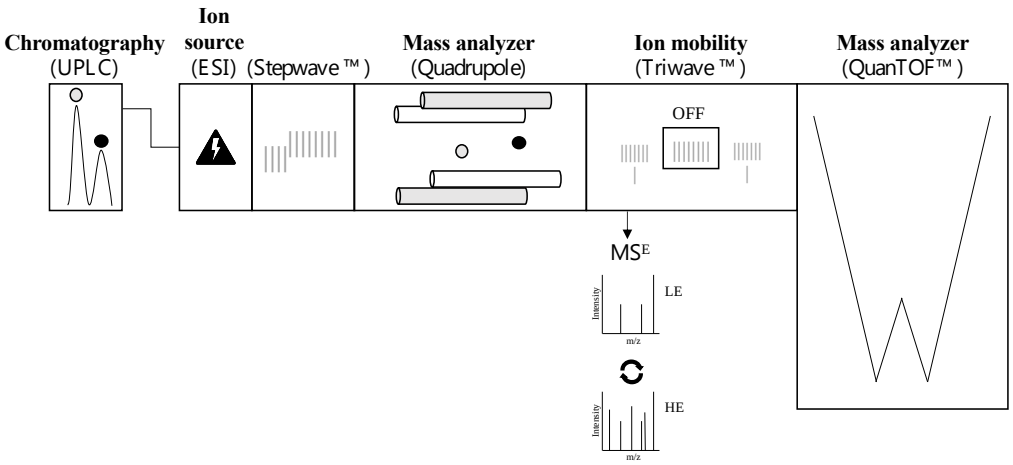


Figure 1.10. Synapt G2-Si MSE workflow.

Protein digests were fractionated with an on-line ultra-high-performance liquid chromatography system, where peptides are resolved based on their hydrophobicity. As peptides elute off the column, they become ionized via ESI where they are then guided through the Stepwave™ ion guide, where neutrals will be lost and ions will travel through to the quadrupole, where they can be fragmented. When operating under the TOF mode, ions entering the Triwave™ will experience low and high collision energies in the trap T-wave creating fragment ions accelerated out of the transfer T-wave and the QuanTOF™ mass analyzer, where data will be collected and prepared for processing.

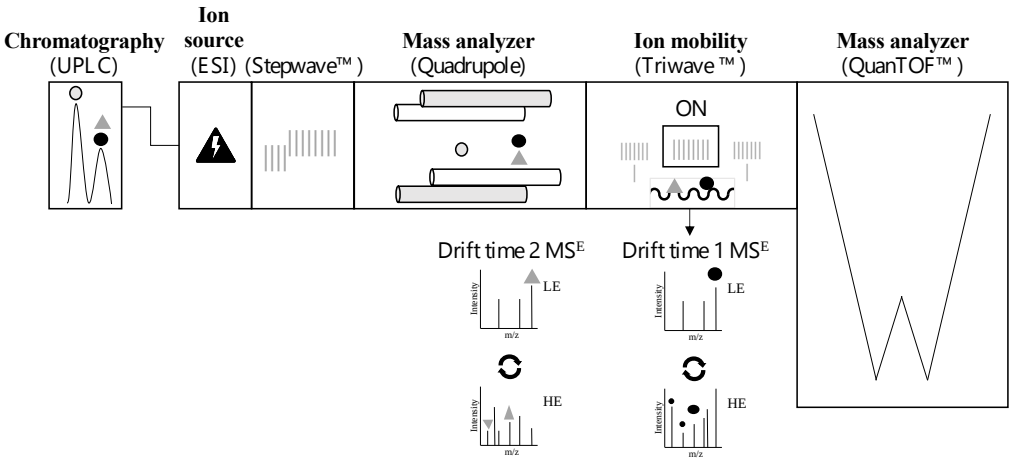


Figure 1.11. Synapt G2-Si HDMS^E workflow.

Protein digests generated were fractionated with an on-line ultra-high-performance liquid chromatography system, where peptides are differentially depending on their hydrophobicity. As peptides elute off the column, they become ionized via ESI where they are then guided through the Stepwave™ ion guide, where neutrals will be lost and ions will travel through to the quadrupole, where they can be fragmented. When operating under the HDMS^E mode, ions entering the Triwave™ will experience low and high collision energies in the trap T-wave ensures ions are trapped prior to undergoing IMS in the IMS cell by collision with nitrogen gas. Mobility separated ions will then be accelerated out of the transfer T-wave and into the mass analyzer, where data will then be collected and ready for processing.

There are a multitude of parameters controlling various parts of the MS instrument used in this study. The trap and transfer T-waves can also operate as collision cells as well as the ion guides described above. When acting as collision cells, the voltages in each of the cells can be increased. The Synapt allows for independent control for the voltage of the trap and transfer T-waves. When operating the trap as a collision cell, fragment ions are generated and enter the IMS T-wave region allowing for separation in this second travelling wave device and then enter the transfer T-wave which can operate simply as an ion guide, allowing the fragment ions to pass to the TOF detector. Alternatively, the trap T-wave can operate as an ion guide, allowing ions generated to be mobility-separated in the IMS T-wave, and then fragmented at the transfer T-wave, with fragment ions possessing the same drift time as their coinciding precursor ions. The time-aligned parallel (TAP) fragmentation can also be used,[56] where both the trap and transfer T-waves both operate as collision cells, producing fragment ions in the trap T-wave, mobility separation in the IMS T-wave, followed by further fragmentation forming second generation fragment ions at the transfer T-wave. The versatility of this instrument has championed for the Synapt G2-Si as a powerful tool in the proteomic, lipidomic, and metabolomic applications.[161]

Following ejection from the transfer cell, ions enter the QuanTof™ system where a high-field pusher accelerates through to the dual stage reflectron and ion mirror until reaching the ion detection system. The pusher and reflectron reduces ion turnaround times and improves focusing of high energy ions, respectively. The detection system utilizes an ultrafast electron multiplier which is processed by a hybrid analog-to-digital converter (ADC) detector.[162] When the machine is operated in a HD-DDA way utilizing wideband enhancement, the TOF pusher can synchronize the acceleration of ions exiting IMS and

entering the oa-TOF, allowing for close to a 100% duty cycle defined as the percentage of time a mass analyzer spends on an ion of interest during a single experimental cycle.[163]

Once these measurements are completed, data can be examined using software such as MassLynx and further processed using programs such as ProteinLynx Global SERVER (PLGS) and Progenesis QI. More targeted approaches can be processed using specific software, such as TargetLynx for monitoring transitions of specific ions. These programs utilize multiple criteria obtained from measurements, such as retention time, *m/z*, and ion mobility.[162]

PLGS uses two algorithms, Apex3D and Pep3D, to process raw data obtained from the mass spectrometer, working both with $HDMS^E$ and TOF modes. The first algorithm Apex3D, selects a list of ions that are above a user defined threshold, integrating ion current signals across their chromatographic elution.[164] Prior to using the second algorithm, Pep3D, the precursor fragment ion association is unknown, but using the exact mass and retention time (EMRT) of the fragment ion can putatively assign a precursor-fragment ion link.[164] Using this system, high-resolution MS data was produced allowing for further understanding of specific proteins, their regulation, modifications, as well as the identification of proteins for their use in other applications such as disease biomarkers.

REFERENCES

1. Andy T. Kong, Felipe V. Leprevost, Dmitry M. Avtonomov, Dattatreya Mellacheruvu, and A. I. N. MSFragger: ultrafast and comprehensive peptide identification in shotgun proteomics. *Nat. Methods* **14**, 513–520 (2017).
2. Cole, J., Hanson, E. J., James, D. C., Dockrell, D. H. & Dickman, M. J. Comparison of data-acquisition methods for the identification and quantification of histone post-translational modifications on a Q Exactive HF hybrid quadrupole Orbitrap mass spectrometer. *Rapid Commun. Mass Spectrom.* **33**, 897–906 (2019).
3. Moseley, M. A. *et al.* Scanning Quadrupole Data-Independent Acquisition, Part A: Qualitative and Quantitative Characterization. *J. Proteome Res.* **17**, 770–779 (2018).
4. Perez, C. J., Bagga, A. K., Prova, S. S., Yousefi Taemeh, M. & Ifa, D. R. Review and perspectives on the applications of mass spectrometry imaging under ambient conditions. *Rapid Commun. Mass Spectrom.* **33**, 27–53 (2019).
5. Heidemann, J. *et al.* Further insights from structural mass spectrometry into endocytosis adaptor protein assemblies. *Int. J. Mass Spectrom.* **447**, 1–9 (2020).
6. Park J, Piehowski PD, Wilkins C, Zhou M, Mendoza J, Fujimoto GM, Gibbons BC, Shaw JB, Shen Y, Shukla AK, M. R. Informed-Proteomics: Open Source Software Package for Top- down Proteomics. *Nat. Methods* **14**, 909–914 (2017).
7. Giansanti, P., Tsiatsiani, L., Low, T. Y. & Heck, A. J. R. Six alternative proteases for mass spectrometry-based proteomics beyond trypsin. *Nat. Protoc.* **11**, 993–1006 (2016).
8. Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. & Whitehouse, C. M. Electrospray ionization for mass spectrometry of large biomolecules. *Science.* **246**, 64–71 (1989).
9. Chait, B. T. Shufang Niu, Wenzhu Zhang, and Brian T. Chait. *Am. Soc. Mass Spectrom.* **0305**, 1–7 (1997).
10. Karas, M. *et al.* Principles and applications of matrix-assisted UV-laser desorption/ionization mass spectrometry. *Anal. Chim. Acta* **241**, 175–185 (1990).
11. Zenobi, R. & Knochenmuss, R. ChemInform Abstract: Ion Formation in MALDI Mass Spectrometry. *ChemInform* **30**, no-no (2010).
12. Bae, Y. J., Shin, Y. S., Moon, J. H. & Kim, M. S. Degree of ionization in MALDI of peptides: Thermal explanation for the gas-phase ion formation. *J. Am. Soc. Mass Spectrom.* **23**, 1326–1335 (2012).
13. Karas, M., Glückmann, M. & Schäfer, J. Ionization in matrix-assisted laser desorption/ionization: Singly charged molecular ions are the lucky survivors. *J. Mass Spectrom.* **35**, 1–12 (2000).
14. Jaskolla, T. W. & Karas, M. Compelling evidence for lucky survivor and gas phase protonation: The unified MALDI analyte protonation mechanism. *J. Am. Soc. Mass*

Spectrom. **22**, 976–988 (2011).

15. Yates, J. R. Mass spectrometry and the age of the proteome. *J. Mass Spectrom.* **33**, 1–19 (1998).

16. Boesl, U. TIME-OF-FLIGHT MASS SPECTROMETRY: INTRODUCTION TO THE BASICS Ulrich. *Mass Spectrom. Rev.* **36**, 86–109 (2016).

17. Miller, P. E. & Denton, M. B. The quadrupole mass filter: Basic operating concepts. *J. Chem. Educ.* **63**, 617–622 (1986).

18. Morris, M., Thibault, P. & Boyd, R. K. Characterization of a high-pressure quadrupole collision cell for low-energy collision-indneed dissociation. *J. Am. Soc. Mass Spectrom.* **5**, 1042–1063 (1994).

19. Javahery, G. & Thomson, B. A segmented radiofrequency-only quadrupole collision cell for measurements of ion collision cross section on a triple quadrupole mass spectrometer. *J. Am. Soc. Mass Spectrom.* **8**, 697–702 (1997).

20. Makarov, A. Electrostatic axially harmonic orbital trapping: A high-performance technique of mass analysis. *Anal. Chem.* **72**, 1156–1162 (2000).

21. Scigelova, M. & Makarov, A. Orbitrap mass analyzer - Overview and applications in proteomics. *Proteomics* **1**, 16–21 (2006).

22. Shliaha, P. V. *et al.* Middle-Down Proteomic Analyses with Ion Mobility Separations of Endogenous Isomeric Proteoforms. *Anal. Chem.* **92**, 2364–2368 (2020).

23. Liu, R., Li, Q. & Smith, L. M. Detection of large ions in time-of-flight mass spectrometry: Effects of ion mass and acceleration voltage on microchannel plate detector response. *J. Am. Soc. Mass Spectrom.* **25**, 1374–1383 (2014).

24. Imrie, D. C., Pentney, J. M. & Cottrell, J. S. A Faraday cup detector for high-mass ions in matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **9**, 1293–1296 (1995).

25. Gilar, M., Olivova, P., Daly, A. E. & Gebler, J. C. Two-dimensional separation of peptides using RP-RP-HPLC system with different pH in first and second separation dimensions. *J. Sep. Sci.* **28**, 1694–1703 (2005).

26. McCalley, D. V. & Guillarme, D. Evaluation of additives on reversed-phase chromatography of monoclonal antibodies using a 1000 Å stationary phase. *J. Chromatogr. A* **1610**, 460562 (2020).

27. Bobály, B., Beck, A., Fekete, J., Guillarme, D. & Fekete, S. Systematic evaluation of mobile phase additives for the LC-MS characterization of therapeutic proteins. *Talanta* **136**, 60–67 (2015).

28. Pappin, D. J. C., Hojrup, P. & Bleasby, A. J. Rapid identification of proteins by peptide-mass fingerprinting. *Curr. Biol.* **3**, 327–332 (1993).

29. Sleno, L. & Volmer, D. A. Ion activation methods for tandem mass spectrometry. *J. Mass Spectrom.* **39**, 1091–1112 (2004).

30. Kim, M.S., Kandasamy, K., Chaerkady, R., and Pandey, A. Assessment of

resolution parameters for CID-based shotgun proteomic experiments on the LTQ-Orbitrap mass spectrometer. *J. Am. Soc. Mass Spectrom.* **21**, 1606–1611 (2010).

31. Mitchell Wells, J. & McLuckey, S. A. Collision-induced dissociation (CID) of peptides and proteins. *Methods Enzymol.* **402**, 148–185 (2005).

32. Fabris, D., Kelly, M., Murphy, C., Wu, Z. & Fenselau, C. High-Energy Collision-Induced Dissociation of Multiply Charged Polypeptides Produced by Electrospray. *Society* **4**, 652–661 (1993).

33. Harrison, A. G. TO b OR NOT TO b: THE ONGOING SAGA OF PEPTIDE b IONS. *Mass Spectrom. Rev.* **28**, 640–654 (2009).

34. Grossmann, J. *et al.* Implementation and evaluation of relative and absolute quantification in shotgun proteomics with label-free methods. *J. Proteomics* **73**, 1740–1746 (2010).

35. Kirkpatrick, D. S., Gerber, S. A. & Gygi, S. P. The absolute quantification strategy: A general procedure for the quantification of proteins and post-translational modifications. *Methods* **35**, 265–273 (2005).

36. Lange, V., Picotti, P., Domon, B. & Aebersold, R. Selected reaction monitoring for quantitative proteomics: A tutorial. *Mol. Syst. Biol.* **4**, (2008).

37. Anderson, L. & Hunter, C. L. Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins. *Mol. Cell. Proteomics* **5**, 573–588 (2006).

38. Mollah, Sa. *et al.* Targeted mass spectrometric strategy for global mapping of ubiquitination on proteins. *Rapid Commun. Mass Spectrom.* **21**, 3357–3364 (2007).

39. Gillet, L. C. *et al.* Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: A new concept for consistent and accurate proteome analysis. *Mol. Cell. Proteomics* **11**, 1–17 (2012).

40. Venable, J. D., Dong, M. Q., Wohlschlegel, J., Dillin, A. & Yates, J. R. Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nat. Methods* **1**, 39–45 (2004).

41. Bern, M., Finney, G., Hoopmann, M.R.,, Merrihew, G., Toth, M.J., and M.J. MacCoss. Deconvolution of Mixture Spectra from Ion-Trap Data- Independent-Acquisition Tandem Mass Spectrometry Marshall. *Anal. Chem.* **82**, 1–19 (2010).

42. Plumb, R. S., Johnson, K. A., Rainville, P., Smith, B. W., Wilson, I. D., Castro-Perez, J. M., & Nicholson, J. K. UPLC/MSE; a new approach for generating molecular fragment information for biomarker structure elucidation. *Rapid Commun. Mass Spectrom.* **20**, 1989–1994 (2006).

43. Zhao, X., Wei, J. & Yang, M. Simultaneous analysis of iridoid glycosides and anthraquinones in morinda officinalis using UPLC-QqQ-MS/MS and UPLC-Q/TOF-MSE. *Molecules* **23**, (2018).

44. Zhu, X., Chen, Y. & Subramanian, R. Comparison of information-dependent acquisition, SWATH, and MS All techniques in metabolite identification study

employing ultrahigh-performance liquid chromatography-quadrupole time-of-flight mass spectrometry. *Anal. Chem.* **86**, 1202–1209 (2014).

45. Castro-Perez, J. M. *et al.* Comprehensive LC-MSE lipidomic analysis using a shotgun approach and its application to biomarker detection and identification in osteoarthritis patients. *J. Proteome Res.* **9**, 2377–2389 (2010).

46. Geromanos, S. J. *et al.* The detection, correlation, and comparison of peptide precursor and product ions from data independent LC-MS with data dependant LC-MS/MS. *Proteomics* **9**, 1683–1695 (2009).

47. Hoaglund-Hyzer, C. S., Li, J. & Clemmer, D. E. Mobility labeling for parallel CID of ion mixtures. *Anal. Chem.* **72**, 2737–2740 (2000).

48. May, J. C. & McLean, J. A. Ion mobility-mass spectrometry: Time-dispersive instrumentation. *Anal. Chem.* **87**, 1422–1436 (2015).

49. Wu, C., Siems, W. F., Asbury, G. R. & Hill, H. H. Electrospray ionization high-resolution ion mobility spectrometry-mass spectrometry. *Anal. Chem.* **70**, 4929–4938 (1998).

50. Marchand, A., Livet, S., Rosu, F. & Gabelica, V. Drift Tube Ion Mobility: How to Reconstruct Collision Cross Section Distributions from Arrival Time Distributions? *Anal. Chem.* **89**, 12674–12681 (2017).

51. Pringle, S. D. *et al.* An investigation of the mobility separation of some peptide and protein ions using a new hybrid quadrupole/travelling wave IMS/oa-ToF instrument. *Int. J. Mass Spectrom.* **261**, 1–12 (2007).

52. Giles, K. *et al.* Applications of a travelling wave-based radio-frequency-only stacked ring ion guide. *Rapid Commun. Mass Spectrom.* **18**, 2401–2414 (2004).

53. Helm, D. *et al.* Ion mobility tandem mass spectrometry enhances performance of bottom-up proteomics. *Mol. Cell. Proteomics* **13**, 3709–3715 (2014).

54. Silva, J. C. *et al.* Simultaneous qualitative and quantitative analysis of the Escherichia coli proteome: A sweet tale. *Mol. Cell. Proteomics* **5**, 589–607 (2006).

55. Silva, J. C., Gorenstein, M. V., Li, G. Z., Vissers, J. P. C. & Geromanos, S. J. Absolute quantification of proteins by LCMSE: A virtue of parallel MS acquisition. *Mol. Cell. Proteomics* **5**, 144–156 (2006).

56. Bruderer, R. *et al.* Optimization of experimental parameters in data-independent mass spectrometry significantly increases depth and reproducibility of results. *Mol. Cell. Proteomics* **16**, 2296–2309 (2017).

57. Maxson & Mitchell. Site specific analysis of changes in the glycosylation of proteins in liver cirrhosis using data independent workflow with soft fragmentation. *Anal. Bioanal. Chem.* **409**, 619–627 (2017).

58. Craig, R. & Beavis, R. C. TANDEM: Matching proteins with tandem mass spectra. *Bioinformatics* **20**, 1466–1467 (2004).

59. Z Noor, Z., Ahn, S.B., Baker, M.S., Ranganathan, S. and Mohamedali, A. Mass spectrometry–based protein identification in proteomics—a review. *Brief. Bioinform.* **00**, 1–19 (2019).

60. Mann, M. & Wilm, M. Error-Tolerant Identification of Peptides in Sequence Databases by Peptide Sequence Tags. *Anal. Chem.* **66**, 4390–4399 (1994).

61. Mørtz, E. *et al.* Sequence tag identification of intact proteins by matching tandem mass spectral data against sequence data bases. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 8264–8267 (1996).

62. Chen, Y., Chen, W., Cobb, M. H. & Zhao, Y. PTMap - A sequence alignment software for unrestricted, accurate, and full-spectrum identification of post-translational modification sites. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 761–766 (2009).

63. Kertesz-Farkas, A., Reiz, B., P. Myers, M. & Pongor, S. Database Searching in Mass Spectrometry Based Proteomics. *Curr. Bioinform.* **7**, 221–230 (2012).

64. Elias, J. E. & Gygi, S. P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* **4**, 207–214 (2007).

65. Kim, H., Lee, S. & Park, H. Target-small decoy search strategy for false discovery rate estimation. *BMC Bioinformatics* **20**, 1–6 (2019).

66. Moore, R. E., Young, M. K. & Lee, T. D. Qscore: An algorithm for evaluating SEQUEST database search results. *J. Am. Soc. Mass Spectrom.* **13**, 378–386 (2002).

67. Käll, L., Storey, J. D., MacCoss, M. J. & Noble, W. S. Assigning significance to peptides identified by tandem mass spectrometry using decoy databases. *J. Proteome Res.* **7**, 29–34 (2008).

68. Elias, J. E. & Gygi, S. P. Target-Decoy Search Strategy for Mass Spectrometry-Based Proteomics Joshua. *Methods Mol. Biol.* **604**, 55–71 (2010).

69. Reisdorph, N. A., Walmsley, S. & Reisdorph, R. A perspective and framework for developing sample type specific databases for LC/MS-based clinical metabolomics. *Metabolites* **10**, (2020).

70. Ong, S. E. & Mann, M. Mass Spectrometry–Based Proteomics Turns Quantitative. *Nat. Chem. Biol.* **1**, 252–262 (2005).

71. S. P. Gygi, B. Rist, 4, S. A. Gerber, F. Turecek, M. H. Gelb, & R. A. Access Quantitative analysis of complex protein mixtures using isotope-coded affinity tags Nature Biotechnology. *Nature* **17**, 994–999 (1999).

72. Shiio, Y. *et al.* Quantitative proteomic analysis of chromatin-associated factors. *J. Am. Soc. Mass Spectrom.* **14**, 696–703 (2003).

73. Ross, P. L. *et al.* Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics* **3**, 1154–1169 (2004).

74. Dayon, L. *et al.* Relative quantification of proteins in human cerebrospinal fluids by

MS/MS using 6-plex isobaric tags. *Anal. Chem.* **80**, 2921–2931 (2008).

75. Rose, C. M. *et al.* Highly Multiplexed Quantitative Mass Spectrometry Analysis of Ubiquitylomes. *Cell Syst.* **3**, 395-403.e4 (2016).

76. Ong, S. E. *et al.* Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* **1**, 376–386 (2002).

77. Martinović, S., Veenstra, T. D., Anderson, G. A., Paša-Tolic, L. & Smith, R. D. Selective incorporation of isotopically labeled amino acids for identification of intact proteins on a proteome-wide level. *J. Mass Spectrom.* **37**, 99–107 (2002).

78. Unwin, R. D. *et al.* Multiple reaction monitoring to identify sites of protein phosphorylation with high sensitivity. *Mol. Cell. Proteomics* **4**, 1134–1144 (2005).

79. Eshraghi, M., Gombar, R., De Repentigny, Y., Vacratsis, P. O. & Kothary, R. Pathologic alterations in the proteome of synaptosomes from a mouse model of spinal muscular atrophy. *J. Proteome Res.* **18**, 3042–3051 (2019).

80. Wang, H. & Hanash, S. Mass spectrometry based proteomics for absolute quantification of proteins from tumor cells. *Methods* **81**, 34–40 (2015).

81. Chen, G., Zhang, Y., Trinidad, J. C. & Dann, C. Distinguishing Sulfotyrosine Containing Peptides from their Phosphotyrosine Counterparts Using Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **29**, 455–462 (2018).

82. McNulty, D. E. & Annan, R. S. Hydrophilic interaction chromatography reduces the complexity of the phosphoproteome and improves global phosphopeptide isolation and detection. *Mol. Cell. Proteomics* **7**, 971–980 (2008).

83. Leitner, A. Phosphopeptide enrichment using metal oxide affinity chromatography. *TrAC - Trends Anal. Chem.* **29**, 177–185 (2010).

84. Bailey, L. S., Alves, M., Galy, N., Patrick, A. L. & Polfer, N. C. Mechanistic insights into intramolecular phosphate group transfer during collision induced dissociation of phosphopeptides. *J. Mass Spectrom.* **54**, 449–458 (2019).

85. Potel, C. M., Lemeer, S. & Heck, A. J. R. Phosphopeptide Fragmentation and Site Localization by Mass Spectrometry: An Update. *Anal. Chem.* **91**, 126–141 (2019).

86. Alonso, A. & Pulido, R. The extended human PTPome: A growing tyrosine phosphatase family. *FEBS J.* **283**, 1404–1429 (2016).

87. Yu, Z. H. & Zhang, Z. Y. Regulatory Mechanisms and Novel Therapeutic Targeting Strategies for Protein Tyrosine Phosphatases. *Chem. Rev.* **118**, 1069–1091 (2018).

88. Begley, M. J. *et al.* Molecular basis for substrate recognition by MTMR2, a myotubularin family phosphoinositide phosphatase. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 927–932 (2006).

89. Begley, M. J. *et al.* Crystal Structure of a Phosphoinositide Phosphatase, MTMR2: Insights into Myotubular Myopathy and Charcot-Marie-Tooth Syndrome. *Mol. Cell* **12**, 1391–1402 (2003).

90. Hille, E. J. D. and B. Understanding phosphoinositides: rare, dynamic, and essential

membrane phospholipids. *Biochem. J.* **476**, 1–23 (2019).

91. Dickson, E. J. *et al.* Dynamic formation of ER-PM junctions presents a lipid phosphatase to regulate phosphoinositides. *J. Cell Biol.* **213**, 33–48 (2016).

92. Berger, P., Schaffitzel, C., Berger, I., Ban, N. & Suter, U. Membrane association of myotubularin-related protein 2 is mediated by a pleckstrin homology-GRAM domain and a coiled-coil dimerization module. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 12177–12182 (2003).

93. Kim, S. J. & Ryu, S. E. Structure and catalytic mechanism of human protein tyrosine phosphatome. *BMB Rep.* **45**, 693–699 (2012).

94. Backer, J. M. The intricate regulation and complex functions of the Class III phosphoinositide 3-kinase Vps34. *Biochem. J.* **473**, 2251–2271 (2016).

95. Natsuko Jin, Michael J. Lang, and L. S. W. Phosphatidylinositol 3,5-bisphosphate: Regulation of cellular events in space and time. *Inf. Theory Quantum Phys.* **44**, 177–184 (2016).

96. Stephens, L. R., Hughes, K. T. & Irvine, R. F. Pathway of phosphatidylinositol(3,4,5)- trisphosphate synthesis in activated neutrophils. *Nature* **351**, 33–39 (1991).

97. Insall, R. H. & Weiner, O. D. PIP3, PIP2, and Cell Movement - Similar Messages, Different Meanings? *Dev. Cell* **1**, 743–747 (2001).

98. Tan, X., Thapa, N., Choi, S. & Anderson, R. A. Emerging roles of PtdIns(4,5)P2 - beyond the plasma membrane. *J. Cell Sci.* **128**, 4047–4056 (2015).

99. Pendaries, C. *et al.* Emerging roles of phosphatidylinositol monophosphates in cellular signaling and trafficking. *Adv. Enzyme Regul.* **45**, 201–214 (2005).

100. Poli, A. *et al.* Phosphatidylinositol 5 phosphate (Pi5p): From behind the scenes to the front (nuclear) stage. *Int. J. Mol. Sci.* **20**, 1–17 (2019).

101. Nunès, J. A. & Guittard, G. An emerging role for PI5P in T cell biology. *Front. Immunol.* **4**, 1–7 (2013).

102. Raess, M. A. *et al.* Expression of the neuropathy-associated MTMR2 gene rescues MTM1-associated myopathy. *Hum. Mol. Genet.* **26**, 3736–3748 (2017).

103. Prasenjit Manna and Sushil K. Jain. Phosphatidylinositol-3,4,5-Triphosphate and Cellular Signaling: Implications for Obesity and Diabetes Prasenjit. *Cell Physiol. Biochem.* **35**, 1253–1275 (2015).

104. Brautigan, D. L. & Shenolikar, S. Protein Serine/Threonine Phosphatases: Keys to Unlocking Regulators and Substrates. *Annu. Rev. Biochem.* **87**, 921–964 (2018).

105. Juvvadi, P. R. *et al.* Harnessing calcineurin-FK506-FKBP12 crystal structures from invasive fungal pathogens to develop antifungal agents. *Nat. Commun.* **10**, 1–18 (2019).

106. Suganuma, M. *et al.* Calyculin A, an Inhibitor of Protein Phosphatases, a Potent Tumor Promoter on CD-I Mouse Skin. *Cancer Res.* **50**, 3521–3525 (1990).

107. Ishihara, H. *et al.* Calyculin A and okadaic acid: Inhibitors of protein phosphatase activity. *Biochem. Biophys. Res. Commun.* **159**, 871–877 (1989).

108. Azzi, J. R., Sayegh, M. H. & Mallat, S. G. Calcineurin Inhibitors: 40 Years Later, Can't Live Without …. *J. Immunol.* **191**, 5785–5791 (2013).

109. Liu, J. *et al.* Inhibition of T Cell Signaling by Immunophilin-Ligand Complexes Correlates with Loss of Calcineurin Phosphatase Activity. *Biochemistry* **31**, 3896–3901 (1992).

110. Kimura, K. I., Aburai, N., Yoshida, M. & Ohnishi, M. Sanguinarine as a potent and specific inhibitor of protein phosphatase 2C in vitro and induces apoptosis via phosphorylation of p38 in HL60 cells. *Biosci. Biotechnol. Biochem.* **74**, 548–552 (2010).

111. Swatek, K. N. & Komander, D. Ubiquitin modifications. *Cell Res.* **26**, 399–422 (2016).

112. Hristova, V., Sun, S., Zhang, H. & Chan, D. W. Proteomic analysis of degradation ubiquitin signaling by ubiquitin occupancy changes responding to 26S proteasome inhibition. *Clin. Proteomics* **17**, 1–11 (2020).

113. Fosdahl, A. M. *et al.* ErbB3 interacts with Hrs and is sorted to lysosomes for degradation. *Biochim. Biophys. Acta - Mol. Cell Res.* **1864**, 2241–2252 (2017).

114. Udeshi, N. D. *et al.* Refined preparation and use of anti-diglycine remnant (k-ε-gg) antibody enables routine quantification of 10,000s of ubiquitination sites in single proteomics experiments. *Mol. Cell. Proteomics* **12**, 825–831 (2013).

115. Wagner, S. A. *et al.* A Proteome-wide, Quantitative Survey of In Vivo Ubiquitylation Sites Reveals Widespread Regulatory Roles. *Mol. Cell. Proteomics* **10**, M111.013284 (2011).

116. Kim, W. *et al.* Systematic and quantitative assessment of the ubiquitin-modified proteome. *Mol. Cell* **44**, 325–340 (2011).

117. Peng, J. *et al.* A proteomics approach to understanding protein ubiquitination. *Nat. Biotechnol.* **21**, 921–926 (2003).

118. David Meierhofer, Xiaorong Wang, Lan Huang, and P. K. Quantitative Analysis of global Ubiquitination in HeLa Cells by Mass Spectrometry. *J. Proteome Res.* **7**, 4566–4576 (2008).

119. Xu, G., Paige, J. S. & Jaffrey, S. R. Global analysis of lysine ubiquitination by ubiquitin remnant immunoaffinity profiling. *Nat. Biotechnol.* **28**, 868–873 (2010).

120. Danielsen, J. M. R. *et al.* Mass spectrometric analysis of lysine ubiquitylation reveals promiscuity at site level. *Mol. Cell. Proteomics* **10**, 1–12 (2011).

121. Sadanandom, A., Bailey, M., Ewan, R., Lee, J. & Nelis, S. The ubiquitin-proteasome system: Central modifier of plant signalling. *New Phytol.* **196**, 13–28 (2012).

122. Hicke, L. Protein regulation by monoubiquitin. *Nat. Rev. Mol. Cell Biol.* **2**, 195–201 (2001).

123. Chau, V. *et al.* Multiubiquitin Chain Is Confined to Specific Lysine in a Targeted

Short-Lived Protein. **243**, 1576–1583 (1989).

124. Glickman, M. H. & Ciechanover, A. The ubiquitin-proteasome proteolytic pathway: Destruction for the sake of construction. *Physiol. Rev.* **82**, 373–428 (2002).

125. Ohtake, F., Saeki, Y., Ishido, S., Kanno, J. & Tanaka, K. The K48-K63 Branched Ubiquitin Chain Regulates NF-κB Signaling. *Mol. Cell* **64**, 251–266 (2016).

126. Komander, D. & Rape, M. The Ubiquitin Code. *Annu. Rev. Biochem.* **81**, 203–229 (2012).

127. Vierstra, R. D. The ubiquitin-26S proteasome system at the nexus of plant biology. *Nat. Rev. Mol. Cell Biol.* **10**, 385–397 (2009).

128. Chen, J. *et al.* Autophagy Induced by Proteasomal DUB Inhibitor NiPT Restricts NiPT-Mediated Cancer Cell Death. *Front. Oncol.* **10**, 1–13 (2020).

129. Liu, C. Y. *et al.* The hippo tumor pathway promotes TAZ degradation by phosphorylating a phosphodegron and recruiting the SCFβ-TrCP E3 ligase. *J. Biol. Chem.* **285**, 37159–37169 (2010).

130. Ye, X. *et al.* Recognition of phosphodegron motifs in human cyclin E by the SCF Fbw7 ubiquitin ligase. *J. Biol. Chem.* **279**, 50110–50119 (2004).

131. Chong-Shan Shi, J. H. K. Traf6 and A20 Regulate Lysine 63-Linked Ubiquitination of Beclin 1 Controlling TLR4-Induced Autophagy Chong-Shan. *Sci. Signal.* **3**, 1–23 (2010).

132. Hill, S. M., Wrobel, L. & Rubinsztein, D. C. Post-translational modifications of Beclin 1 provide multiple strategies for autophagy regulation. *Cell Death Differ.* **26**, 617–629 (2019).

133. Wang, G. *et al.* K63-linked ubiquitination in kinase activation and cancer. *Front. Oncol.* **2**, 1–13 (2012).

134. Fisher, R. D. *et al.* Structure and ubiquitin binding of the ubiquitin-interacting motif. *J. Biol. Chem.* **278**, 28976–28984 (2003).

135. Levkowitz, G. *et al.* c-Cb1/Sli-1 regulates endocytic sorting and ubiquitination of the epidermal growth factor receptor. *Genes Dev.* **12**, 3663–3674 (1998).

136. Meister, M. *et al.* Regulation of cargo transfer between ESCRT-0 and ESCRT-I complexes by flotillin-1 during endosomal sorting of ubiquitinated cargo. *Oncogenesis* **6**, (2017).

137. Mueller, T. D., Kamionka, M. & Feigon, J. Specificity of the Interaction between Ubiquitin-associated Domains and Ubiquitin. *J. Biol. Chem.* **279**, 11926–11936 (2004).

138. Cao, X. *et al.* Histone H4K20 Demethylation by Two hHR23 Proteins. *Cell Rep.* **30**, 4152-4164.e6 (2020).

139. Poondla, N., Chandrasekaran, A. P., Kim, K. S. & Ramakrishna, S. Deubiquitinating enzymes as cancer biomarkers: New therapeutic opportunities? *BMB Rep.* **52**, 181–189 (2019).

140. Pfoh, R., Lacdao, I. K. & Saridakis, V. Deubiquitinases and the new therapeutic opportunities offered to cancer. *Endocr. Relat. Cancer* **22**, T35–T54 (2015).

141. March, E. & Farrona, S. Plant deubiquitinases and their role in the control of gene expression through modification of histones. *Front. Plant Sci.* **8**, 1–14 (2018).

142. Verma, S., Dixit, R. & Pandey, K. C. Cysteine proteases: Modes of activation and future prospects as pharmacological targets. *Front. Pharmacol.* **7**, 1–12 (2016).

143. Ndubaku, C. & Tsui, V. Inhibiting the deubiquitinating enzymes (DUBs). *J. Med. Chem.* **58**, 1581–1595 (2015).

144. Cowell, I. G., Ling, E. M., Swan, R. L., Brooks, M. L. W. & Austin, C. A. The deubiquitinating enzyme inhibitor PR-619 is a potent DNA topoisomerase II poison. *Mol. Pharmacol.* **96**, 562–572 (2019).

145. van der Wal, L. *et al.* Improvement of ubiquitylation site detection by Orbitrap mass spectrometry. *J. Proteomics* **172**, 49–56 (2018).

146. Goldberg, A. L. Development of proteasome inhibitors as research tools and cancer drugs. *J. Cell Biol.* **199**, 583–588 (2012).

147. Chen, J.-J., Chou, C.-W., Chang, Y.-F. & Chen, C.-C. Proteasome Inhibitors Enhance TRAIL-Induced Apoptosis through the Intronic Regulation of DR5: Involvement of NF-κB and Reactive Oxygen Species-Mediated p53 Activation. *J. Immunol.* **180**, 8030–8039 (2008).

148. Park, H. S., Jun, D. Y., Han, C. R., Woo, H. J. & Kim, Y. H. Proteasome inhibitor MG132-induced apoptosis via ER stress-mediated apoptotic pathway and its potentiation by protein tyrosine kinase p56 lck in human Jurkat T cells. *Biochem. Pharmacol.* **82**, 1110–1125 (2011).

149. Luo, D. *et al.* MG132 selectively upregulates MICB through the DNA damage response pathway in A549 cells. *Mol. Med. Rep.* **19**, 213–220 (2019).

150. Emmerich, C. H. & Cohen, P. Optimising methods for the preservation, capture and identification of ubiquitin chains and ubiquitylated proteins by immunoblotting. *Biochem. Biophys. Res. Commun.* **466**, 1–14 (2015).

151. Maor, R. *et al.* Multidimensional protein identification technology (Mud(PIT) analysis of ubiquitinated proteins in plants. *Mol. Cell. Proteomics* **6**, 601–610 (2007).

152. Hjerpe, R. *et al.* Efficient protection and isolation of ubiquitylated proteins using tandem ubiquitin-binding entities. *EMBO Rep.* **10**, 1250–1258 (2009).

153. Hjerpe, R. & Rodríguez, M. S. Efficient approaches for characterizing ubiquitinated proteins. *Biochem. Soc. Trans.* **36**, 823–827 (2008).

154. Zhang, N. *et al.* Comprehensive profiling of lysine ubiquitome reveals diverse functions of lysine ubiquitination in common wheat. *Sci. Rep.* **7**, 1–14 (2017).

155. Denis, N. J., Vasilescu, J., Lambert, J. P., Smith, J. C. & Figeys, D. Tryptic digestion of ubiquitin standards reveals an improved strategy for identifying ubiquitinated proteins by mass spectrometry. *Proteomics* **7**, 868–874 (2007).

156. Kim, D. Y., Scalf, M., Smith, L. M. & Vierstra, R. D. Advanced proteomic analyses yield a deep catalog of ubiquitylation targets in Arabidopsis. *Plant Cell* **25**, 1523–1540 (2013).

157. Xie, H. *et al.* Global Ubiquitome Profiling Revealed the Roles of Ubiquitinated Proteins in Metabolic Pathways of Tea Leaves in Responding to Drought Stress. *Sci. Rep.* **9**, 1–12 (2019).

158. Xie, X., Kang, H., Liu, W. & Wang, G. L. Comprehensive profiling of the rice ubiquitome reveals the significance of lysine ubiquitination in young leaves. *J. Proteome Res.* **14**, 2017–2025 (2015).

159. Giles, K., Williams, J. P. & Campuzano, I. Enhancements in travelling wave ion mobility resolution. *Rapid Commun. Mass Spectrom.* **25**, 1559–1566 (2011).

160. Richardson, K., Langridge, D. & Giles, K. Fundamentals of travelling wave ion mobility revisited: I. Smoothly moving waves. *Int. J. Mass Spectrom.* **428**, 71–80 (2018).

161. Paglia, G. & Astarita, G. Metabolomics and lipidomics using traveling-wave ion mobility mass spectrometry. *Nat. Protoc.* **12**, 797–813 (2017).

162. Shliaha, P. V., Bond, N. J., Gatto, L. & Lilley, K. S. Effects of traveling wave ion mobility separation on data independent acquisition in proteomics studies. *J. Proteome Res.* **12**, 2323–2339 (2013).

163. Bateman, R. H. *et al.* A novel precursor ion discovery method on a Hybrid Quadrupole Orthogonal Acceleration for Studying Protein Phosphorylation. *J. Am. Soc. Mass Spectrom.* **13**, 792–803 (2002).

164. Bond, N. J., Shliaha, P. V., Lilley, K. S. & Gatto, L. Improving qualitative and quantitative performance for MSE-based label-free proteomics. *J. Proteome Res.* **12**, 2340–2353 (2013).

CHAPTER 2

CHARACTERIZATION OF PHOSPHOPEPTIDE POSITIONAL ISOMERS ON THE TRANSCRIPTIONAL CO-ACTIVATOR TAZ

*Justin Roberto[1], Catherine E. Sykes[1], and Panayiotis O. Vacratsis[1,]**

[1]Department of Chemistry and Biochemistry, University of Windsor, 401 Sunset Ave., Windsor, Ontario N9B 3P4, Canada

INTRODUCTION

The transcriptional co-activator with PDZ-binding motif (TAZ) was originally identified as a novel binding partner of the highly ubiquitous 14-3-3 proteins.[1] TAZ functions as a transcriptional co-activator interacting with various transcription factors such as Runx2, PPARγ, TEAD, and PAX,[2-4] regulating a myriad of cellular processes including the epithelial-to-mesenchymal transition (EMT), neural crest development, stem cell differentiation, and progression of tumorigenesis.[5-8]

Due to these pleiotropic effects, understanding mechanisms regulating TAZ activity is paramount. One such mechanism is the phosphorylation dependent sequestration of TAZ in the cytoplasm by 14-3-3 proteins.[9] In response to stimulation of the Hippo signaling pathway, the serine/threonine kinase Mst1/2 phosphorylates and activates the large as tumour suppressor 1/2 kinase (LATS1/2), which phosphorylate TAZ at Ser^{89} and its paralog Yes associated protein (YAP) at Ser^{127}.[5] This phosphorylation event induces 14-3-3 association, sequestering TAZ/YAP in the cytoplasm, and thus spatially attenuating their transcriptional co-activation functions.[1] In the absence of Hippo pathway signaling, TAZ/YAP readily translocate to the nucleus and form heterocomplexes with various transcription factors to modulate expression of target genes.[10] Dysregulation of Hippo signaling has been implicated in a wide array of disorders including tumourigenesis and age-related diseases,[11] indicating the critical maintenance of homeostatic level TAZ/YAP activity.

In addition to Ser^{89} phosphorylation of TAZ, there is additional evidence that regulatory phosphorylation is likely a prevalent mechanism regulating the activity of TAZ. For example, mutagenesis studies have demonstrated that phosphorylation of Ser^{58} and

Ser[62] regulates an N-terminal phosphodegron motif,[12] while there is evidence that phosphorylation of Ser[311] and Ser[314] may regulate a C-terminal phosphodegron motif.[5] Surprisingly, there have been limited studies directly mapping phosphorylation sites on TAZ, with the majority of reported phosphorylation sites implicated by consensus site alignments and mutagenesis experiments.[5,13-15] To this end, we were interested in employing mass spectrometry-based methodology to directly sequence phosphorylation sites on TAZ.

MATERIALS AND METHODS

Plasmids

Wild-type TAZ cDNA (MGC-19891; ATCC) was subcloned into the pFLAG-CMV2 eukaryotic expression vector (Sigma) to generate FLAG-TAZ via PCR amplification using the following primers: forward, 5'-CGGAATTCGATGAATCCGGCCTCG-3' and reverse, 5'-CGTTAAGGATCCTTACAGCCAGGTTAG-3', which incorporated the *Eco*RI and *Bam*HI restriction enzyme recognition sites for subcloning.

Mutant constructs were generated via site-directed mutagenesis with cloned *Pfu* DNA Polymerase (Stratagene). FLAG-TAZ S93A was generated using the following primers: forward, 5'-GTCGCCCGCGGCCCTGCAGCTG-3' and reverse, 5'-CAGCTGCAGGGCCGCGGGCGAC-3'. The FLAG-TAZ S93D mutant was generated using the following primers: forward, 5'-GTCGCCCGCGGACCTGCAGCTGGG-3' and reverse, 5'-CCCAGCTGCAGGTCCGCGGGCGAC-3'. The 14-3-3 binding mutant of TAZ, FLAG-TAZ S89A, was produced using the following primers: forward, 5'-CTCGCACGCGTCGCCCGCGTC-3' and reverse, 5'-GACGCGGGCGACGCGTGCGAG-3'. All DNA constructs were verified by automated DNA sequencing (ACGT).

Cell culture

Human embryonic kidney (HEK) 293 cells were obtained from American Type Tissue Culture Collection (ATCC). HEK 293 cells were grown as a monolayer in Dulbecco's Modified Eagle's Medium/Nutrient F-12 Ham (D-MEM/F-12) (Invitrogen), supplemented with 10% fetal bovine serum (FBS) (Invitrogen), 2 mM L-glutamine, and

antibiotics (100 units/mL penicillin, 100 µg/mL streptomycin) (Sigma), at 5% CO_2 and 37˚C. For phosphorylation site mapping, FLAG-TAZ was overexpressed in HEK293 cells using lipid-mediated transfection with linear polyethylenimine (PEI) (25 kDa) (Sigma). Six hours post-transfection, cells were washed and replaced with fresh media. To increase the stoichiometry of phosphorylation, cells were treated with the phosphatase inhibitor calyculin A (Calbiochem) for 30 minutes prior to cell lysis. For serum starvation experiments, cells were incubated in serum free medium for 30 minutes prior to cell lysis. Cells were lysed in buffer containing 50 mM Tris-HCl (pH 7.4), 150 mM NaCl with 0.1% Triton X-100 supplemented with protease and phosphatase inhibitors. Soluble lysate was separated from cellular debris by centrifugation at 14,000 x g for 15 minutes at 4˚C.

Immunoprecipitation and in-gel trypsin digestion

FLAG-TAZ was isolated from cellular lysates using anti-FLAG M2 mouse monoclonal magnetic beads (Sigma) and an overnight incubation at 4˚C. Samples were washed three times in 50 mM Tris-HCl (pH 7.4), 150 mM NaCl with 1% Triton X-100 and resuspended in SDS/PAGE loading dye prior to analysis. During serum starvation experiments, immunoprecipitants were washed as above and eluted using FLAG peptide (100 µg/mL in 0.5 M Tris-HCl, 1 M NaCl, pH 7.5) (Genscript) for 30 minutes at 4˚C with shaking. Eluted proteins were boiled and reduced with SDS/PAGE loading dye.

Immunoprecipitated TAZ was resolved on 12% SDS/PAGE gels and stained with Imperial Coomassie Blue stain (Thermo Scientific) overnight. Gels were destained with Milli Q water (Millipore) and bands corresponding to TAZ were excised from the gel with a scalpel and cut into smaller pieces. Gel pieces were destained in a 2:1 acetonitrile/ 50 mM ammonium bicarbonate (Fisher) solution, at 37˚C for 15 minutes. Gel pieces were then

dehydrated and dried by addition of the 100% acetonitrile for 10 minutes at room temperature. Shrunken gel pieces were then dried via vacuum centrifugation (Thermo Scientific). Gel pieces were rehydrated in trypsin digestion buffer containing 13 ng/µl sequence grade trypsin (Promega) in 50 mM ammonium bicarbonate (pH 8) and incubated overnight at 37˚C. Peptides were extracted by adding 150 µL of extract buffer (2:1 acetonitrile : 5% formic acid (Sigma)) and incubated at 37˚C for 15 minutes. Extraction was repeated once. Peptide extracts were concentrated via rotovaping and reconstituted in 0.1% trifluoroacetic acid (TFA) (Sigma) in mass spectrometry grade water (Fisher). Following extraction, peptides were purified using 100 cc HLB Oasis columns (Waters). Briefly, columns were activated with 100% acetonitrile, and equilibrated with 0.1% TFA. Following peptide loading, columns were washed with 0.1% TFA and peptides were finally eluted with 60% acetonitrile and 0.1% TFA. Desalted peptides were dried to completion and reconstituted in 0.2% formic acid.

Mass Spectrometry

Peptides were analyzed on a Waters Synapt G2-S*i* mass spectrometer interfaced with a nanoAcquity ultra-performance liquid chromatography (UPLC) system. For initial phosphopeptide identification, peptides (4 µl) were loaded onto a Waters HSS T3 75 µm x 150 µm C18 reverse phase column at a flow rate of 0.3 µL/minute. Peptides were separated using a 3-85 % acetonitrile gradient in 0.1% formic acid over 90 minutes. Eluting peptides were analyzed in DIA mode with ion mobility separation ($HDMS^E$), alternating between low and high energy scans in positive ion mode. Glu1-fibrinopeptide B (m/z 785.8427) was used as a lock mass external calibrant. Data was collected using MassLynx (version 4.1) and analyzed using ProteinLynx Global Server. For the Parallel Reaction Monitoring

targeted mass spectrometry measurements, synthetic phosphopeptides representing peptides 87-114 of TAZ (Bio Basic) or TAZ tryptic peptides generated from immunoprecipitated TAZ from cells were analyzed using the high definition multiple reaction monitoring mode (HD-MRM) option available on the SYNAPT system. The predominant +4 charge state (m/z 694.3334) of the 87-114 positional phosphopeptide isomers was selected for targeted analysis. Wideband Enhancement Fragmentation (WEF) was employed to improve sensitivity of diagnostic fragment ions and reduce complexity by focusing on singly charged fragment ions. Calibration of pusher pulse synchronization with ion mobility drift time was accomplished with Glu1-fibrinopeptide B (m/z 785.8427).

Co-Immunoprecipitation

HEK 293 cells transfected with cDNA encoding FLAG-TAZ phosphovariants were incubated for 24 hours at 37 ˚C. Cells were then lysed and immunoprecipitated with anti-FLAG magnetic beads as described above. Samples were analyzed by immunoblotting using mouse anti-FLAG (Sigma), rabbit anti-1433 (Santa Cruz), rabbit anti-actin (Sigma), and rabbit anti-phospho YAP (Ser^{127}) (Cell Signaling) to detect phosphorylation at Ser^{89} of TAZ. Densitometry analysis was performed using Image J. Measurements were normalized to wild type FLAG-TAZ levels. Statistical analysis was performed using the Student's *t*-test (GraphPad), with P-values < 0.05 being considered statistically significant.

Immunofluorescence microscopy

U2OS cells were seeded at 80, 000 cells/mL in 1 mL chambers (BioBasic). Cells were transfected with Lipofectamine® 3000 (Invitrogen) and 2 µg of FLAG-TAZ-phosphovariants S89A, S93A, S93D, and wildtype in antibiotic-free media. Six hours post-transfection, cells were washed with PBS and fresh media was added to cells. Following

an 18-hour incubation at 37 ˚C, cells were washed with PBS twice and fixed using 3.7% paraformaldehyde (Fisher) for 15 minutes at room temperature and washed with PBS. Cells were permeabilized using 0.15% Triton X-100 in TBS for 2 minutes at room temperature. Cells were blocked for 1 hour at room temperature with 5% BSA in PBS. Mouse anti-FLAG (Sigma) was diluted to 1:500 in TBS supplemented with 1% BSA and primary antibody incubation occurred for 1.5 hours at room temperature in 1% BSA. Cells were washed three times with PBS and incubated with horse anti-mouse-FITC secondary antibody (Vector Labs) for 1 hour at room temperature. Once dry, coverslips were mounted with 50% glycerol and sealed. Fluorescence microscopy images were generated on a Leica DMI6000 using a 40X objective. Nuclear/cytoplasmic ratio measurements were recorded using the Intensity Ratio Nuclei Cytoplasm plugin within Image J (NIH). Measurements were performed on 30 cells per sample from three independent experiments. Statistical analysis was performed using the student's t-test (GraphPad), with differences considered statistically significant at P values < 0.001.

RESULTS AND DISCUSSION

HEK 293 cells expressing FLAG-TAZ were treated with and without the phosphatase inhibitor calyculin A to see the effect of limiting phosphorylation turnover on TAZ by immunoblot analysis. As shown in **Figure 2.1A**, calyculin A treatment induced a significant mobility shift on wild type TAZ and to a lesser extent TAZ S89A. The observation that mutation of the well characterized 14-3-3 associated phosphorylation site didn't abolish the mobility shift indicates that TAZ is likely modified at additional phospho-acceptor sites. To initiate comprehensive phosphorylation site mapping, FLAG-TAZ was immunoprecipitated from 293 cells using FLAG magnetic affinity resin. TAZ peptides were generated from a Coomassie-stained protein band using in-gel trypsin digestion and prepared for mass spectrometry analysis using solid phase extraction purification. Extracted peptides were separated on a nanoAcquity ultra-performance liquid chromatography (UPLC) system interfaced in-line with a SYNAPT G2S*i* quadrupole-time of flight mass spectrometer for tandem mass spectrometry analysis. In order to improve peptide separation and maximize sequence coverage, we employed the ion mobility feature of the SYNAPT system and conducted the analysis in data independent acquisition mode (DIA).[16] Five biological replicates were performed and phosphopeptides were reported if reproducibly observed in at least three biological replicates (**Appendix A**).

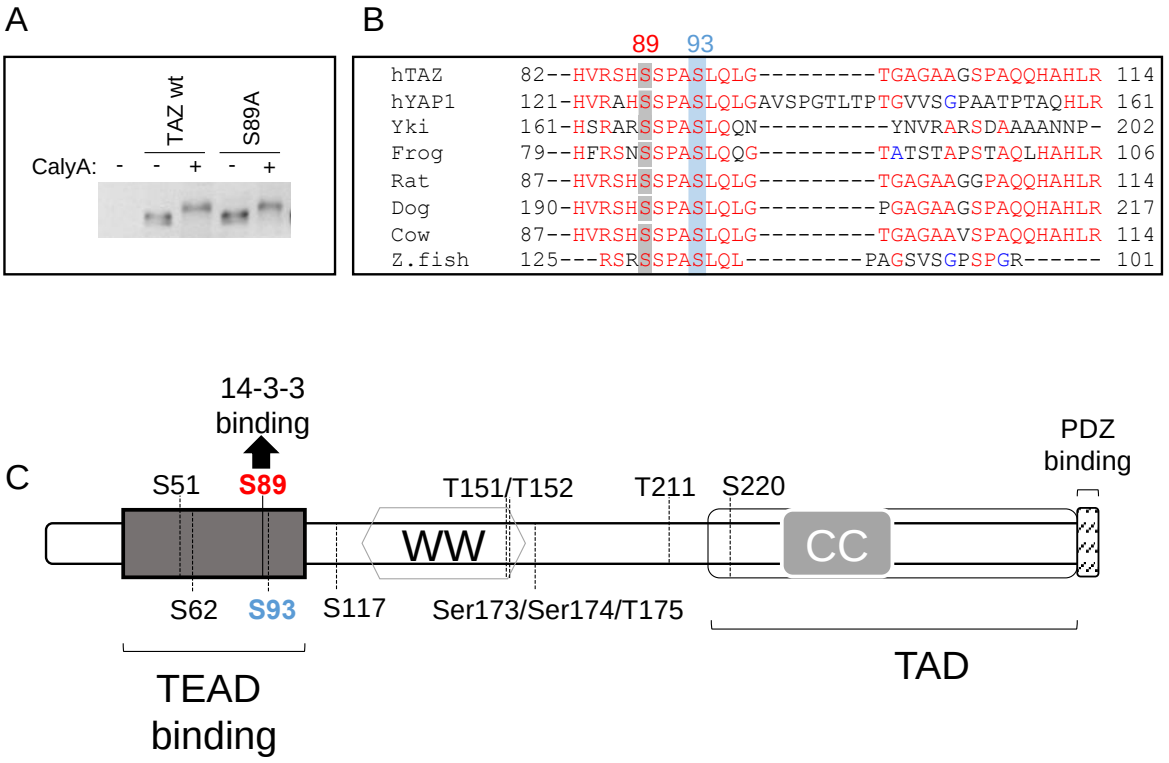


Figure 2.1. TAZ is a multiphosphorylated protein.

(**A**) Immunoblot analysis of HEK 293 cells transiently transfected with FLAG-TAZ and FLAG-TAZ S89A with and without calyculin A treatment (CalyA).

(**B**) Conservation of the Ser^{89} and Ser^{93} phosphorylation sites amongst TAZ orthologues and human YAP. Shown is a multiple sequence alignment of the TAZ tryptic peptide representing amino acids 87-114 of human TAZ. (**C**) Schematic of phosphorylation sites identified by data independent acquisition mass spectrometry and mapped onto a domain schematic representation of the TAZ protein.

Collectively, a number of phosphopeptides were identified including several novel phosphorylation sites mapped to TAZ functional domains and motifs (**Figure 2.1C**). Most notable was the dynamic phosphorylation of a peptide (amino acids 87-114) containing the 14-3-3 binding site ($pSer^{89}$) (**Figure 2.1B**). Processing of the DIA data from numerous replicates concluded the presence of positional phosphoisomers in the singly phosphorylated 87-114 peptide with Ser^{93} being a putative phosphorylation site in addition to the eminent phosphorylation site at Ser^{89}. Moreover, we reproducibly detected a doubly modified form of this peptide in which the processing algorithm assigned phosphorylation at Ser^{89} and Ser^{93}.

The proximity of the Ser^{93} phosphorylation site to the functional Ser^{89} phosphorylation site is an intriguing finding. Prior to examining the functional consequence of Ser^{93} phosphorylation, we wanted to confirm accurate mapping of the positional phosphoisomers. DIA analysis is useful for collecting comprehensive MS data that potentially can be passed over during data dependent acquisition (DDA) modes.[17] However, unambiguous assignment of PTM sites can be challenging if the fragmentation data from the MS2 scan is overly complex. Therefore, we employed a targeted DDA approach to positively confirm our DIA data and also resolve diagnostic fragment ion masses for the individual phosphorylation events. We initiated these studies by analysing synthetic phosphopeptides with the identical amino acid sequence as the 87-114 TAZ peptide with single phosphorylation modifications at sites corresponding to either Ser^{89} or Ser^{93}. This allowed for optimization of chromatography conditions, ion mobility conditions, and collision energy parameters. We attempted to resolve the positional phosphoisomers using optimized chromatography and ion mobility conditions. However,

under all conditions tested, the phosphorylated peptides co-eluted and displayed very similar drift times. We thus transitioned our focus towards elucidating diagnostic fragment ions that would be suitable to confirm phosphorylation site assignment within a mixture of the positional phosphoisomer peptides. This was accomplished by application of the so-called wideband enhancement fragmentation (WEF) technique.[18] WEF attempts to synchronize the exiting of ions from the ion mobility separation stage and the timing of the pusher pulse into the time of flight mass analyser. This can potentially improve the duty cycle of the instrument and increase sensitivity of fragment ion detection. Following optimization and WEF calibration, we compared the fragmentation of the synthetic phosphopeptide positional isomers with or without WEF activation. As seen in **Figure 2.2**, fragmentation under WEF conditions was far superior for both phosphopeptides allowing for identification of diagnostic fragment ions that could uniquely identify phosphorylation of Ser^{89} and Ser^{93}. As was evident in the DIA data, the m/z values of y-ions allowed for clear monitoring of the phosphorylation status of Ser/Thr residues in the C-terminal half of the peptide. In contrast, close proximity of several Ser residues in the N-terminal half of the peptide required specific b-ions to be detectable for confident phosphorylation site assignment. The presence of a fragment ion at m/z 225 in both phosphoisomers corresponds to the b_2 ion and is diagnostic for unphosphorylated Ser^{87}. Moreover, the presence of fragment ions at m/z 392 (b_3 + phosphorylation) and m/z 294 (b_3 - H_3PO_4) are diagnostic for phosphorylation at Ser^{89}. Finally, in combination with the C-terminal y-ions, detection of m/z 399 was the most intense diagnostic ion for phosphorylation of Ser^{93} as this fragment ion corresponds to unphosphorylated b_4, which rules out phosphorylation at serines 87, 89, and 90, leaving Ser^{93} as the only remaining phosphorylatable site.

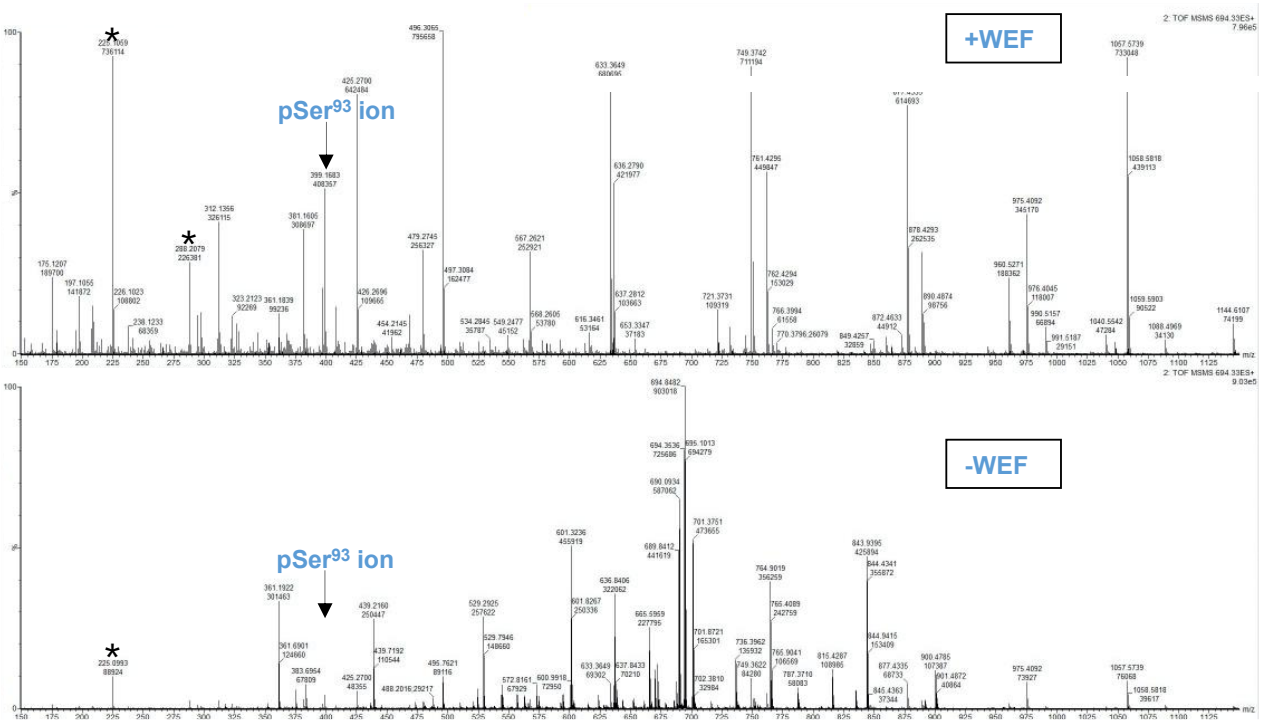
Importantly, the detection of the diagnostic fragment ions was significantly and reproducibly elevated when WEF was activated demonstrating the utility of this acquisition mode for increasing confidence in mapping post translational modifications including phosphorylation sites.

Having identified diagnostic fragment ions for phosphorylation of Ser^{89} and Ser^{93} on synthetic peptides, we next applied the WEF method to a biological sample, isolating full length TAZ from 293 cells under both serum and serum starved conditions. It has been shown that lysophosphatidic acid (LPA) and sphingosine-1-phosphate (S1P) present in serum induces G-protein coupled receptor signaling leading to the inhibition of LATS1/2,[19] the kinase responsible for phosphorylating Ser^{89} of TAZ. Therefore, application of the WEF method on TAZ isolated from cells subjected to serum deprivation was employed to confirm the diagnostic fragment ions in a biological sample and resolve alterations in Ser^{89} phosphorylation in response to an extracellular stimulus.

As shown in **Figure 2.3A**, immunoblot analysis using a phosphoantibody that has specificity for phosphorylated Ser^{89} of TAZ revealed an increase in Ser^{89} phosphorylation in response to serum deprivation. The remaining sample was subjected to anti-FLAG immunoprecipitation to enrich for FLAG-TAZ and processed for mass spectrometry analysis using the WEF acquisition method (**Figure 2.3B**). Importantly, the diagnostic fragment ions were detected indicating the optimized WEF method performed on synthetic peptides is transferable to full length TAZ generated from cellular samples. Moreover, relative quantitation between differential serum conditions was measured by analyzing multiple biological replicates and normalizing the relative intensity of phospho-diagnostic fragment ions against a prominent unphosphorylated fragment ion (m/z 225, b_2).

Interestingly, phosphorylation of Ser^{89} increased approximately 63% in response to the serum deprivation conditions used. In contrast phosphorylation of Ser^{93} under these conditions decreased approximately 35% suggesting that the two phosphorylation events are independently regulated (**Figure 2.3B, C**).

A



B

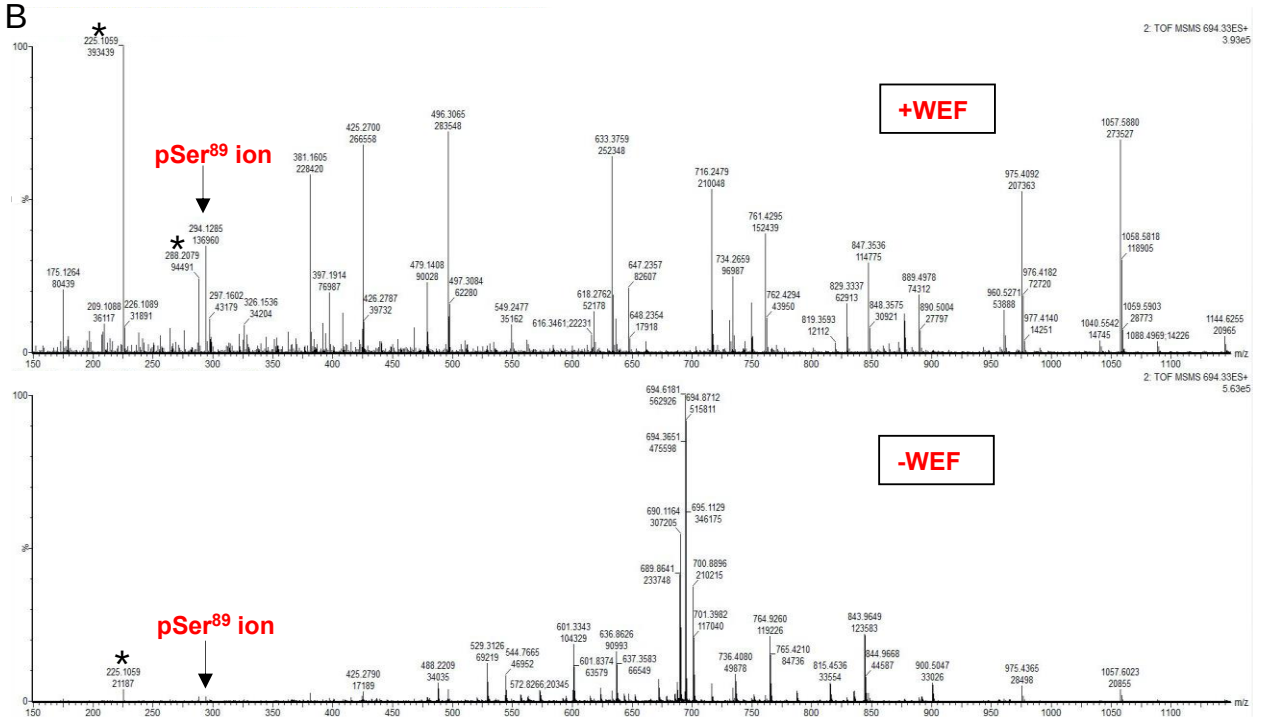
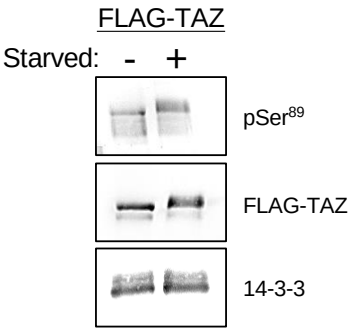


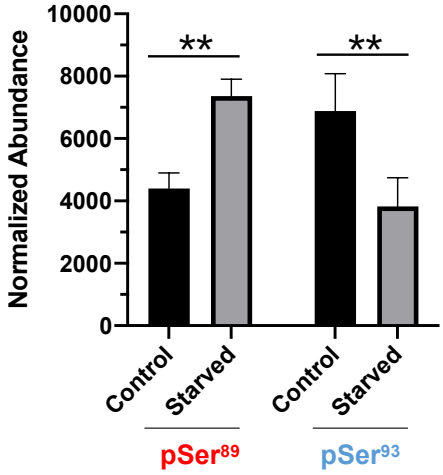
Figure 2.2. Wideband enhancement fragmentation technique dramatically improves sequencing of positional phosphopeptide isomers.

Product ion spectra for synthetic phosphopeptides (1ng on column) representing amino acids 87-114 of TAZ phosphorylated at Ser^{93} (**A**) or Ser^{89} (**B**). Targeted parallel reaction monitoring measurements were acquired with (upper panels) and without (lower panels) Wideband Enhancement (WEF) synchronization of the ion pusher for singly charged fragment ions. Arrows indicate diagnostic fragment ions specific for Ser^{93} and Ser^{89} phosphorylation. Asterisks indicate unphosphorylated fragment ions common to the positional phosphopeptide isomers.

A


FLAG-TAZ
Starved: - +
pSer89
FLAG-TAZ
14-3-3

B


Normalized Abundance
10000
8000
6000
4000
2000
0
**
**
Control
Starved
Control
Starved
pSer89
pSer93

C

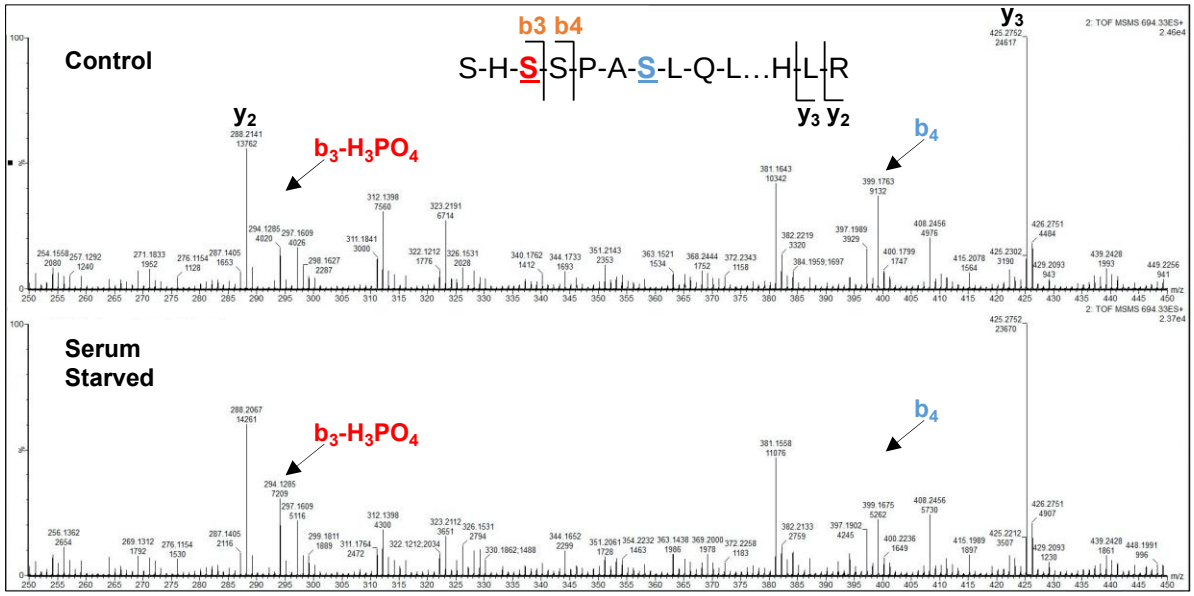

Control
b3 b4
S-H-S-S-P-A-S-L-Q-L…H-L-R
y3 y2
y3
y2
b3-H3PO4
b4
Serum
Starved
b3-H3PO4
b4

Figure 2.3. Serum deprivation increases Ser^{89} phosphorylation and decreases Ser^{93} phosphorylation.

FLAG-TAZ was immunoprecipitated from HEK 293 that were grown in 10% FBS (Control) or serum deprived for 30 min (Starved). (**A**) Measurement of Ser^{89} phosphorylation in response to serum deprivation was monitored by immunoblot analysis using a phosphoantibody with specificity towards Ser^{89}. (**B**) Quantitation of diagnostic fragment ions specific for phosphorylation of Ser^{89} (left) or Ser^{93} (right) in response to serum deprivation. Integrated areas were normalized to specific unphosphorylated fragment ions within the 87-114 peptide. Results from three biological replicates were analyzed by a two tailed t-test, (**$P < 0.05$ was considered significant). (**C**) Representative product ion spectra displaying the alteration in diagnostic fragment ions ($pSer^{89}$, b_3-H_3PO_4) and ($pSer^{93}$, b_4) in response to serum deprivation. Each spectrum is the summed result of 30 cumulative scans.

The LATS1/2 phosphorylation of Ser^{89} mediates 14-3-3 association and results in cytoplasmic sequestration of TAZ. This sequestration ability is a critical mechanism regulating TAZ transcriptional co-activation activities.[20] The close proximity of Ser^{93} to Ser^{89} led us to hypothesize that phosphorylation of Ser^{93} may regulate the phosphorylation of Ser^{89} and/or the association of 14-3-3. Phosphomimetic mutants were generated at Ser^{93} and analyzed alongside wild-type TAZ (FLAG-TAZ) and the 14-3-3 binding mutant of TAZ (FLAG-TAZ S89A) in HEK 293 cells. The resulting anti-FLAG immune complexes were resolved by SDS/PAGE and subjected to immunoblot analysis (**Figure 2.4A**). As expected, FLAG-TAZ S89A was unable to co-immunoprecipitate 14-3-3 to any detectable level compared to wild type TAZ. The phosphomutant FLAG-TAZ S93A was observed to co-immunoprecipitate 14-3-3 to a similar extent as wild-type TAZ. Interestingly, the ability of FLAG-TAZ S93D to associate with 14-3-3 was significantly diminished (**Figure 2.4A**, upper panel), suggesting the function of Ser^{93} phosphorylation may be to regulate the 14-3-3 sequestration of TAZ. To test if Ser^{93} phosphorylation mediates its affect by attenuating Ser^{89} phosphorylation, the phospho-Ser^{89} levels were measured by immunoblot analysis (**Figure 2.4A**, second panel). Surprisingly, TAZ S93D did not display significant alteration of Ser^{89} phosphorylation, suggesting that phosphorylation of Ser^{93} disrupts the TAZ-14-3-3 interaction independently of phosphorylation of Ser^{89}. Collectively, these findings suggest that the phosphorylation of Ser^{93} attenuates the interaction between TAZ and 14-3-3 and may represent a novel mechanism regulating TAZ function through reversal of TAZ cytoplasmic sequestration.

The transcriptional coactivating function of TAZ requires the protein to localize to the nucleus where it partners with various transcription factors to regulate gene expression

in response to various extracellular stimuli.[21] The LATS1/2 kinase mediated phosphorylation of Ser^{89} provides a binding motif for 14-3-3 and promotes sequestration of TAZ in the cytoplasm.[5] Therefore, cytoplasmic/nuclear localization determination is considered a suitable readout for examining regulators of TAZ biological function. To test the effects of phosphorylation of Ser^{93} on TAZ localization, we analyzed the subcellular localization of FLAG-TAZ phosphomimetic mutants by fluorescent microscopy (**Figure 2.4D,E**). As expected, TAZ S89A displays stronger nuclear localization when compared to wild type TAZ, demonstrating the central role phosphorylation of Ser^{89} has on the nuclear/cytoplasmic distribution of TAZ. The subcellular localization pattern of TAZ S93A was similar to wild type TAZ showing both cytoplasmic and nuclear localization. We found that cells expressing TAZ S93D displayed an increased nuclear localization pattern compared to wild type TAZ, consistent with the results from the co-immunoprecipitation studies. Taken together, the localization patterns support the model whereby phosphorylation of Ser^{93} promotes liberation of 14-3-3 association and increased nuclear localization. However, it was clear that the nuclear localization pattern of TAZS93D was not as pronounced as TAZS89A. This may indicate that phosphorylation of Ser^{93} is not sufficient for nuclear translocation, while other mechanistic steps (e.g. dephosphorylation of Ser^{89}) are required for efficient nuclear targeting of TAZ.

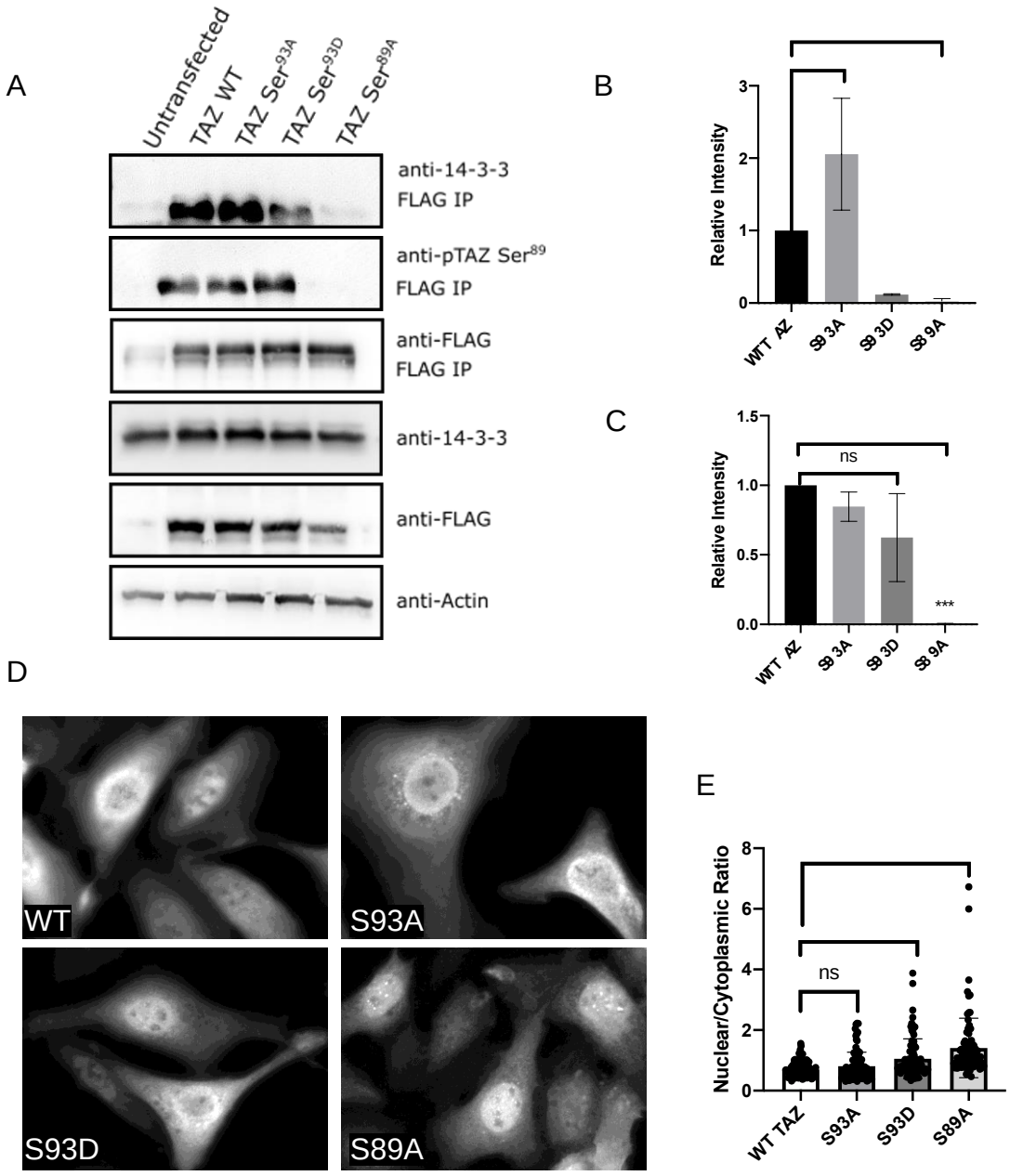


Figure 2.4. Phosphorylation of Ser[93] attenuates TAZ association with 14-3-3.

(**A**) Co-immunoprecipiatation of FLAG-TAZ phosphovariants with 14-3-3. HEK 293 cells expressing the indicated FLAG-TAZ phosphovariant, were subjected to anti-FLAG immunoprecipitation and probed with anti-14-3-3 (upper panel), anti-phospho Ser[89] (second panel) and anti-FLAG (third panel). Immunoblots of whole cell lysates as loading controls were probed with the indicated antibodies. Co-immunoprecipitation of 14-3-3 (**B)**

and pSer^{89} (**C**) was quantitated by densitometry and normalized to wild type FLAG-TAZ levels. Means ± standard deviation from three independent experiments are shown. P-values were calculated using a Student's t-test and compared to wild type TAZ (*P < 0.05, **** P < 0.001). (**D**) Cytoplasmic/nucleus distribution of TAZ phosphovariants. U2OS cells expressing the indicated FLAG-TAZ phosphovariant were analyzed by immunofluorescence microscopy. Images were collected using a 40X objective. The scale bars are 10 μm. (**E**) The Intensity Ratio Nuclei Cytoplasm tool within Image J was utilized to quantify nucleocytoplasmic alterations between phosphovariants of TAZ. Means ± standard deviation of three independent experiments (n=30 cells) is shown. P-values were calculated by a Student's *t*-test and compared with wild type TAZ, with differences considered statistically significant at P values < 0.001 (****).

CONCLUSION

In conclusion, we have performed a comprehensive mass spectrometry-based screen of phosphorylation sites on the transcriptional co-activator TAZ. With over a dozen sites confirmed under the conditions used, it is clear that TAZ is a multiphosphorylated protein. Detailed studies examining cellular context, extracellular stimuli, and participating kinases/phosphatases of the individual modification events will likely yield novel mechanistic insights into how reversible phosphorylation dynamically regulates the pleiotropic biological activities of TAZ.

REFERENCES

1. Kanai, F., Marignani, P. A., Sarbassova, D., Yagi, R., Hall, R. A., Donowitz, M., Hisaminato, A., Fujiwara, T., Ito, Y., Cantley, L. C., & Yaffe, M. B. (2000). TAZ: A novel transcriptional co-activator regulated by interactions with 14-3-3 and PDZ domain proteins. *EMBO Journal*, ***19***(24), 6778–6791.

2. Cui, C. Bin, Cooper, L. F., Yang, X., Karsenty, G., & Aukhil, I. (2003). Transcriptional Coactivation of Bone-Specific Transcription Factor Cbfa1 by TAZ. *Molecular and Cellular Biology*, ***23***(3), 1004–1013.

3. Zhang, H., Liu, C. Y., Zha, Z. Y., Zhao, B., Yao, J., Zhao, S., Xiong, Y., Lei, Q. Y., & Guan, K. L. (2009). TEAD transcription factors mediate the function of TAZ in cell growth and epithelial-mesenchymal transition. *Journal of Biological Chemistry*, ***284***(20), 13355–13362.

4. Murakami, M., Tominaga, J., Makita, R., Uchijima, Y., Kurihara, Y., Nakagawa, O., Asano, T., & Kurihara, H. (2006). Transcriptional activity of Pax3 is co-activated by TAZ. *Biochemical and Biophysical Research Communications*, ***339***(2), 533–539.

5. Lei, Q.-Y., Zhang, H., Zhao, B., Zha, Z.-Y., Bai, F., Pei, X.-H., Zhao, S., Xiong, Y., & Guan, K.-L. (2008). TAZ Promotes Cell Proliferation and Epithelial-Mesenchymal Transition and Is Inhibited by the Hippo Pathway. *Molecular and Cellular Biology*, ***28***(7), 2426–2436.

6. Manderfield, L. J., Engleka, K. a, Aghajanian, H., Gupta, M., Yang, S., Li, L., Baggs, J. E., Hogenesch, J. B., Olson, E. N., & Epstein, J. a. (2014). Pax3 and Hippo Signaling Coordinate Melanocyte Gene Expression in Neural Crest. *Cell Reports*, ***9***(5), 1885–1895.

7. Guo, L., Cai, T., Chen, K., Wang, R., Wang, J., Cui, C., Yuan, J., Zhang, K., Liu, Z., Deng, Y., Xiao, G., & Wu, C. (2018). Kindlin-2 regulates mesenchymal stem cell differentiation through control of YAP1/TAZ. *Journal of Cell Biology*, ***217***(4), 1431–1451.

8. Li, Z., Wang, Y., Zhu, Y., Yuan, C., Wang, D., Zhang, W., Qi, B., Qiu, J., Song, X., Ye, J., Wu, H., Jiang, H., Liu, L., Zhang, Y., Song, L. N., Yang, J., & Cheng, J. (2015). The Hippo transducer TAZ promotes epithelial to mesenchymal transition and cancer stem cell maintenance in oral cancer. *Molecular Oncology*, ***9(***6), 1091–1105.

9. Freeman, A. K., & Morrison, D. K. (2011). 14-3-3 Proteins: Diverse functions in cell proliferation and cancer progression. *Seminars in Cell and Developmental Biology*, ***22***(7), 681–687.

10. Remue, E., Meerschaert, K., Oka, T., Boucherie, C., Vandekerckhove, J., Sudol, M., & Gettemans, J. (2010). TAZ interacts with zonula occludens-1 and -2 proteins in a PDZ-1 dependent manner. *FEBS Letters*, ***584***(19), 4175–4180.

11. Yeung, Y. T., Guerrero-Castilla, A., Cano, M., Muñoz, M. F., Ayala, A., & Argüelles, S. (2019). Dysregulation of the Hippo pathway signaling in aging and cancer. *Pharmacological Research*, ***143***(February), 151–165.

12. Huang, W., Lv, X., Liu, C., Zha, Z., Zhang, H., Jiang, Y., Xiong, Y., Lei, Q. Y., & Guan, K. L. (2012). The N-terminal phosphodegron targets TAZ/WWTR1 protein for SCF

β-TrCP-dependent degradation in response to phosphatidylinositol 3-kinase inhibition. *Journal of Biological Chemistry*, ***287***(31), 26245–26253.

13. Liu, C. Y., Zha, Z. Y., Zhou, X., Zhang, H., Huang, W., Zhao, D., Li, T., Chan, S. W., Lim, C. J., Hong, W., Zhao, S., Xiong, Y., Lei, Q. Y., & Guan, K. L. (2010). The hippo tumor pathway promotes TAZ degradation by phosphorylating a phosphodegron and recruiting the SCFβ-TrCP E3 ligase. *Journal of Biological Chemistry*, ***285***(48), 37159–37169.

14. Jang, E. J., Jeong, H., Han, K. H., Kwon, H. M., Hong, J.-H., & Hwang, E. S. (2012). TAZ Suppresses NFAT5 Activity through Tyrosine Phosphorylation. *Molecular and Cellular Biology*, ***32***(24), 4925–4932.

15. Zhang, L., Chen, X., Stauffer, S., Yang, S., Chen, Y., & Dong, J. (2015). CDK1 phosphorylation of TAZ in mitosis inhibits its oncogenic activity. *Oncotarget*, ***6(***31), 31399–31412.

16. Steevensz, A., Gombar, R., Vergilino, R., Cristescu, M. E., Vacratsis, P. O. (2018) Proteomic Profile of Daphnia pulex using Data-Independent Acquisition Mass Spectrometry and Ion Mobility Separation. *Proteomics **18***, e1700460.

17. Gillet, L. C., Navarro, P., Tate, S., Röst, H., Selevsek, N., Reiter, L., Bonner, R., & Aebersold, R. (2012). Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: A new concept for consistent and accurate proteome analysis. *Molecular and Cellular Proteomics*, ***11***(6), 1–17.

18. Helm, D. et al. (2014) Ion Mobility Tandem Mass Spectrometry Enhances Performance of Bottom-Up Proteomics, *Molecular and Cellular Proteomics **13***(12), 3709-15.

19. Yu, F., Zhao, B., Panupinthu, N., Jewell, J. L., Lian, I., Lloyd, H., Zhao, J., Yuan, H., Tumaneng, K., Li, H., & Fu, X. (2013). Regulation of the Hippo-YAP pathway by G-protein coupled receptor signaling. *Cell*, ***150***(4), 780–791.

20. Zhou, X., & Lei, Q. Y. (2016). Regulation of TAZ in cancer. *Protein and Cell*, ***7(***8), 548–561.

21. Totaro, A., Panciera, T., & Piccolo, S. (2018). YAP/TAZ upstream signals and downstream responses. *Nature Cell Biology*, ***20***(8), 888–899.

CHAPTER 3

LABEL-FREE QUANTITATIVE PROTEOMIC ANALYSIS OF EXTRACELLULAR VESICLES DERIVED FROM MUSCULAR ATROPHY CELLS

Justin Roberto,[1] Kathy L. Poulin,[2] Robin Parks,[2] and Panayiotis O. Vacratsis[1]

[1] Department of Chemistry and Biochemistry, University of Windsor, Windsor, Ontario, Canada N9B 3P4

[2] Regenerative Medicine Program, Ottawa Hospital Research Institute, Ottawa, Ontario, Canada K1H 8L6

[*]Correspondence should be addressed to:

Panayiotis O. Vacratsis

Department of Chemistry and Biochemistry, University of Windsor

Windsor, Ontario, Canada N9B 3P4

E-mail: vacratsi@uwindsor.ca

INTRODUCTION

Spinal muscular atrophy (SMA) is an inherited neuromuscular disorder leading to degeneration of spinal motor neurons and muscle atrophy. The loss of motor neurons in the brain stem and spinal cord lead to muscle weakness, and in the most severe form of the disease in infants leads to death.[1] SMA is inherited in an autosomal recessive manner, where approximately 1 in 50 people are carriers affecting 1 in 10,000 newborns.[2] SMA is divided into 3 pediatric types based on age of onset and severity of disease.[3] Type I SMA with onset as early as birth, often is lethal within the first 2 years of life.[4] Type II SMA, the intermediate form, is typically diagnosed within the first few years of life with patients displaying normal gait and decreased life expectancy of childhood to late adulthood depending on disease severity. Type III SMA, is the least severe form of pediatric SMA detected at ~18 months, with patients having a normal life expectancy but experiencing various neuromuscular deficiencies.[5] There is also an adult-onset form of the disease, SMA type IV, that usually manifests in the second or third decade of life.

Genetically SMA is caused by mutation of the survival of motor neuron 1 (*smn1*) gene causing low amounts of the full-length protein. Functional SMN is a 294 amino acid protein, which has a prevalent role in small nuclear ribonucleoprotein (snRNP) assembly.[5,6] The exact mechanism of how SMN depletion causes SMA is unknown, as it is involved in a variety of neuronal functions. Diagnosis of SMA afflicted patients occurs through genetic testing, a majority identifying with homozygous loss of both *smn1* copies, or heterozygotes with one *smn1* deletion and mutations in the other copy.[5] In these patients, disease severity inversely correlates with the amount of SMN protein produced from the paralogous SMN2 gene. SMN2 contains a single nucleotide substitution that affects mRNA splicing and

primarily produces a truncated, unstable protein. However, the SMN2 locus is variably amplified and the small amount of full-length protein that is produced from each copy of the SMN2 gene brings the patient closer to normal levels.

New treatment and biomarker detection strategies are required to treat and diagnose those afflicted with disease. One biomarker detection method which has shown promise for a variety of disorders, including SMA,[8] is the isolation and examination of extracellular vesicles (EVs). EVs are secreted by a wide variety of cells and contain a mixture of lipids, RNA, and proteins as cargo. EVs can be isolated from a variety of bodily fluids and are approximately 30-100 nm in size encased within a lipid membrane.[7] EVs can participate in cell-cell communication shuttling cargo to other cells, influencing a myriad of biological processes.[9] Early studies heavily focused on the presence of miRNA found within EVs and their potential for both diagnosis and treatment.[9-10] To date, EV biomarkers for diseases including amyotrophic lateral sclerosis (ALS), kidney disease and various cancers have shown increasing promise as tools for understanding disease mechanisms and as prognostic monitoring tools.[11-14]

Proteomic profiling of EV-derived, disease-relevant samples is an emerging field for discovering potential disease biomarkers, therapeutic targets, and advancing mechanistic understanding of disease progression. In this study, we have utilized label-free quantitative proteomics to identify differences in EV protein cargo produced by fibroblasts derived from patients with SMA compared to unaffected volunteer control cells. Approximately 100 statistically significant protein alterations were identified including a number of proteins that have been previously implicated in the etiology of SMA.

MATERIALS AND METHODS

Cell culture

Human fibroblast cell lines were obtained from Coriell Institute for Medical Research (Camden, New Jersey). These cell lines include AG08498 (normal control, male, 1 YR – age of the individual when the fibroblast sample was obtained); AG07095 (normal control, male, 2 YR); GM05565 (normal control, male, 3 YR); GM00232 (SMA type I, male, 7 MO, homozygous deletion of exons 7 and 8 of the SMN1 gene, 2 copies of SMN2 gene); GM22592 (SMA type II, male, 1 YR, homozygous deletion of exons 7 and 8 of the SMN1 gene, 3 copies of SMN2 gene); and GM03813 (SMA type II, male, 3 YR, homozygous deletion of exons 7 and 8 of the SMN1 gene, 3 copies of SMN2 gene). We also utilized GM03814, the clinically unaffected mother of GM03813 (1 copy of the SMN1 gene and 5 copies of the SMN2 gene, age at sampling not available). SMN2 gene copy number was described in Stabley et al. (2015). Cells were propagated in Dulbecco's Modified Eagle Medium (DMEM, Sigma-Aldrich) containing 10% (v/v) Fetal Bovine Serum (FBS, Sigma Aldrich), 2 mM GlutaMAX (Invitrogen) and 1x antibiotic-antimycotic (Invitrogen).

Preparation of extracellular vesicle samples

For quantitative proteomic analysis, we utilized cell lines GM03813 and GM03814, which were derived from a male patient affected with SMN type II and his unaffected, carrier mother. Fibroblast cell lines were seeded into 10x 15 cm dishes at 10^7 cells per dish (approximately 50% confluency). Next day, the cells were switched to DMEM supplemented with 5% vesicle-depleted FBS, 2 mM GlutaMAX, and 1X antibiotic-antimycotic. The vesicle-depleted FBS was prepared by centrifuging overnight (approximately 18 h) at 100,000×g at 4˚C, and the resulting supernatant was carefully

decanted and vacuum-filtered through a 0.22 μM filter and stored at 4˚C until use. Seventy-two hrs after the cells were switched to the vesicle-depleted medium, the medium was collected and EV were isolated by differential centrifugation. The medium (approximately 200 ml per cell line) was centrifuged at 300xg for 10 min at 4˚C to remove cells and large cellular debris. The resulting supernatant was centrifuged at 2000xg for 10 min at 4˚C to remove apoptotic bodies. The supernatant was centrifuged at 100,000xg for 70 min at 4˚C to collect a mixed population of large (microparticles) and small (exosome) EV. The resulting pellets were washed once in PBS by combining and resuspending in a total of 13 ml PBS, centrifuging at 100,000xg for 70 min at 4˚C, and finally resuspending in 100 μl of PBS. The sample was stored at -80^{o}C until analyzed. EV size and concentration was determined using a Zetaview PMX110 Multiple Parameter Particle Tracking Analysis (ParticleMetrix, Meerbusch, Germany) in size mode. EV were diluted in PBS to the working range of the system (10^6–10^9 particle/ml), and videos were captured with the ZetaView software (version 8.02.28, Meerbusch, Germany) using 11 camera positions, a 2-second video length, and a camera frame rate of 30 fps at 21˚C. EV protein concentration was determined using the Bradford assay (BioRad, Hercules, CA).

EV for immunoblot analysis were prepared by precipitation with polyethylene glycol (PEG). Briefly, fibroblast cells were seeded in a 10 cm dish at 5x10^6 cells per plate in complete medium. Next day, the cells were switched to DMEM supplemented with 5% vesicle-depleted FBS, 2 mM GlutaMAX, and 1x antibiotic-antimycotic. Seventy-two hrs later, the medium was collected and centrifuged at 2000xg for 10 min at room temperature to remove cell debris. The supernatant was combined with an equal volume of 16% PEG

(MW 8000) prepared in 137 mM NaCl. The mixture was incubated overnight at 4˚C, centrifuged at 1,500xg for 30 min at 4˚C, and resuspended in 100 μl PBS.

Immunoblot Analysis

For preparation and analysis of cellular protein content, cells were lysed in 2xSDS–PAGE protein loading buffer (62.5 mM Tris, pH 6.8, 10% glycerol, 2% SDS, 0.1% bromophenol blue, 5% 2-mercaptoethanol). For EV analysis, 16 μl of EV isolated by PEG precipitation was combined with 4 μl of 5xSDS–PAGE protein loading buffer. The protein samples were boiled for 10-20 min, and separated by electrophoresis on a 10% SDS–polyacrylamide gel. The separated proteins were transferred to a polyvinylidene difluoride membrane (Millipore) using the Trans-Blot SD Semi-Dry Transfer System (Bio-Rad). The membrane was blocked with 5% bovine serum albumin (BSA) in tris-buffered saline containing 0.2% Tween 20 (Thermo Fisher Scientific) and probed with antibodies diluted in the 5% BSA solution. The following primary antibodies were used: rabbit anti-flotillin-2 (1:1,000, Cell Signaling, MA), mouse anti-SMN (1:10,000 for cell lysates or 1:1,000 for EV, BD Transduction, Mississauga, Ontario) or rabbit anti-SMN (Santa Cruz, Dallas TX), or mouse anti-Tubulin (1:10,000, EMD Millipore, Etobicoke, Ontario). Following exposure to antibody, the membranes were washed three times in TBST and incubated with the appropriate secondary antibodies conjugated to horseradish peroxidase (HRP). Blots were developed using the Immobilon Classico Western HRP Substrate (Millipore).

Tryptic peptide sample preparation

EVs isolated as described above were prepared for SDS-PAGE separation based on particle counts obtained from the Particle Tracking Analysis described above. An equal number of EVs between control and SMA-derived samples were prepared for SDS-PAGE

analysis by resuspension in 5% SDS buffer and boiled for 5 minutes. Solubilized samples were separated on 10% SDS-PAGE gels using the short-stacking technique.[8] Briefly, solubilized EVs were resolved until the dye front migrated 2 cm into the separating gel. Proteins were then visualized with Imperial Coomassie stain (Thermo Fisher).

Four 1 cm x 1 cm gel slices per lane were excised and prepared for MS analysis. Gel pieces were destained using acetonitrile (ACN) **(**Burdick and Jackson) and 50 mM NH_4HCO_3 (Fisher Scientific) for 15 minutes at 37˚C. Following destaining, gel pieces were dehydrated by the addition of neat ACN and a brief incubation at room temperature. Post-dehydration, gel pieces were fully dried by vacuum centrifugation. Gel pieces were rehydrated by addition of 1.5 µg of trypsin (Promega) in 50 mM NH_4HCO_3 on ice for 1.5 hours, and then incubated overnight at 37˚C. Extraction of the tryptic peptides was accomplished by incubating 150 µL of extract solution containing 2.5% Formic acid (Fluka) and 50% ACN with the gel pieces at 37˚C for 15 minutes. Extraction was repeated and all extracts were pooled in siliconized tubes and vacuum concentrated.

Lyophilized tryptic peptides were reconstituted in 100 µL of 0.1% trifluoroacetic acid (TFA) (Fluka) and purified by solid phase extraction using Oasis CLB 1cc columns (Waters). Oasis columns were preconditioned using 100 µL of ACN with 0.1% TFA, washed twice with 100 µL of 0.1% TFA, prior to loading of EV-derived peptides. Columns were washed twice with 100 µL of 0.1% TFA. A stepwise elution method was utilized, first using 100 µL of 30% ACN in 0.1% TFA, followed by 100 µL of 50% ACN in 0.1% TFA, and finally twice with 150 µL of 80% ACN in 0.1% TFA. Desalted peptides were then vacuum concentrated to dryness.

Mass Spectrometry and data analysis

Peptides were reconstituted in 20 µL of 0.1% formic acid (FA) and loaded onto a nanoAcquity ultra performance liquid chromatography (UPLC) system for LC-MS^E (Waters). Three biological replicates from control cells and three biological replicates from SMA-derived cells were analyzed. Peptides were injected onto a 1.8 µm HSS T3 75 µm x 150 mm reverse-phase column (Waters) at a flow rate of 0.3 µL/min using a nanoAcquity UPLC system. Mobile phase A (0.1% formic acid in water) and mobile phase B (ACN with 0.1% formic acid) were used to equilibrate the column. Elution of peptides occurred using an 85 min ACN gradient (3-30% B for 60 min, 30-50% B for 15 min, 85% B for 10 min) and directly sprayed into a SYNAPT G2-S*i* mass spectrometer (Waters) using a cone voltage of 30 V and a capillary voltage of 3 kV. The instrument was operated in data independent acquisition mode with ion mobility activated. Low energy scans were captured at 4 eV, while high energy scans were captured using a 20-45 eV ramp in positive resolution mode scanning from 50 to 2000 *m/z* at a rate of 0.8 s. Calibration was achieved using [Glu1]-fibrinopeptide B (50 fmol/µL) in the lockmass channel at *m/z* 785.8427 for a doubly charged positive ion. Raw data was collected using Mass Lynx version 4.1 and have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository.

Progenesis QI (Nonlinear Dynamics) was used for data processing, including chromatogram alignments, data normalization, and peptide identification using the human UniProtKB/SwissProt database (20,614 proteins), acquired May 23, 2014. Raw data was processed according to the following parameters: low energy noise reduction of 135 counts, high energy reduction noise reduction threshold of 30 counts, and an intensity threshold of 750 counts. Lock mass correction occurred post-acquisition. Protein identification was

achieved using the following parameters: a minimum of 3 fragment ions per peptide, a minimum of 7 fragment ion matches per protein, and a minimum of 1 peptide match per protein. The maximum false discovery rate of less than 1% was estimated by using a decoy database. A maximum of two missed cleavages following trypsin digestion was permitted and the variable modification of methionine oxidation +15.9949 was added. The Hi3 method of quantification originally described by Silva et al., 2006,[9] was used on the identified proteins, where comparisons of the 3 most abundant peptides from the samples and spiked reference were used to determine protein levels. To accomplish this, 37.5 fmol of the internal peptide standard Hi3 (Waters) derived from the chaperone protein ClpB (P63284), was spiked into all biological replicate samples. The ClpB protein sequence was added into the human Uniprot database and used to convert normalized abundance values to on-column absolute fmol measurements as previously described.[10] The relative fold changes of proteins identified in both normal control and SMA-derived EVs were determined. Fold changes were calculated by making a SMA/normal ratio of abundance and log2 transforming these values. A two-tailed t-test was used to calculate significance and were -log10 transformed. Volcano plots were created using GraphPad Prism Version 8.2.1. A P-value of 0.001 was used to indicate significance.

RESULTS AND DISCUSSION

Isolated extracellular vesicles from spinal muscular atrophy cells show high purity and EV markers

Previously work has suggested EVs may represent a promising source for biomarker discovery in spinal muscular atrophy disease. To investigate this hypothesis further, EVs were isolated from patient derived fibroblast cells and control cells for use in proteomic analysis (**Figure 3.1**). EVs were isolated using differential centrifugation methods and characterized using Particle Tracking Analysis (**Figure 3.2**). Particle Tracking Analysis measured the diameter of the isolated particles to be approximately 140 nm, consistent with reported size of microparticles and exosome vesicles. Collectively, these results indicate that the preparation methods were suitable for downstream proteome investigation.

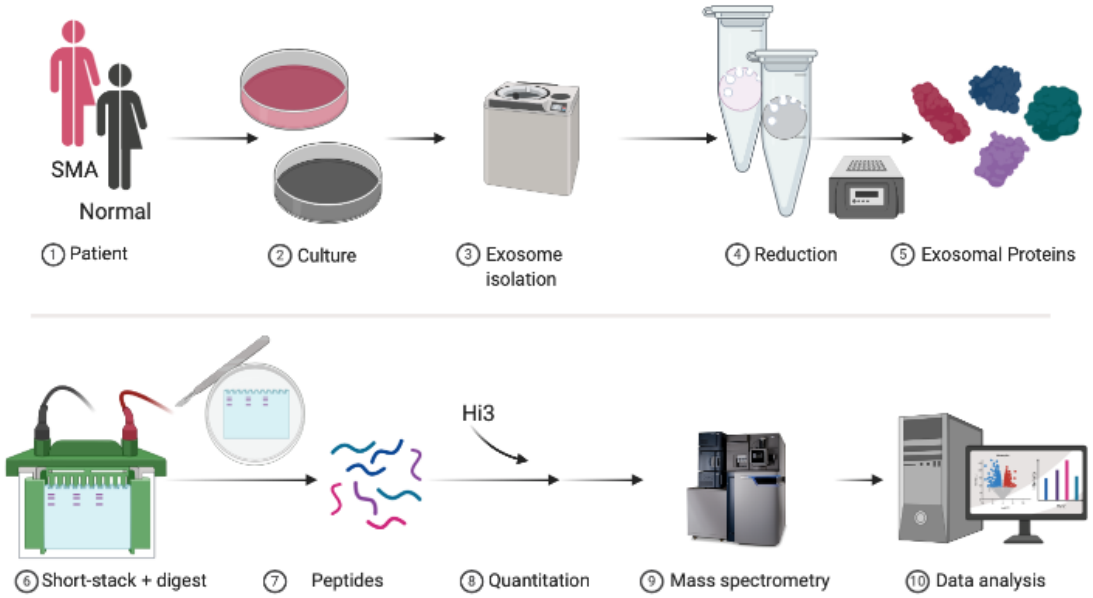


Figure 3.1. Schematic for Spinal Muscular Atrophy extracellular vesicle proteomic workflow.

Patient derived fibroblasts and unaffected control cells were cultured as described in "Experimental Section". Extracellular vesicles were isolated using differential centrifugation and characterized by Particle tracking analysis. Samples were prepared for proteomic analysis by the short stack SDS-PAGE method and subjected to in-gel trypsin digestion. Label-free quantitative proteomics was accomplished using the data independent acquisition mode and ion mobility separation technique. Hi3 internal standards were used to convert normalized data to fmol on-column. Data analysis was performed using Progenesis software package. Made using Biorender (https://biorender.com).

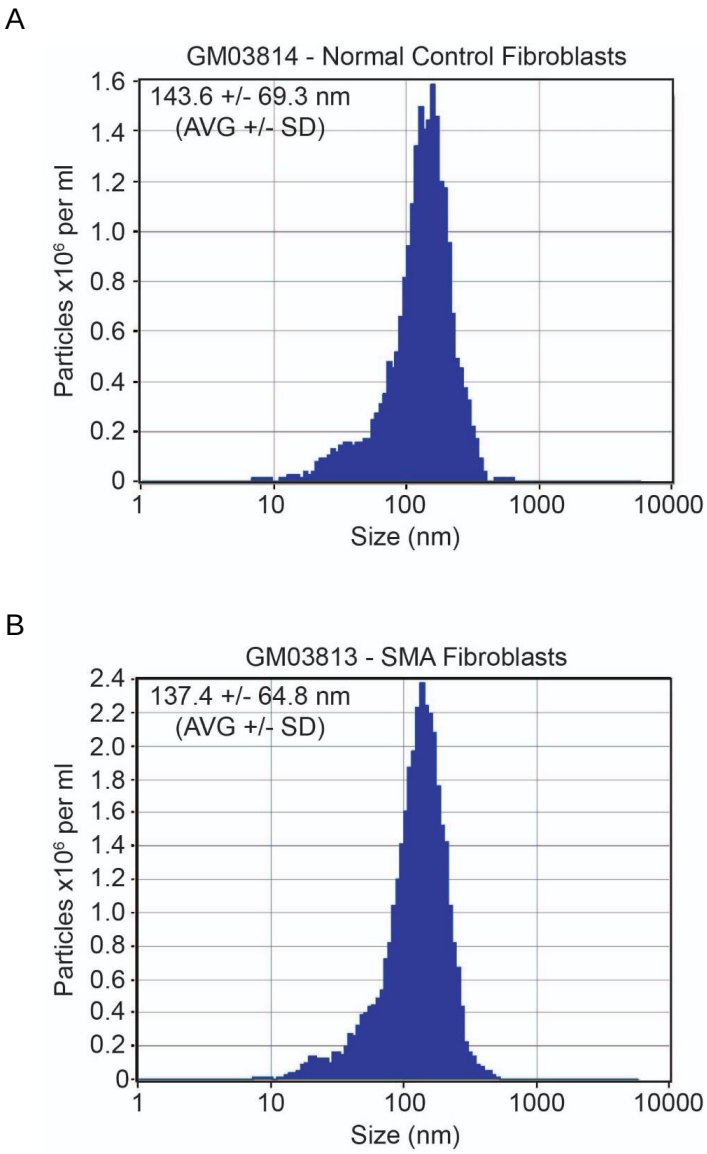


Figure 3.2. Particle tracking analysis of extracellular vesicles isolated from SMA cells. Extracellular vesicles from (**A**) control cells and (**B**) SMA cells were subjected to Particle tracking analysis to determine size and concentration (particles per ml). Video measurements were captured using 11 camera positions, a 2-second video length, and a camera frame rate of 30 fps at 21 °C.

Proteome analysis of extracellular vesicles from SMA fibroblasts

Following EV characterization, proteome profiling was performed. Samples were prepared for mass spectrometry using a so-called short stack SDS-PAGE approach.[8] This approach involves running the samples approximately 2 cm into the separating gel before halting electrophoresis and visualizing migration via Coomassie staining. The 2 cm region is then excised and subjected to in-gel trypsin digestion and sample clean-up prior to tandem mass spectrometry. EV protein identification was undertaken using data-independent acquisition mass spectrometry and incorporated ion mobility separation to improve resolution and increase protein identification yield.[11] Proteome measurements were performed on six biological replicates (3 control samples and 3 SMA samples). Furthermore, to improve statistical rigor, three technical replicates were acquired for each biological replicate (18 total runs). Collectively, 684 proteins were confidently identified across all samples with 659 proteins being identified in every technical replicate tested (**Appendix B**). Gene ontology was performed to classify the total identified proteins in terms of cellular component, molecular function, and biological process (**Figure 3.3**). The analysis revealed overrepresentation of numerous ontology categories associated with EV assembly, vesicle-mediated transport, membrane docking, and exocytosis. This finding is consistent with the Particle Tracking analysis data indicating that the sample preparation method effectively isolated EVs. Additionally, we compared out acquired dataset to the top 100 proteins curated on ExoCarta (http://exocarta.org/index.html) to further validate the EV content of our samples. By comparing our complete dataset to these 100 proteins,

there was an overlap of 86 proteins (**Figure 3.4**), including well-characterized markers CD9, CD81, flotillin, Hsc71, enolase, and VCP.

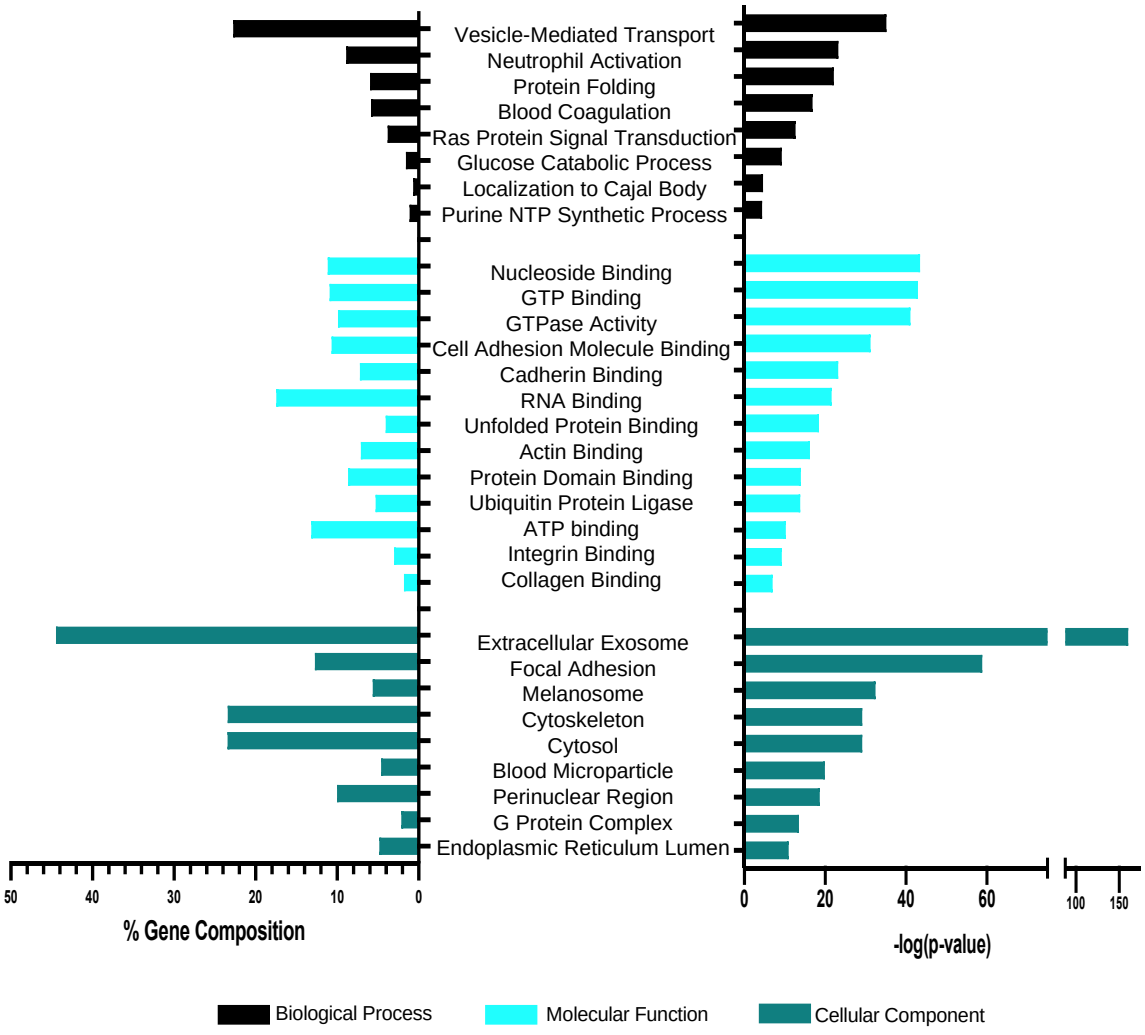


Figure 3.3. Gene ontology analysis of total extracellular vesicle proteome dataset.

Gene ontology analysis using the Panther classification system with % gene composition (left bars) and enrichment of terms (right bars) shown. Gene ontology was mapped for proteins in relation to biological process (top), molecular function (middle), and cellular component (bottom). Gene ontology terms of notable significance are shown on the y-axis.

In addition to verification of the exosomal characteristics of our sample, we were interested in determining if any of the protein cargo were known SMA biomarkers. A panel of SMA biomarkers (BforSMA) has been developed from a collaboration between Myriad Rules-Based Medicine and the SMA Foundation (https://smafoundation.org). Furthermore, independent studies evaluating the BforSMA biomarker panel have been conducted in SMA patients using blood as the sample source. These studies have validated the biomarker panel and have correlated the expression of certain biomarkers with clinical outcomes such as changes in motor function.[12,13] As EVs are an attractive source for biomarker monitoring, we examined if any of the BforSMA biomarkers were detected in our EV samples. If detected, we were also interested in evaluating protein levels between control and SMA cells to determine if their expression pattern corresponded to reported alterations observed clinically in SMA patients.[12] Interestingly, we detected 22 out of 84 discovery phase SMA biomarkers associated with Modified Hammersmith Functional Motor Scale (MHFMS) SMA motor scores (**Figure 3.4**). Importantly, biomarkers detected in our EV sample which satisfied the criteria for quantitative analysis displayed protein alteration patterns that were consistent with previous reports from patient blood samples.[12,13] Most notably were Endoglin and dipeptidyl peptidase 4 (DPP4) which are components of a commercial SMA biomarker panel.[12,13] These results indicate that EVs represent an optional sample source for measuring standardized biomarkers for SMA.

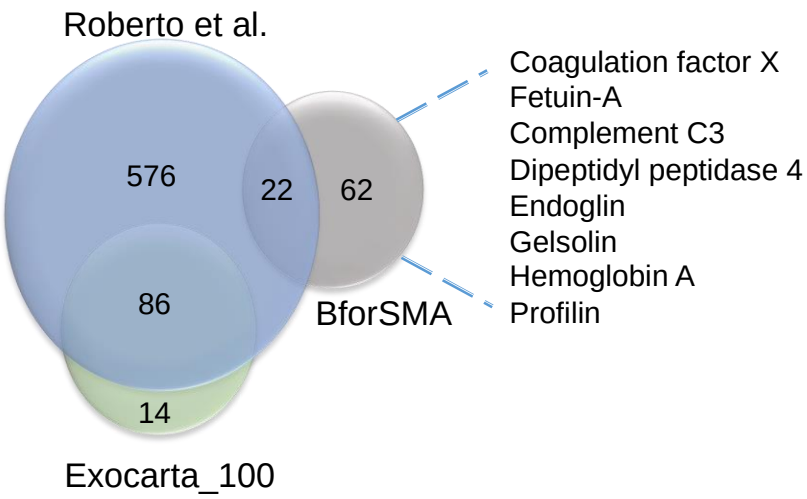


Figure 3.4. Extracellular vesicle proteome contains known SMA biomarkers.

Venn diagram displaying the overlap of proteins from the current study (Roberto et al.), the top 100 annotated exosome proteins commonly identified (Exocarta_100) (http://exocarta.org), and discovery phase panel of SMA biomarkers (BforSMA).[117] Listed are proteins from the BforSMA biomarker panel that show statistically significant protein alterations in the current study.

Extracellular vesicles from SMA cells display significant protein alterations, revealing putative biomarkers

In addition to previously characterized SMA biomarkers, we were interested in identifying protein alterations within the EV samples as a first step towards discovering novel putative SMA biomarkers. Only proteins that were confidently identified in all technical and biological replicates and had $\geq$3 unique peptides associated with their identification were qualified for quantitative analysis. Statistical analysis of the label-free quantitative proteomic data utilized a Student's *t*-test to classify significant differences of on-column protein levels between control and SMA cells. This initial screen identified 133 proteins whose protein levels were concluded to be statistically significant (P-values <0.05) (**Appendix B**). This group of proteins were then assessed additional statistical criteria including ANOVA analysis (P values < 0.001) and fold change differences (>1.5). This assessment yielded 116 proteins that met all three criteria.

A volcano plot was constructed by inputting the $\log_{10}$ P-value as a function of $\log_2$ fold-change to visually display the protein alteration data (**Figure 3.5**). Most proteins cluster in a region at the base of the volcano. This is expected as it indicates that the majority of proteins identified in the EV samples were not significantly different in terms of proteins levels. Of the altered proteins, 94 out of the 116 proteins were observed to be significantly higher in control cells while 21 proteins were higher in SMA cells (**Appendix B**). It is unclear why there is substantially more proteins downregulated in SMA cells compared to upregulated. The majority of total proteins identified in the study were found to be present at similar levels. Thus, it is unlikely that this discrepancy is due to a defect in EV dynamics or sample preparation. Considering the role of SMN1 in translation

regulation, it is possible that loss of SMN1 impacts the steady state levels of a subset of the proteome. This is consistent with current hypotheses that have suggested the physiological significance of low SMN1 protein levels results in defective mRNA metabolism.[14]

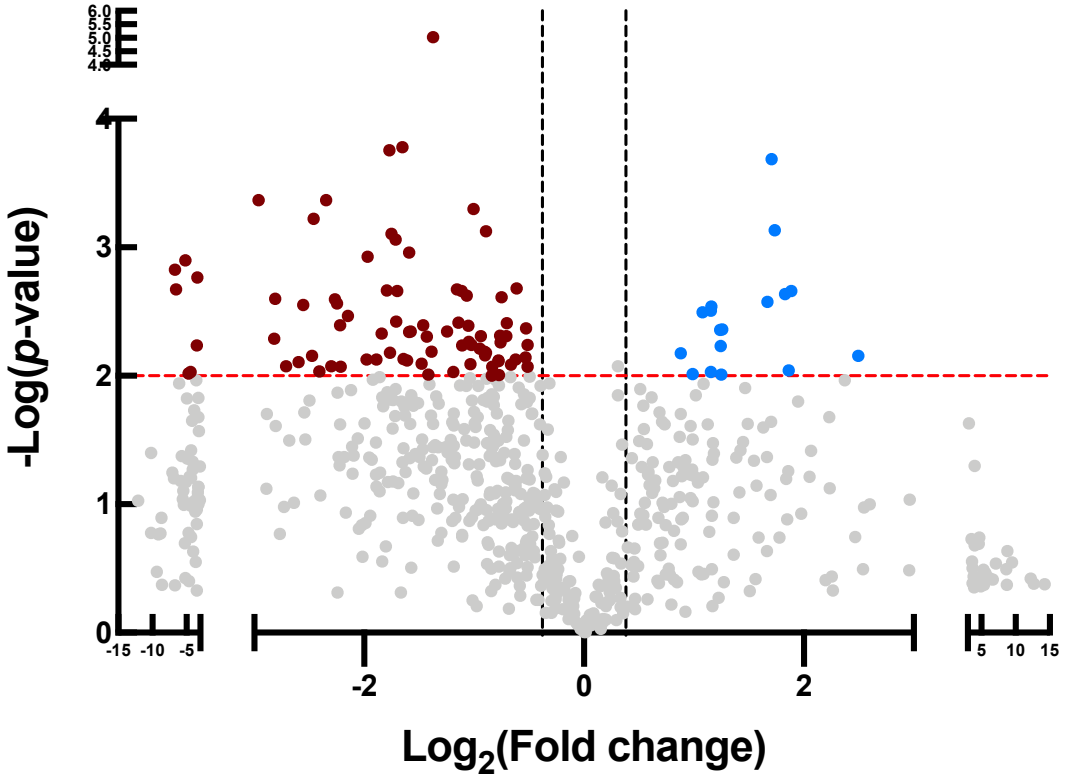


Figure 3.5. Volcano plot depicting differentially abundant extracellular vesicle proteins. The volcano plot displays -log10 corrected *t*-test values (y-axis) vs. log2 fold change (x-axis). Data points that are above the horizontal dashed red line (P values below 0.01) and are to the left (red) and right (blue) of the vertical dashed black lines (abundance ratios above 1.5) represent proteins with statistically significant protein alteration levels.

Gene ontology enrichment analysis of the cohort of proteins whose protein levels were altered did not result in any statistically significant findings. However, inspection of the protein alterations identified several intriguing differences that corroborates previous reports and reveals novel promising avenues for future research (**Figure 3.6**). For example, Plastin-3 was identified to be 4.75 fold lower in the EVs of SMA cells compared to control (**Figure 3.6**). Plastin-3 is an actin-binding protein and has been extensively studied as a potential gene therapy candidate. Attention has been granted to Plastin-3 due to the observation that asymptomatic women carriers of SMN1 deletions have been found to exhibit overexpression of Plastin-3.[15] Moreover, using various SMA animal models, overexpression of Plastin-3 reversed SMA phenotypes associated with SMN1 deficiency.[16] Therefore, our finding of Plastin-3 levels being significantly lower in EVs derived from SMA cells is consistent with the working model that low Plastin-3 corresponds to SMA phenotype severity.[1,15] Furthermore, the finding that Plastin-3 can be monitored in EVs is indicative of its potential as a prognostic biomarker for disease diagnostics using EVs.

Another interesting finding was the statistically significant alteration of the inactive protein tyrosine kinase 7 (PTK7). PTK7 functions in planar cell polarity non-canonical Wnt signal transduction, where it acts as a co-receptor for the Frizzled family of Wnt receptors.[17] There is increasing evidence that packaging components of the Wnt signaling pathway in EVs for spreading to peripheral tissues is a common mechanism for the morphogenic effects of Wnt signal transduction.[18] Moreover, upregulation of Wnt signaling components has been described in transgenic mouse models of ALS,[19] while a recent study has postulated a role for dysregulation of the Wnt signaling pathway in SMA subtypes that involve lamin perturbations.[20] In our study, PTK7 levels were found to be

2.3-fold lower in our SMA EV sample. A decrease in a Frizzled co-receptor (PTK7) that mediates non-canonical Wnt signaling is thought to induce higher levels of canonical Wnt signaling.[21,22] Thus our PTK7 result supports the working models of dysfunctional Wnt signaling in ALS and SMA. Interestingly, we also observed alterations in other Wnt pathway components in our EV dataset (**Appendix D**). The Wnt-3 ligand was over 2-fold higher in the SMA sample while the Frizzled related protein 1 (FRP1), which inhibits Wnt ligand-receptor association, was over 3-fold lower in the SMA sample. These proteins were not included in our protein alteration list since they were identified with less than 3 unique peptides. However, their peptide confidence scores were very high, with statistically low variation. Collectively, our findings highlight the potential of measuring perturbations of Wnt signaling components by EV proteomics and adds to the body of experimental evidence that dysfunctional Wnt signaling may play a significant role in the etiology of SMA disease.

One of the more striking gene ontology terms that was revealed in our dataset was enrichment of proteins associated with regulating localization to Cajal bodies (**Figure 3.3**). Cajal bodies are found in the nucleus and are sites of small nuclear ribonucleoprotein (snRNP) biogenesis and telomere maintenance.[23] SMN protein functionally localizes to Cajal bodies,[24] while a failure of SMN to access these nuclear sites leads to reorganization of Cajal bodies in SMA type I.[25] Our EV proteomic dataset contained six T-complex protein 1 (TCP-1) subunits (alpha, beta, gamma, delta, theta, and zeta), which constitutive components of the chaperonin complex TCP-1 Ring Complex (TRiC).[26] TRiC is critical for trafficking regulatory proteins to Cajal bodies.[27] Interestingly, the protein levels of two TCP-1 subunits (alpha (3-fold) and beta (2.4-fold)) were significantly lower in the SMA

EV samples (**Figure 3.6**). This result is consistent with a study that discovered TCP-1 alpha and beta to be downregulated in skeletal muscle from 'severe' SMA mice.[28] Also, the antisense oligonucleotide therapeutic agent, Nusinersen, has been shown to co-localize with TCP-1 beta, and is postulated to mediate Nusinersen nuclear accumulation.[29,30] Taken together, TCP-1 alpha and beta subunits may represent critical biomarkers for both diagnostic and therapeutic monitoring in SMA disease.

While most of the protein alterations detected in our dataset were downregulated in the SMA sample, there were several promising findings whose protein levels were elevated, including known SMA biomarkers Fetuin-A (5.2-fold), and Coagulation factor X (2.4-fold) (**Figure 3.6**). Another protein that was significantly elevated in the SMA sample was the insulin-like growth factor binding protein 3 (IGFBP3). IGFBP3 was found to have a 3.5-fold increase in the SMA sample compared to the control sample, suggesting IGFBP3 may represent a novel SMA biomarker. To support the quantitative proteomic analysis, we examined the abundance levels of IGFBP3 in the control and SMA cell lines by immunoblot analysis (**Figure 3.7**). In addition, levels of IGFBP3 were probed in unique SMA cell lines to validate the trend in distinct patient derived cells lines. As shown in **Figure 3.7**, IGFBP3 was observed to be dramatically elevated in SMA cell lysate and EV preparations compared to control in all cell line pairs tested. The altered levels of IGFBP3 is intriguing as a growing body of experimental evidence points to a role of the IGF signaling pathway in SMA and other neuromuscular disorders. The human genome encodes for 6 IGFBP family members all with similar architecture, having a cysteine-rich N-terminal domain, a variable linker region, and a cysteine-rich C-terminal domain.[31] IGBP3 is the most prevalent family member present in human sera and has been implicated

in other neurological diseases such as ALS by regulating the half-life of its binding partner IGF-1. Although IGFBP3 lengthens the half-life of IGF-1 and aids in shuttling to target tissues, high levels of IGFBP3 can also attenuate IGF-1 signaling through sequestration from its receptor.[32,33] This sequestration mechanism is important since IGF-1 signaling is critical for the survival of motor neurons and decreased levels have been observed in SMA-model mice.[34,35] Interestingly, IGF-1 receptor levels also display subtype specific expression patterns, with SMA type I showing higher levels of IGF-1 receptor levels than SMA type III.[36] Therefore, together with the markedly higher levels in the SMA EV samples, IGFBP3 represents a promising novel SMA biomarker candidate.

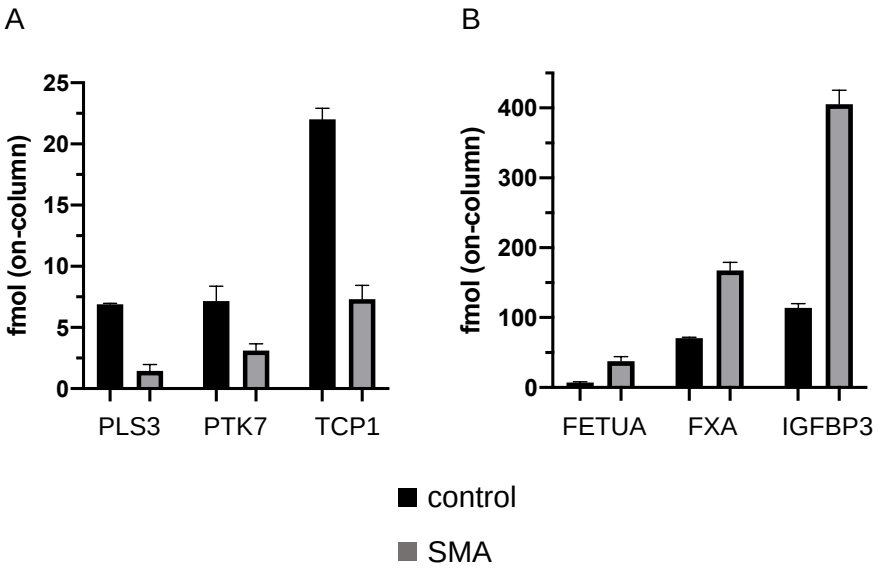


Figure 3.6. Significant protein alterations in extracellular vesicles from SMA cells. Proteins of interest found to be statistically significant in terms of abundance ratio differences are shown. Proteins are organized by lower abundance in (A) SMA cells and (B) higher abundance in SMA cells. Protein abundances are shown as mean fmol (on-column) + standard deviation. All protein comparisons shown were statistically significant, $P < 0.001$. Proteins shown are the following: Plastin-3 (PLS3), Inactive protein tyrosine kinase 7 (PTK7), T-complex protein 1 (TCP1), Fetuin-A (FETUA), Coagulation factor X (FXA), and insulin-like growth factor binding protein 3 (IGFBP3).

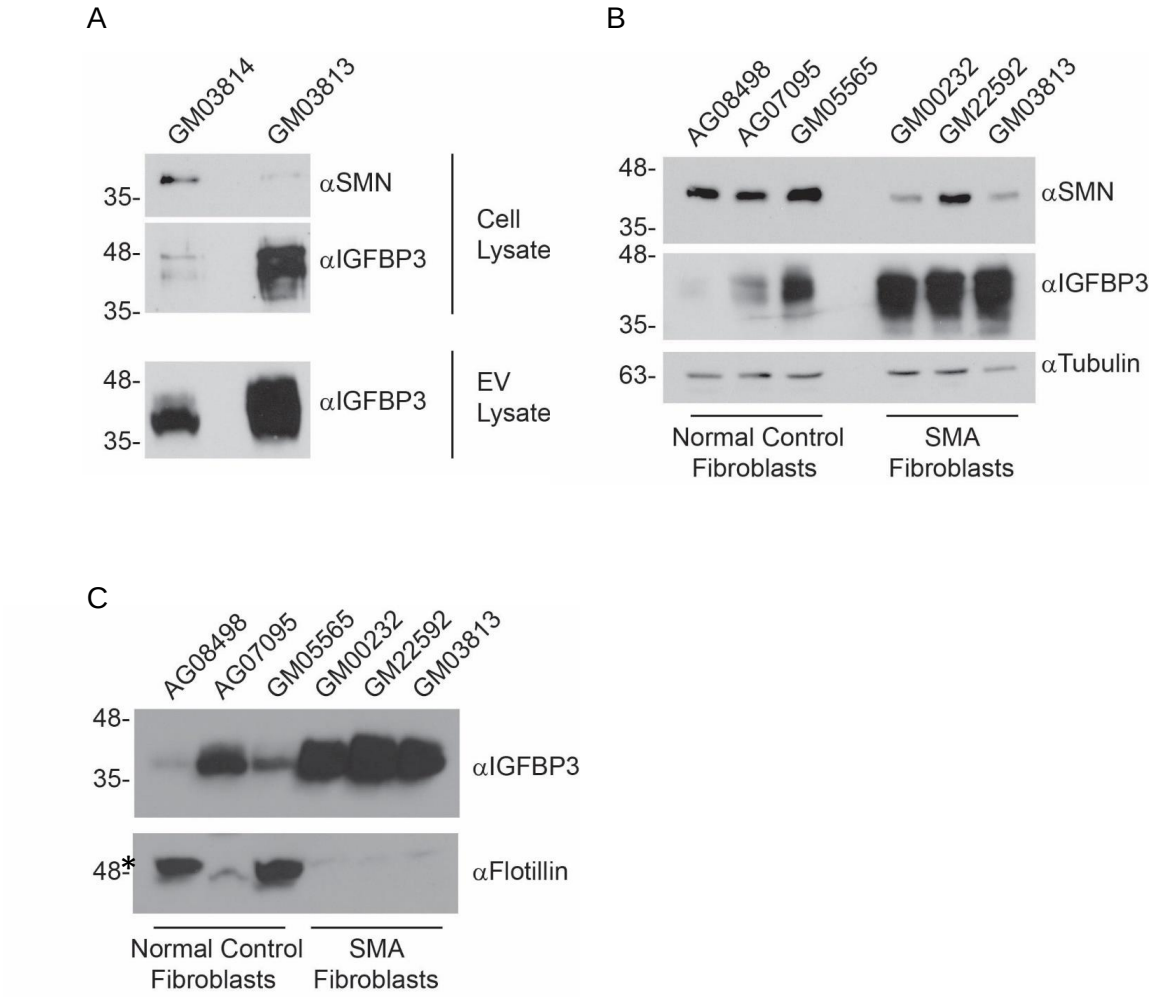


Figure 3.7. Validation of IGFBP3 as a novel biomarker in the cell lysate and extracellular vesicles of SMA cells.

Representative images of immunoblotting experiments on patient derived SMA cells and unaffected controls showing the significant elevated protein levels of IGFBP3 in SMA samples. (**A**) aliquot of extracellular vesicle samples (EV lysate) that were the subject of the current proteomic study (bottom panel) and corresponding cell lysate (upper panels). GMO3814 is the unaffected control cell line, while GMO3813 is the SMA cell line. Representative immunoblots of cell lysates (**B**) and extracellular vesicles (**C**) from indicated SMA patient derived fibroblasts. Levels of SMN are shown as a SMA control,

tubulin as a cell lysate loading control (B, lower panel), and flotillin as an extracellular vesicle loading control (C, lower panel). The asterisk highlights non-specific IgG bands observed in the extracellular vesicle

CONCLUSION

In conclusion, proteomic analysis of EVs isolated from SMA fibroblasts revealed numerous established and potential novel SMA biomarkers. It is important to note that recent work has highlighted the multi-organ nature of SMA disease.[37] This feature of SMA underscores the potential of utilizing EVs as a potential tool for clinical prognostic testing and for further understanding the molecular mechanisms underlying this fatal disease. Many of the therapeutic efforts for treating SMA patients involve elevating the protein levels of SMN. Therefore, it will be important in the future to utilize EV targeted proteomics to investigate if increasing SMN protein levels in turn restores steady state levels of altered proteins identified in this study. This would strongly support their functional relevance as SMA biomarkers and EVs as a promising source for monitoring SMA and other neurodegenerative diseases.

REFERENCES

1. Kariyawasam, D. S. T., D'silva, A., Lin, C., Ryan, M. M. & Farrar, M. A. Biomarkers and the development of a personalized medicine approach in spinal muscular atrophy. *Front. Neurol.* **10**, 1–12 (2019).
2. Ruhno, C. *et al.* Complete sequencing of the SMN2 gene in SMA patients detects SMN gene deletion junctions and variants in SMN2 that modify the SMA phenotype. *Hum. Genet.* **138**, 241–256 (2019).
3. Gavrilov, D. K., Shi, X., Das, K., Gilliam, T. C. & Wang, C. H. Differential SMN2 expression associated with SMA severity [3]. *Nat. Genet.* **20**, 230–231 (1998).
4. Bach, J. R., Niranjan, V. & Weaver, B. Spinal muscular atrophy type 1: A noninvasive respiratory management approach. *Chest* **117**, 1100–1105 (2000).
5. Lefebvre, S. *et al.* Identification and characterization of a spinal muscular atrophy-determining gene. *Cell* **80**, 155–165 (1995).
6. Chaytow, H., Huang, Y. T., Gillingwater, T. H. & Faller, K. M. E. The role of survival motor neuron protein (SMN) in protein homeostasis. *Cell. Mol. Life Sci.* **75**, 3877–3894 (2018).
7. Johnstone, R. M., Adam, M., Hammond, J. R., Orr, L. & Turbide, C. Vesicle formation during reticulocyte maturation. Association of plasma membrane activities with released vesicles (exosomes). *J. Biol. Chem.* **262**, 9412–9420 (1987).
8. Wang, Z., Hill, S., Luther, J. M., Hachey, D. L. & Schey, K. L. Proteomic analysis of urine exosomes by multidimensional protein identification technology (MudPIT). *Proteomics* **12**, 329–338 (2012).
9. Silva, J. C., Gorenstein, M. V., Li, G. Z., Vissers, J. P. C. & Geromanos, S. J. Absolute quantification of proteins by LCMSE: A virtue of parallel MS acquisition. *Mol. Cell. Proteomics* **5**, 144–156 (2006).
10. Finamore, F. *et al.* Impact of high glucose concentration on aspirin-induced acetylation of human serum albumin: An in vitro study. *EuPA Open Proteomics* **3**, 100–113 (2014).
11. Helm, D. *et al.* Ion mobility tandem mass spectrometry enhances performance of bottom-up proteomics. *Mol. Cell. Proteomics* **13**, 3709–3715 (2014).
12. Kobayashi, D. T. *et al.* SMA-MAP: A Plasma Protein Panel for Spinal Muscular Atrophy. *PLoS One* **8**, (2013).
13. Finkel, R. S. *et al.* Candidate proteins, metabolites and transcripts in the biomarkers for spinal muscular atrophy (BforSMA) clinical study. *PLoS One* **7**, (2012).
14. Li, D. K., Tisdale, S., Lotti, F. & Pellizzoni, L. SMN control of RNP assembly: From post-transcriptional gene regulation to motor neuron disease. *Semin. Cell Dev. Biol.* **32**, 22–29 (2014).
15. Oprea, Gabriela E., Sandra Kröber, Michelle L. McWhorter, Wilfried Rossoll, Stefan Müller, Michael Krawczak, Gary J. Bassell, Christine E. Beattie, and B. W.

Plastin 3 Is a Protective Modifier of Autosomal Recessive Spinal Muscular Atrophy. *Science.* **320**, 524–527 (2008).

16. Kaifer, K. A. *et al.* Plastin-3 extends survival and reduces severity in mouse models of spinal muscular atrophy. *JCI Insight.* **2**, 1–10 (2017).

17. Peradziryi, H. *et al.* PTK7/Otk interacts with Wnts and inhibits canonical Wnt signalling. *EMBO J.* **30**, 3729–3740 (2011).

18. Gross, J. C. & Zelarayán, L. C. The Mingle-Mangle of Wnt Signaling and Extracellular Vesicles: Functional Implications for Heart Research. *Front. Cardiovasc. Med.* **5**, 1–8 (2018).

19. Chen, Y. *et al.* Wnt signaling pathway is involved in the pathogenesis of amyotrophic lateral sclerosis in adult transgenic mice. *Neurol. Res.* **34**, 390–399 (2012).

20. Fuller, D. Š. and H. R. Molecular Crosstalk Between Non-SMN-Related and SMN-Related Spinal Muscular Atrophy Darija. *Hum. Mol. Genet.* **15**, 1–3 (2020).

21. Hayes, M. *et al.* Ptk7 promotes non-canonical Wnt/PCP-mediated morphogenesis and inhibits wnt/β- catenin-dependent cell fate decisions during vertebrate development(Development, 140, (1807-1818)). *Dev.* **140**, 2245 (2013).

22. Lu, X. *et al.* PTK7/CCK-4 is a novel regulator of planar cell polarity in vertebrates. *Nature* **430**, 93–98 (2004).

23. Sawyer, I.A., Sturgill, D., Sung, M.H., Hager, G.L. and Dundr, M. Cajal body function in genome organization and transcriptome diversity Iain. *Bioessays* **38**, 1197–1208 (2016).

24. Hebert, M. D., Szymczyk, P. W., Shpargel, K. B. & Gregory Matera, A. Coilin forms the bridge between Cajal bodies and SMN, the spinal muscular atrophy protein. *Genes Dev.* **15**, 2720–2729 (2001).

25. Tapia, O. *et al.* Reorganization of Cajal bodies and nucleolar targeting of coilin in motor neurons of type i spinal muscular atrophy. *Histochem. Cell Biol.* **137**, 657–667 (2012).

26. Knee, K. M., Sergeeva, O. A. & King, J. A. Human TRiC complex purified from HeLa cells contains all eight CCT subunits and is active in vitro. *Cell Stress Chaperones* **18**, 137–144 (2013).

27. Freund, A., Zhong, F. L., Venteicher, A. S., Meng, Z., Veenstra, T. D., Frydman, J., & Artandi, S. E. Proteostatic control of telomerase function through TRiC- mediated folding of TCAB1. *Cell* **159**, 1389–1403 (2014).

28. Mutsaers, C. A., Lamont, D. J., Hunter, G., Wishart, T. M. & Gillingwater, T. H. Label-free proteomics identifies Calreticulin and GRP75/Mortalin as peripherally accessible protein biomarkers for spinal muscular atrophy. *Genome Med.* **5**, (2013).

29. Li, Q. Nusinersen as a therapeutic agent for spinal muscular atrophy. *Yonsei Med. J.* **61**, 273–283 (2020).

30. Liang, X. H., Shen, W., Sun, H., Prakash, T. P. & Crooke, S. T. TCP1 complex

proteins interact with phosphorothioate oligonucleotides and can co-localize in oligonucleotide-induced nuclear bodies in mammalian cells. *Nucleic Acids Res.* **42**, 7819–7832 (2014).

31. Jafari, E. *et al.* In silico interaction of insulin-like growth factor binding protein 3 with insulin-like growth factor 1. *Res. Pharm. Sci.* **13**, 332–342 (2018).

32. Baxter, R. C., Bayne, M. L. & Cascieri, M. A. Structural determinants for binary and ternary complex formation between insulin-like growth factor-I (IGF-I) and IGF binding protein-3. *J. Biol. Chem.* **267**, 60–65 (1992).

33. Allard, J. B. & Duan, C. IGF-binding proteins: Why do they exist and why are there so many? *Front. Endocrinol. (Lausanne).* **9**, 1–12 (2018).

34. Yimin Hua, Kentaro Sahashi, Frank Rigo, Gene Hung, Guy Horev, Frank Bennett, A. R. K. Peripheral SMN restoration is essential for long-term rescue of a severe SMA mouse model. *Nature* **478**, (2012).

35. Murdocca, M. *et al.* IPLEX administration improves motor neuron survival and ameliorates motor functions in a severe mouse model of spinal muscular atrophy. *Mol. Med.* **18**, 1076–1085 (2012).

36. Millino, C. *et al.* Different atrophy-hypertrophy transcription pathways in muscles affected by severe and mild spinal muscular atrophy. *BMC Med.* **7**, 1–9 (2009).

37. Wan, B. *et al.* A severe mouse model of spinal muscular atrophy develops early systemic inflammation. *Hum. Mol. Genet.* **27**, 4061–4076 (2018).

CHAPTER 4

REGULATION OF MTMR2 ENDOSOMAL DYNAMICS THROUGH REVERSIBLE PHOSPHORYLATION

INTRODUCTION

MTMR2 and Charcot-Marie-Tooth disease

Charcot-Marie-Tooth disease (CMT) is one of the most common PNS neuropathies, affecting approximately one in 2500 people.[1] The disease was first described by Charcot and Marie in 1886, and has since been found to manifest itself in various forms, depending on genetic mutations to different genes. Loss of the myotubularin-related protein 2 (MTMR2) causes CMT4B1 in Schwann cells.[2] The seminal study reporting MTMR2—or any dual-specificity phosphatase (DUSP)—being responsible for a peripheral neuropathy found regions of MTMR2 that are commonly mutated. Some mutations found include a nonsense mutation, 1276C>T (Gln426>stop), in an Italian individual, and an in-frame deletion 1480–1593 deletion (Phe494–Glu531 deletion) or an 826G>T (Glu276>stop) from an afflicted Saudi Arabian.[3] To further study MTMR2's role in CMT4B1, MTMR2 knockout mice were developed.[4] These mice exhibit myelin outfoldings and eventual axonal degeneration,[4] consistent with patients suffering from MTMR2-derived CMT4B, where multiple myelin sheaths form onion bulb-like structures and occasional vocal cord paresis and chest deformities.[5]

Both MTMR13 and MTMR5, catalytically inactive members of the MTM family, interact with MTMR2 via coiled coil motifs, regulating the phosphatase activity of MTMR2 (**Figure 4.1**).[6,7,8,9,10] Mutations of these inactive MTMs cause CMT subtypes: CMT4B2 and CMT4B3, for MTMR5 and MTMR13,[11,12,13,14] respectively. Understanding the biological context behind these phenotypes is important, with studies showing MTMR2 can regulate the levels of the pseudophosphatases and vice vera.[10,15] During sciatic nerve maturation, MTMR2 levels remain constant, while the two pseudophosphatases levels

fluctuate, with MTMR13 protein levels being highest during the initial myelination period, and MTMR5 levels highest during late stage radial sorting.[15] As these MTMs and other endolysosomal genes are involved in CMT, it is critical to elucidate the exact mechanisms through which these genes and gene products regulate myelination and CMT.

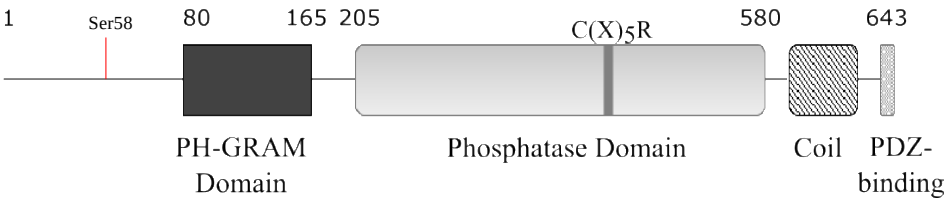


Figure 4.1. MTMR2 protein schematic.

A schematic highlighting the different domains and motifs within MTMR2. Ser^{58} represents serine residue 58, the N-terminal phosphorylation site of the protein. The catalytic region encompasses residues 75–579 and contains two important domains of MTMR2. The first being the localization domain, pleckstrin homology-glucosyl transferase and myotubularin (PH-GRAM) domain,[16] which is composed of amino acid residues 80–185, and allows MTMR2 to localize to PI found on endosomes.[17] The second domain found in the catalytic region is the protein tyrosine phosphatase domain, comprising residues 205–580. Within the PTP domain resides the catalytic cysteine, Cys^{417}. The third region of MTMR2 houses residues 582–643, which includes the C-terminal coiled-coil motif, a highly solvent accessible region important for homo- and heterodimerization with other MTMR proteins, primarily MTMR5 and 13.[9,10,16] Residues 640–643 comprise the PDZ binding motif which is thought to mediate interactions with other PDZ motif-containing proteins such as PSD-95, a protein involved in synapse regulation.[18,19] PH-GRAM = Pleckstrin homology-glucosyltransferase and myotubularins; $C(X)_5R$ the PTP active site residues; Coil = coiled coil; PDZ-binding PSD-95 = post-synaptic density.

MTMR2 and endosomes

MTMR2's substrate, PI3P is critical to the activity of vesicular structures known as endosomes. Clathrin-mediated/dependent endocytosis is a prevalent and well-characterized form of endocytosis responsible for introducing materials into the cell.[20] This complex process is achieved through the concerted effort of many proteins, which ultimately leads to cargo recycling or degradation by transport to the trans-Golgi or the lysosome, respectively. When a vesicle buds from the plasma membrane and enters the cytoplasm, it can fuse with endosomes, where early endosomes (EEs) are thought to be the main sorting station for endocytosed vesicles.[21] The phosphatidylinositols are prevalent on endocytic structures, where various phospho-forms of these lipids can be detected through time. EEs are decorated with PI3P, generated from PI3-kinases through the help of the Rab5 GTPase, which recruits specific proteins that facilitate the fusion process, like the early endosome antigen 1 (EEA1).[22,23] Maturation of EEs to late endosomes occurs through the switch of the Rab5 to Rab7 GTPase found on the endosome, altering the subset of effector proteins and structure of the endosome itself.[24] The maturation of these endosomes determines the fate of the internalized cargo either being recycled back to the plasma membrane or targeted for degradation.[25,26,27] Intricate regulation of these vesicles and their cargo have importance in almost every facet of regulation. For example, PI signaling during the myelination process aiding in regulation of Schwann cell function.[28]

The interplay of PI3K in myelin damage and repair

The PI3K and mitogen-activated protein kinase (MAPK) cascades are critical for rapid cellular response, leading to the activation of multiple genes influencing events, including myelination.[29] The PI3K and MAPK pathways operate in Schwann cells and the

myelination program, with PI3K accelerating Schwann cell maturation and myelination.[30] In contrast, the MAPK pathway hinders Schwann cell maturation and myelination.[30] Cells employ demyelination to adjust to neuronal damage, known as 'Wallerian degeneration'.[31] Through this process, neurons undergoing demyelination typically lose myelin while preserving axons at the injury site while axons downstream of the injured nerve undergo Wallerian degeneration. Schwann cells interacting with these axons dedifferentiate to a progenitor-like state and clear axonal and myelin debris with the help of macrophages. An inflammatory response occurs where inflammatory cells enter the nerve at sites of damage and spread throughout the length of the distal stump.[32] The Raf/MEK/ERK signaling nexus has been shown to activate dedifferentiation of myelinated Schwann cells.[33] Neuregulin 1 (NRG1), a growth factor homologous to EGF, is activated following nerve injury signaling through the Raf kinase pathway. It is believed that ERK1/2 are the effectors influencing myelin integrity, where increasing the strength and/or altering the timing of ERK1/2 activity is critical for these events in both the central and peripheral nervous systems.[29]

ERK-mediated phosphorylation of MTMR2 and its influence on endosomal trafficking

EGF and the MAPK pathway influences MTMR2 localization to and away from the endosome through extracellular-regulated kinase 1/2 (ERK1/2). When activated by EGF stimulation, ERK causes an increase in phosphorylation of MTMR2 at serine 58 ($MTMR2^{pSer58}$).[34,35] Phosphorylation of $MTMR2^{Ser58}$ leads to MTMR2's cytoplasmic localization, serving as a critical form of MTMR2 regulation. A negative feedback loop exists where dephosphorylated MTMR2 is able to localize to endosomes and deplete PI3P, allowing for endosomal signaling to occur, eventually leading to an increase in ERK1/2

activity.[35] As the MAPK pathway is involved in the myelination program, understanding this phosphorylation is critical.[36] A second site on MTMR2, serine 631, was identified as a phosphorylation site regulating MTMR2's localization to early endosome subtypes using phosphodeficient and phosphomimetic mutants.[35] To date, the kinase responsible for MTMR2^{S631} phosphorylation or the phosphatase(s) responsible for MTMR2 Ser^{58} or Ser^{631} dephosphorylation have not been determined. The identity of these enzymes and understanding MTMR2's reversible phosphorylation may provide a critical link between CMT disease onset and MTMR2 activity or localization.

Protein-protein interactions of MTMR2

MTMR2 is involved in the regulation of endosomal dynamics through its participation in a myriad of protein-protein interactions. MTMR2's role as a regulator of the myelination process is evidenced in MTMR2's protein-protein interactions in Schwann cells and axons, where it has been shown to interact with Disks large 1 (DLG1) and neurofilament light chain (NFL) in myelinating and non-myelinating Schwann cells. DLG1, is expressed in neuronal cells, binding and transporting glutamate receptors to the membrane through its membrane-associated guanylate kinase function.[37] MTMR2 interacts directly with DLG1 utilizing a PDZ domain of DLG1 at Schwann cell paranodes, where myelin outfoldings arise.[4,38] When MTMR2 is bound to the DLG1 complex, myelination is stopped, preventing further myelin outfolding formation, ultimately regulating myelin homeostasis.[38] NFL, a protein allowing for intracellular transport to axons and dendrites, causes CMT4E when mutated, and is also able to interact with MTMR2.[39,40] MTMR2's involvement in the endolysosomal system is further shown in its interaction with vacuolar protein sorting 34 (Vps34)/Vps15 PI3K complex, critical for

generating PI3P on endosomes. A report found that both wildtype MTMR2 and catalytically inactive mutants could be co-enriched with this complex.[41] It is evident that proper functioning and interaction of MTMR2 with other proteins is critical for the regulation of endosomal dynamics and phosphoinositide regulation.

Understanding the regulation and interactions of the lipid phosphatase MTMR2 may link the endosomal PI3P-dependent signaling events to the onset of CMT4B. It is the goal of this work to examine MTMR2 regulation, through protein-protein interactions and post-translational modifications, as well as its effect on the endosomal landscape and its effectors. Through treatment of MTMR2 with phosphatase inhibitors and generation of MTMR2-BioID constructs, we have identified potential Schwann cell-specific regulatory events acting on MTMR2 and proteins affected by MTMR2 function at the endosome.

MATERIALS AND METHODS

Transformation of plasmids into competent *E. coli cells*

DH5α competent *E. coli* cells were used to generate multiple copies of plasmid DNA. 40 µL of *E. coli* cells were incubated with 1–2 µg of plasmid DNA on ice for 30 minutes. Following this incubation, cells were heat shocked at 42˚C for 45 seconds followed by a 2-minute incubation on ice. 500 µL of LB media was then added and placed in a shaking incubator for 1 hour at 37˚C. Cells were then pelleted at 12,000 rpm for 1 minute and 400 µL of media was removed. The cells were then resuspended in the remaining 100 µL of LB media and plated on LB agar plates containing the appropriate antibiotics. Colonies were allowed to grow overnight.

Plasmids

N-terminally FLAG-tagged MTMR2 made as described Kim *et al*. and Taylor *et al*.[42,43] pCDNA3.1-NF-MTMR2 (cDNA KIAA1073) was used in this study. Site directed mutagenesis (SDM) was used to create MTMR2 serine 58 mutations, FLAG-MTMR2^{S58A} and FLAG-MTMR2^{S58E}, made prior to this study.[35] SDM was used on a second site, Ser^{631}, to generate FLAG-MTMR2^{S631A} and FLAG-MTMR2^{S631E}, as well as S58A/631A, S58A 631E, 58E/631A, and 58E/631E FLAG-tagged double mutants.[35] MYC-constructs, MYC-MTMR2^{WT}, MYC-MTMR2^{S58A}, and MYC-MTMR2^{S58E} were also utilized. The pEGFP-2xFYVE construct was a generous gift from Harald Stenmark[44] and DsRed-FYVE, previously generated, was also used.[45] pClneo-HA-TSSC1 and dsRed-TSSC1 were a generous gift from Juan Bonifacino.

Maxiprep of plasmid DNA

Desired plasmid DNA from bacterial cultures using Maxiprep kits from Sigma Aldrich and Qiagen (B34-X4G). Bacterial colonies containing genes of interest were grown in 5 mL of LB media starter cultures supplemented with the appropriate antibiotics and grown for 6–8 hours shaking at 37˚C. Starter cultures were then added to 150 mL LB media with the appropriate antibiotic and incubated for 14–18 hours at 37˚C, at which point cultures would have an optical density between 1.4–1.8. Cultures were then centrifuged at 6,000 *g* for 20 min at 4˚C to pellet cells. Supernatants were discarded and pellets were resuspended according to manufacturer's instructions. DNA was eluted from DNA precipitator columns with either water or TE buffer. DNA concentration and purity were measured using an UV/Vis spectrophotometer (Biochrom, Ultrospec 2100 *pro*).

Cell culture and transfection

A variety of cell lines were used in these studies (obtained from ATCC). Human embryonic kidney cells 293 (HEK293), HeLa, RT4 Schwann cells, and U2OS. were grown in Dulbecco's Modified Eagle's Medium/Nutrient F-12 Ham (DMEM-F12 HAM; Corning) supplemented with 10% FBS (Gibco) and antibiotics (100 units/mL penicillin, 100 µg/mL streptomycin) at 5% CO_2 and 37˚C. Cells were maintained at 70% confluency in 100 mm or 150 mm culture dishes (Starstedt). Cells were grown in DMEM-F12 HAM (Corning) supplemented with 5% FBS (Gibco) and antibiotics (100 units/mL of penicillin and 100 µg/mL of streptomycin; Sigma). Cells were plated 24-hour prior to transfection at a range from 500,000–750,000 cells/mL in antibiotic-free media. At the time of transfection 7.5–10 µg of plasmid DNA was added to 100 µL of non-supplemented media and incubated for 10 minutes in the dark at room temperature, meanwhile 100 µL of media

is added to 20–40 µL of 1 mg/mL PEI. Following this incubation, the contents of the PEI/media mixture is added to the DNA/media, flicked, mixed, and incubated for 10-13 minutes. DNA/PEI mixtures were added onto the plates in a dropwise manner and incubated at 37˚C for 24–48 hours prior to cellular treatment(s) or lysis.

Large-scale transient transfections were used for MS analysis. Here, cells were plated in antibiotic-free media 24-hours prior to transfection at a density of 1.5 million cells/mL determined by manual cell counting using a hemocytometer. Cells were maintained in 150-mm cell culture dishes (Starstedt). Transfection reactions consisted of 20 µg of plasmid DNA incubated with 200 µL media, while 40 µL of PEI was incubated with 200 µL of media. All reactions were incubated in the dark for 10 minutes.

Alkaline phosphatase treatment

In vitro phosphatase assays testing the potential of MTMR2 dephosphorylation occurred in 20-µL volume reactions. Reactions were made on ice composed of 2 µg of total protein was added with 10 units of alkaline phosphatase, 2 µL 10X FAST AP buffer (Thermo Fisher Scientific) and made up with water. Reactions were incubated at 37˚C for 1 hour and quenched by the addition of 20 µL 6X reduced SDS-PAGE loading dye (0.35 M tris-HCl, 10% SDS, 0.3% β-mercaptoethanol, 30% glycerol, 0.012% bromophenol blue, 3% dithiothreitol, pH 6.8) and boiled at 98˚C for 5 min. Resulting products were analyzed via immunoblotting.

The phosphorylation status of MTMR2 was examined using mass spectrometry. To produce a suitable amount of phosphorylated protein, HEK293 cells were transiently transfected with various FLAG-MTMR2 constructs, treated with phosphatase and proteasome inhibitors, lysed, and underwent FLAG-immunoprecipitation.

Cells were treated with 10 μM MG132 for 6 hours prior to lysis. For cells cotreated with MG132 and calyculin A, cells were first treated with MG132, followed by calyculin A after 5.5 hours of the MG132 treatment.

Phosphatase inhibitor treatment

To increase global phosphorylation of proteins in cell culture, cells were treated with 5 nM of calyculin A (CalBioChem) or 1 μM okadaic acid (LC Laboratories) for 30 min at 37˚C or 10 μM cyclosporine A (LC Laboratories) for 24 hours at 37˚C. Following this, cells treated with calyculin A or okadaic lost their ability to adhere to the cultures plated and were transferred to 15 mL conical tubes and spun at 5,000 rpm for 5 minutes at 4˚C. Pellets were then resuspended with cold PBS and pelleted at 5,000 rpm for 5 minutes. Pellets were then prepared for cell lysis.

Cellular lysis

For calyculin A and okadaic acid-treated cells, pellets were then resuspended in lysis buffer (50 mM tris-HCl, 1% Triton X-100, 150 mM NaCl, 0.1% SDS, pH 7.4) supplemented with protease and phosphatase inhibitors and incubated on ice for 5 minutes. Meanwhile, adherent cells, including DMSO-vehicle and cyclosporine A-treated cells, were washed with cold PBS and collected with lysis buffer and incubated on ice for 5 minutes. Lysates were then spun at 15,000 rpm for 10 minutes at 4˚C. Supernatants were collected and protein concentrations were determined via Bradford assay.

FLAG immunoprecipitation

FLAG conjugated to agarose beads (Sigma) were prepared for immunoprecipitation. 10 μL of resin was used per immunoprecipitation, by washing with 500 μL of lysis buffer spinning at 5,000 rpm at 4˚C for 1 minute. Resins were then

resuspended with lysis buffer. Fresh lysates were added directly onto beads and incubated for 1.5 hours rotating at 4˚C. Following this incubation, resins were rinsed with wash buffer (50 mM Tris-HCl, 0.1% Triton X-100, 150 mM NaCl, 0.1% SDS, pH 7.4) 3 times, spun at 5,000 rpm. Following supernatant removal, 35 µL of reduced 6X loading dye is added onto the beads and boiled for 5 minutes.

SDS-PAGE and immunoblotting

Whole cell lysates and immunoprecipitations was evaluated using immunoblotting using sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). Protein samples were resolved on 10% gels for 2 hours 20 minutes at 140 V in 1X SDS-PAGE running buffer (0.192 M, 0.025 M Tris-base, 0.01% SDS, pH 8.3). Following separation, proteins were transferred to polyvinylidene difluoride (PVDF) membrane (Millipore Corp.) or nitrocellulose (Millipore) using cold 1X transfer buffer (20% methanol, 0.192 M glycine, 0.025 M tris-base, pH 8.3) for 1 hour at 100 V. Following transfer, membranes were blocked with 5% skim milk or bovine serum albumin (BSA; Sigma) for phosphoblots, in 1X Tris-Buffered Saline (1X TBS; 0.068 M NaCl, 8.3 mM Tris-base, pH 7.6) containing 0.1% Tween®-20 (1X TBSt; Fisher Scientific). Following blocking at room temperature for 1 hour with gentle agitation, blots were probed overnight with primary antibody at 4˚C with gentle agitation. Primary antibodies used were mouse anti-FLAG (Sigma), mouse anti-MTMR2 (Santa Cruz), mouse anti-SGK3 (Santa Cruz), mouse anti-ubiquitin (P4D1, Cell Signaling Technology) at 1:3,000 in 2.5% milk and TBSt, and rabbit anti-MTMR2^{pSer58}. Rabbit anti-MTMR2^{pSer58} antibody and streptavidin-HRP (Pierce) were made in 2.5% BSA and TBSt at a dilution of 1:1,000 and 1:5,000, respectively. Blots were then washed with TBSt for 5 minutes three times and incubated for 45 minutes at room

with secondary antibody. Goat anti-mouse-HRP (Sigma) secondary antibody was made in 2.5% milk at 1:5,000, while goat anti-rabbit-HRP (BioRad) was made in 2.5% BSA at 1:5,000. Blots were washed with TBSt as above and ready for imaging. bands were visualized via chemiluminescence using the SuperSignal® West Femto Maximum Sensitivity Substrate (Thermo Fisher) according to the manufacturer's instructions.

Production of BirA-MTMR2 fusion constructs

BirA-destination vector possessing a N-terminal FLAG tag, pDEST-pcDNA5_BirA_FLAG N-term, was generously donated by Brian Raught (Sick Kids, Toronto). MTMR2 fusion constructs were generated through the use of gateway cloning (Thermo Fisher). To do this, *att*B sites were added onto wildtype and Ser58 phosphomimetic (S58E) and phosphodeficient (S58A) MTMR2 cDNA templates with the following primers: forward primer 5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGAGAAGAGCTCGAGC-3' and reverse primer 5'-GGGGACCACTTTGTACAAGAAAGCTGGGTATTATACAAGAAAGCTGGGTATTATACAACAGTTTGGAC-3'. PCR reactions generated the MTMR2 coding sequence with *att*B_1/*att*B_2 sites. Following PCR amplification, products were mixed with loading dye and resolved on 1% agarose gel containing ethidium bromide. Gels were visualized, and desired bands were excised. DNA concentration was measured by NanoDrop.

50 ng of each purified amplicon was then mixed with 50 ng of pDONR vector pDONR221 (Invitrogen), to a total final volume of 2 µL, and incubated with 0.5 µL of BP clonase II enzyme (Invitrogen) at 25˚C, overnight. Following this initial recombination reaction, the reaction volume was increased to 10 µL with TE buffer. TOP10 *E. coli* cells

(Invitrogen) were electroporated with 2 µL of the above reaction. Cells were then plated onto LB kanamycin-containing plates. Following an overnight incubation at 37˚C, pENTR221:MTMR2 plasmids were extracted and digested for verification. Destination vectors were created through the addition of 100 ng of pENTR and 100 ng of destination vector for a total volume of 2 µL, added to 0.5 µL of LR clonase II (Invitrogen) and electroporated into TOP10 electrocompetent *E. coli* cells. Resulting pDEST vectors were ampicillin-resistant (100 µg/mL) and grown on LB-ampicillin plates. Colonies were selected, grown, and plasmid DNA was extracted. Resulting FLAG-BirA-MTMR2 vectors sequences were verified through DNA sequencing by AGCT Corporation using in-house primers.

Generation of BirA-MTMR2 mutants

Various MTMR2 Ser58 mutants were produced. The phosphomimetic mutant, $MTMR2^{S58E}$ was generated using the forward primer 5'-CTTCTGCCGACAACTTTGAGCCTGATTTGAG GGTCC-3' and reverse primer 5'-GGACCCTCAAATCAGGCTCAAAGTTGTCGGCAGAAG-3'. Phosphodeficient MTMR2, $MTMR2^{S58A}$, was generated using forward primer 5'-CTGCCGACAACTTTGCTCCTGATTTGAGGGTC-3' and reverse primer 5'-GACCCTCAAATCAGGAGCAAAGTTGTCGG- CAG-3'. Mutants were generated through 50-µL reactions containing 500 nM of both forward and reverse primers, 5 ng of template DNA, 10 µM dNTP mixture, 5 µL Pfu reaction buffer, and 1 µL Pfu polymerase (Promega). Resulting PCR products were digested with 1 µL of DpnI (New England BioLabs Limited) at 37˚C for 1 hour. Digested products were then transformed into

ampicillin resistant *E.coli* competent cells and grown overnight at 37 ˚C. Colonies were then selected and sequenced for mutagenesis (AGCT Corporation).

Transient expression of BirA-fusion proteins

BirA-MTMR2 fusions were predominantly overexpressed in RT4 Schwann cells. Cells were plated at densities of 500,000 cells/mL in antibiotic free media. 24 hours post-plating, cells were transfected using linear polyethyleneimine as described above using a DNA:PEI ration of 1:2, with 7.5 µg of DNA. The initial incubation of DNA with media and PEI with media was done at 10 minutes in the dark, followed by an 11-minute DNA:PEI incubation prior to addition onto plates. 6 hours post-transfection, cells were washed with PBS and fresh media containing antibiotics and serum supplemented with 50 µM L-biotin. Biotin was dissolved in media without serum or antibiotic supplementation. Cells were lysed 24 hours post-transfection (18 hours post-biotin addition). In order to minimize the potential suppression of signals generated by the BirA-MTMR2 fusion constructs in downstream analyses, a preclearing step was implemented. Here, an anti-FLAG M2 IP was performed for 30 minutes at 4 ˚C. The flow-through from this enrichment was subsequently added onto Dynabeads™ MyOne™ streptavidin C1 beads (Invitrogen) preconditioned with lysis buffer. For optimization experiments, 10 µL of Neutravidin beads (Pierce) were used, whereas 15 µL of streptavidin C1 beads per enrichment were utilized in large-scale enrichments. Following an overnight pulldown, 40 µL of reduced 6X loading dye supplemented with excess biotin was added onto beads and boiled or prepared for in-solution trypsin digestion.

Optimization of biotin labeling

Biotin addition was optimized to ensure the maximal labeling of proteins in the cell. The first parameter tested following transfection efficiency (data not shown) was the biotin labeling time, adding biotin 30 mins, 2 hours, or 18 hours prior to lysis. Both 2 and 18 hours of biotin addition allowed for labeling of proteins. To avoid not capturing interactions that occur well before the 2-hour timepoint at which biotin was added, we proceeded with adding biotin 18 hours prior to lysis. The next parameter optimized was the concentration of biotin. Biotin was added at concentrations of 50, 100, and 250 µM for 18 hours. Biotin labeling was detected at all concentrations used. For subsequent experiments, biotin was supplemented at 50 µM.

Localization studies of BirA-MTMR2 fusion constructs

The localization of fusion constructs was evaluated using immunofluorescence assay (IFA). HeLa cells were transfected with FLAG-BirA-MTMR2 fusion constructs at 80,000 cells/mL were plated in 1-mL chambers (BioBasic SP41219) in antibiotic-free media. 24 hours post-plating, 1.5 µg of each construct was transfected into each well using Lipofectamine® 3000 (Invitrogen) in antibiotic-free media. 6 hours post-transfection, cells were washed with PBS and antibiotic-containing media was added to cells. Following an 18-hour incubation at 37˚C, cells were then washed with PBS twice and fixed using 3.7% paraformaldehyde (Fisher) for 15 minutes at room temperature and washed with PBS. Cells were permeabilized using 0.15% Triton X-100 in 1X TBS (0.068 M NaCl, 8.3 mM Tris-base, pH 7.6) for 2 minutes at room temperature and washed with 1X TBS. Fixed cells were blocked for 1 hour at room temperature, shaking, in 5% BSA in PBS. Cells were then incubated in primary antibody (mouse anti-FLAG 1:500 (Sigma), mouse anti-MTMR2

(Santa Cruz), and rabbit anti-Rab5 1:500) for 1.5 hours at room temperature, shaking, in 1% BSA. Cells were washed with PBS thrice and incubated with secondary antibody (horse anti-mouse-FITC (Vector Laboratories) and goat anti-rabbit-AlexaFluor® 568 (Vector Laboratories)) for 1 hour at room temperature. Following the secondary antibody incubation, cells were watched 3X with PBS and nuclei were stained with Hoechst 33342 (Molecular Probes) for 2 minutes at room temperature and washed with PBS and dried before coverslip mounting. Once dry, coverslip was mounted with 50% glycerol/PBS and sealed with nail polish.

Epifluorescence image acquisition

Protein localization was examined using a Leica DMIRB microscope equipped with a Q-imaging CCD camera using LAS AF software. Epifluorescence images were acquired using a 40X oil objective with a DAPI, Texas red, and FITC filter. Images were processed using ImageJ and Inkscape 1.0.

Optimized BioID protocol

Bait constructs from whole cell lysates were precleared by adding 10 µL of FLAG-IP beads (Sigma) for 1.5 hours, rotating at 4˚C. FLAG-IP beads were then magnetized, allowing for collection of the soluble portion. This soluble portion was then added to 15 µL of streptavidin beads overnight, rotating at 4 ˚C. Beads were then washed with IP wash buffer thrice, followed by three PBS washes, and a final wash with 50 mM ammonium bicarbonate.

In-solution digestion of immunoprecipitated material

Identification of proteins captured through BioID were prepared by in-solution digestion followed by MS analysis. FLAG IP and streptavidin beads were resuspended in

50 mM ABC. 1 mL of $CaCl_2$ solution was added (20 mM Tris-HCl, 2 mM $CaCl_2$, pH 8). Beads were then conditioned with three 500-µL washes of 50 mM NH_4HCO_3 (Thermo Fisher Scientific). 100 µL of NH_4HCO_3 was added and supplemented with 500 ng (1:50) of mass spectrometry grade trypsin (Promega). Following overnight digestion, peptides were collected, and an additional 500 ng of trypsin in 50 mM NH_4HCO_3 was added and incubated for 4 hours at 37˚C. Reactions were quenched by the addition of formic acid (Sigma) to a final concentration of 2%. Peptides were then desalted as described above.

In-gel digest of immunoprecipitated material

Immunoprecipitated (IP) materials isolated from cell lines were prepared for sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) separation. IPs were prepared for SDS-PAGE analysis by mixing with reduced SDS loading dye (3% glycerol, 0.03 g bromophenol blue, and 10% SDS), and boiled for 5 minutes. Solubilized samples were then loaded on 10 or 12% gels and separated. Proteins were then visualized through staining with Imperial™ stain (Thermo Fisher 24615).

Whole lanes were excised into multiple 1 cm x 1 cm gel slices and destained using acetonitrile (ACN; Burdick and Jackson) and 50 mM NH_4HCO_3 (Fisher Scientific) for 15 minutes at 37˚C. Following destaining, gel pieces were dehydrated by addition of neat ACN and a brief incubation at room temperature. Post-dehydration, gel pieces were fully dried by vacuum centrifugation (Sorvall SPD121P P1). Gel pieces were rehydrated by addition of ~1.5 µg of trypsin (Promega) in 50 mM NH_4HCO_3 on ice for 1.5 hours, and incubated overnight at 37 ˚C. Extracts were then collected and 150 µL of extract solution [1:2 of 5% Formic acid (Fluka):ACN] was added to the gel pieces and incubated at 37˚C

for 15 minutes, twice. All extracts were pooled in siliconized tubes and vacuum concentrated to ~10 µL and prepared for desalting.

Concentrated tryptic peptides were reconstituted in 100 µL of 0.1% trifluoroacetic acid (TFA, Fluka) and desalted using Oasis CLB 1cc (Waters). Oasis columns were preconditioned by addition of 100 µL of 100% ACN/0.1% TFA, washed twice with 100 µL of 0.1% TFA, and exosome-derived peptides were allowed to bind the columns. Columns were washed twice with 100 µL of 0.1% TFA. A stepwise elution method was utilized, first using 100 µL of 30% ACN/70% 0.1% TFA, followed by 100 µL 50% ACN/50% 0.1% TFA, and finally twice with 150 µL 80% ACN/20% 0.1% TFA. Desalted peptides were then vacuum concentrated to dryness.

Solution pH

For all solutions and pH measurements, an Orion 410A+ pH meter (Thermo Electron Corporation) was used. The pH meter was calibrated using three standard solutions of pH values 3, 7, and 10 (Orion^{TM} pH).

LC-MS/MS settings

Peptides were reconstituted in 20 µL of 0.1% formic acid (FA) and loaded onto nanoAcquity ultra performance liquid chromatography (UPLC) system for LC-MS^{E} (Waters, Milford, MA). Peptides were injected onto a 1.8-µm HSS T3 75 µm x 150 mm reverse-phase column (Waters) at a flow rate of 0.3 µL/min using a nanoAcquity UPLC system. Mobile phase A (0.1% formic acid in water) and mobile phase B (ACN with 0.1% formic acid) were used to equilibrate the column. Peptides were loaded onto the column with a 97:3 (A/B). Elution of peptides occurred using an 85-min ACN gradient (3-30% B 60 min, 30-50% B for 15 min, 85% B for 10 min) and directly sprayed into a SYNAPT

G2-Si mass spectrometer (Waters, Miliford, MA) using a cone voltage of 30 V and a capillary voltage of 3 kV. The instrument was operated in data independent acquisition mode with ion mobility activated. Low energy scans were captured at 4 eV, while high energy scans were captured ramping from 20–45 eV in positive resolution mode scanning from 50 to 2,000 *m/z* at a rate of 0.8 s. Calibration was achieved using [Glu1]-fibrinopeptide B (50 fmol/μL) in the lockmass channel at *m/z* 785.8427 for a doubly charged positive ion. Raw data was collected using Mass Lynx version 4.1 and processed using Progenesis QI (Nonlinear Dynamics) or ProteinLynx Global SERVER (PLGS), allowing for chromatographic alignment, normalization, and peptide identification using the human UniProtKBSwissProt database (20,614 proteins acquired May 23, 2014). For proteins derived from RT4 Schwann cells, the Rat UniprotSwissProt database was used (29,981 acquired October 27, 2017). Raw data was processed according to the following parameters: low energy noise reduction of 135 counts, high energy reduction noise reduction threshold of 30 counts, and an intensity threshold of 750 counts. Lock mass correction occurred post-acquisition. Protein identification according to the following parameters: a minimum of 3 fragment ions per peptide, minimum of 7 fragment ion matches per protein, minimum of 1 peptide match per protein. The maximum false positive rate was set to 4, with a maximum of two missed cleavages following trypsin digestion. The variable modifications of methionine oxidation +15.9949, phosphorylation of Ser/Thr/Tyr +79.9663, and biotinylation +226.0776 were added. Three technical replicates per biological replicate were grouped and compared in Progenesis QI. Conflicts were resolved followed by quantitation using the three most abundant peptides per protein.

RESULTS

MTMR2^{Ser58} is accessible to dephosphorylation

In order to investigate reversible phosphorylation of MTMR2^{Ser58}, we evaluated the accessibility of this site to the general phosphatase, alkaline phosphatase (APP), using MTMR2 derived from cell lysates. Cell extracts were treated with APP and phosphorylation status was measured through immunoblotting with a MTMR2^{pSer58}-specific antibody. MTMR2 treated with alkaline phosphatase displayed a reduction in pSer58 levels compared to control reactions absent of alkaline phosphatase (**Figure 4.2**).

Screening a panel of phosphatase inhibitors

To identify the phosphatase that regulates MTMR2^{pSer58} we overexpressed MTMR2 in HEK293 cells and treated them with a panel of phosphatase inhibitors in attempts to inhibit the elusive phosphatase. FLAG-MTMR2 was enriched for using FLAG IP and analyzed using immunoblotting probing for MTMR2^{pSer58}. No significant increases in phosphorylation could be detected using this method, however mobility shifts were reproducibly observed for cells treated with calyculin A or okadaic acid (**Figure 4.3A** and **B**).

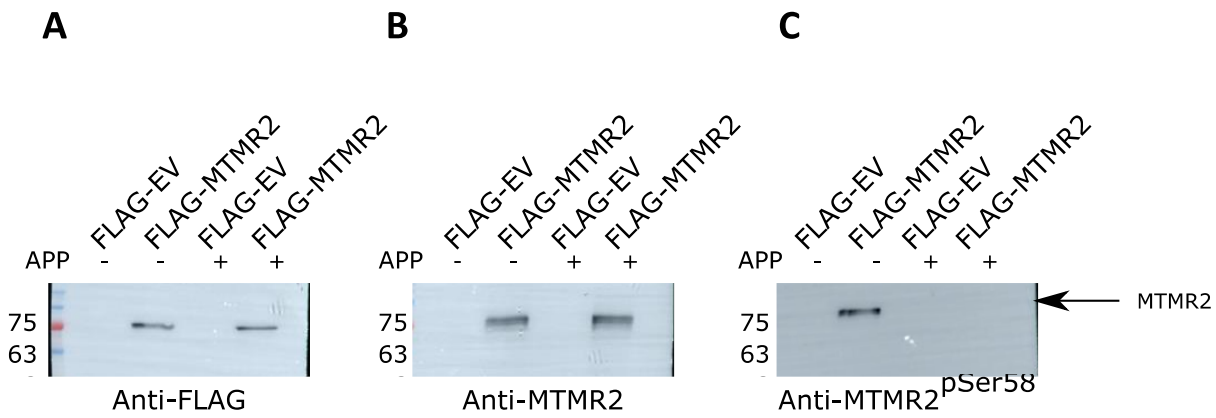
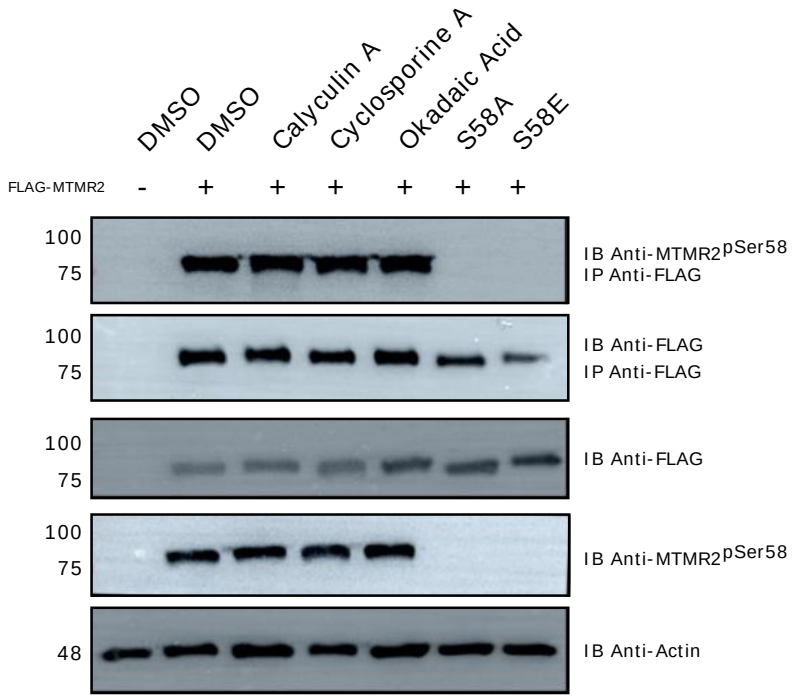


Figure 4.2. MTMR2 pSer^{58} can be dephosphorylated in vitro.

Protein extracts from HEK293 cells transiently expressing FLAG-MTMR2 were treated with 10 units of alkaline phosphatase (APP) and subjected to immunoblotting, probed with (**A**) anti-FLAG, (**B**) anti-MTMR2, and (**C**) anti-MTMR2^{pSer58} to determine the accessibility of serine 58 to dephosphorylation.

A


DMSO
DMSO
Calyculin A
Cyclosporine A
Okadaic Acid
S58A
S58E
FLAG-MTMR2 - + + + + + +
100
75
IB Anti-MTMR2pSer58
IP Anti-FLAG
100
75
IB Anti-FLAG
IP Anti-FLAG
100
75
IB Anti-FLAG
100
75
IB Anti-MTMR2pSer58
48
IB Anti-Actin

B

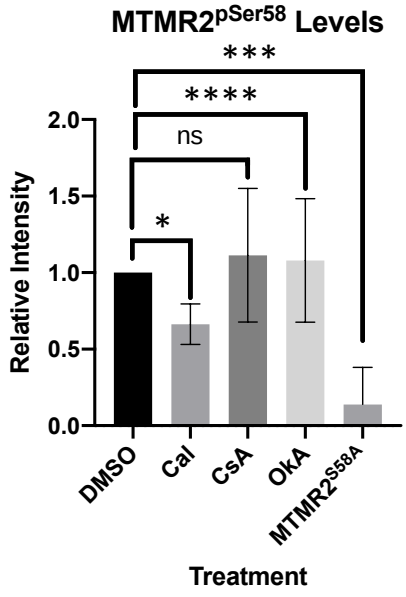

MTMR2pSer58 Levels
***
****
ns
*
2.0
1.5
1.0
0.5
0.0
Relative Intensity
DMSO
Cal
CsA
OkA
MTMR2S58A
Treatment

Figure 4.3. Phosphatase inhibitor screening.

In an attempt to increase MTMR2^{pSer58} levels, HEK293 cells were transiently transfected with FLAG-MTMR2 constructs and treated with a panel of phosphatase inhibitors: calyculin A (Cal), cyclosporine A (CsA), and okadaic acid (OkA). Following treatment, cells were lysed and FLAG-MTMR2 was enriched through IP.

(**A**) Whole cell lysates and IPs were subjected to immunoblotting probing for FLAG-MTMR2 and MTMR2^{pSer58}.

(**B**) MTMR2^{pSer58} levels were quantified from 3 independent transfections and plotted as mean ±standard deviation (S.D.), **P*<0.05, ****P*<0.0001, and *****P*<0.00001 using an unpaired Student's *t* test compared to DMSO treated cells expressing FLAG-MTMR2.

ns = not significant.

Affinity purification of the MTMR2 phosphatase in question

Understanding the MTMR2 interactome at and away from the endosome will provide details as to how MTMR2 is regulated and how the protein itself can regulate the myelination program. To determine the identity of these interacting proteins and to complement our phosphatase inhibitor approach, we attempted to co-enrich MTMR2's phosphatase through expression of FLAG-MTMR2 in Schwann cells, and perform in-solution trypsin digestion to create peptides to be analyzed through LC-MS/MS (**Figure 4.4**). Using this method, we were able to enrich for many proteins consistent with the proteome of Schwann cells. Several peptides belonging to protein phosphatase 1B (PP1B; P35815.1) were identified in both FLAG-MTMR2^{WT} and FLAG-EV expressing cells. Of the proteins identified (**Table 4.1**), MTMR2 was uniquely found in MTMR2-overepxressing cells. The PANTHER classification system was used to generate gene ontology terms of proteins uniquely found in MTMR2-expressing cells following anti-FLAG enrichments. Proteins enriched were found to be primarily located in vesicles and exosomes (**Figure 4.3C**) and involved in organelle localization (**Figure 4.3D**).

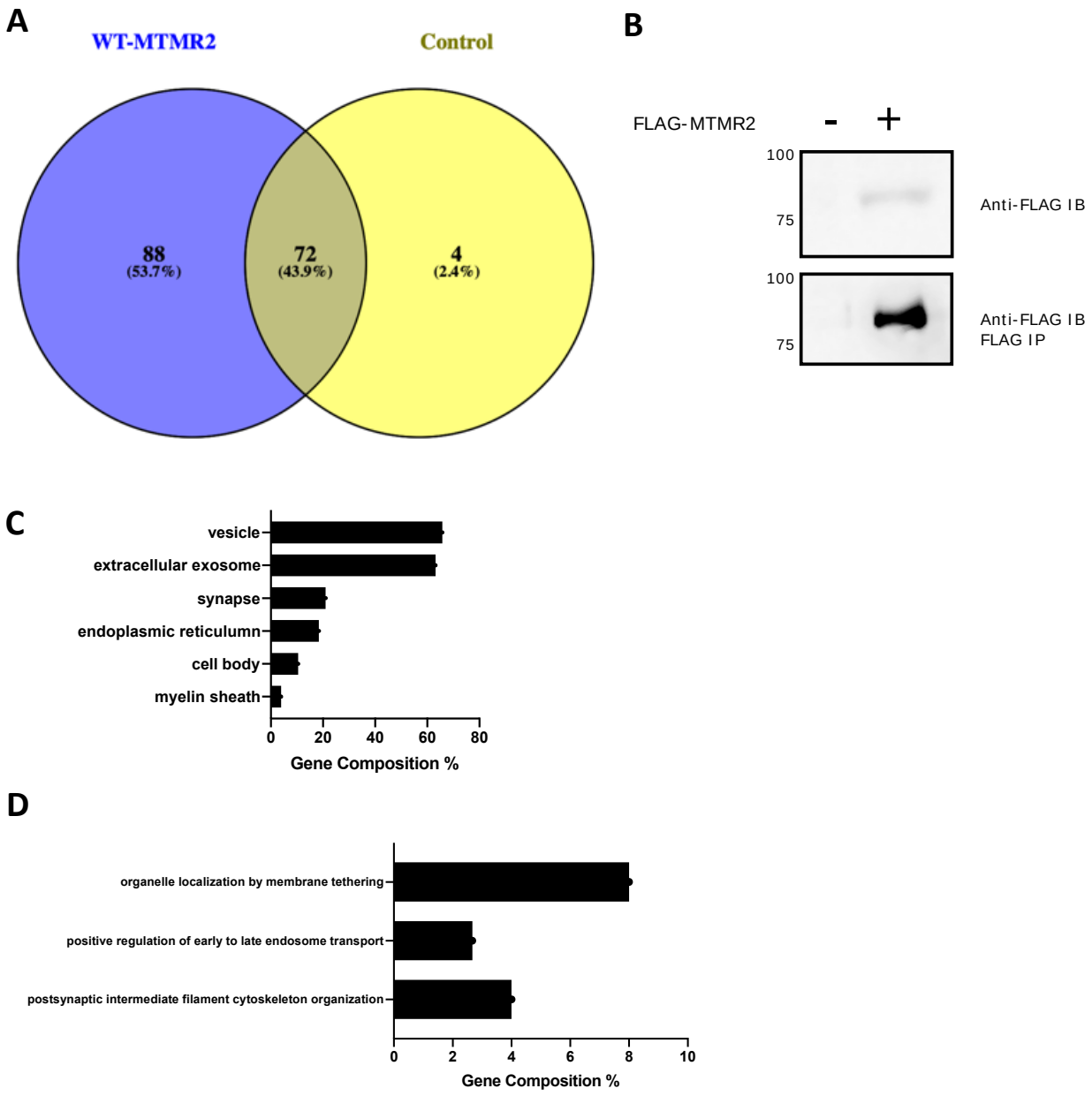


Figure 4.4. Affinity purification followed by mass spectrometry.

FLAG-MTMR2 was expressed in Schwann cells in order to copurify the Ser^{58} phosphatase in question, followed by in-solution digestion to identify the isolated proteins via mass spectrometry.

(**A**) A Venn diagram showing the numbers of proteins found in MTMR2-expressing and untransfected control cells from 2 biological replicates is shown. Venn diagram generated using Venny (https://bioinfogp.cnb.csic.es/tools/venny/index.html).

(**B**) Representative immunoblots of cell lysates prior to IP (top) and after immunoprecipitation but prior to digestion (bottom) were captured using anti-FLAG. Full list of identified proteins is presented in **Table 4.1**.

(**C**) Gene ontology terms related to the cellular component of proteins derived using the PANTHER classification system collected from co-enriched proteins identified in FLAG-MTMR2-expressing cells.

(**D**) Select gene ontology terms generated using the PANTHER classification system relating to the biological function of proteins enriched using FLAG IP were converted to gene composition percentages.

Table 4.1: WT MTMR2 FLAG-IP identifications using AP-MS protein.

A full is of identifications can be seen in **Appendix C**.

Accession	Protein Name	Accession	Protein Name
Q13614.4	Myotubularin-related protein 2	P58775.1	Tropomyosin beta chain
P68370.1	Tubulin alpha-1A chain	Q63610.2	Tropomyosin alpha-3 chain
Q6AYZ1.1	Tubulin alpha-1C chain	P69897.1	Tubulin beta-5 chain
Q6P9V9.1	Tubulin alpha-1B chain	P29457.1	Serpin H1
P12839.4	Neurofilament medium polypeptide	P84245.2	Histone H3.3
P16884.4	Neurofilament heavy polypeptide	Q6LED0.3	Histone H3.1
P19527.3	Neurofilament light polypeptide	Q9Z1Z1.1	Eukaryotic translation initiation factor 2-alpha kinase 3
P23565.2	Alpha-internexin	P62963.2	Profilin-1
Q68FR8.1	Tubulin alpha-3 chain	P62243.2	40S ribosomal protein S8
Q9JLT0.1	Myosin-10	P70501.1	RNA-binding protein 10
P62250.2	40S ribosomal protein S16	P62856.3	40S ribosomal protein S26
Q5XIF6.1	Tubulin alpha-4A chain	Q6P9T8.1	Tubulin beta-4B chain
Q925G0.2	RNA-binding protein 3	P62902.1	60S ribosomal protein L31
Q5BJY9.3	Keratin_ type I cytoskeletal 18	P45592.3	Cofilin-1
Q9WUL0.1	DNA topoisomerase 1	Q5XIB4.1	Ufm1-specific protease 2
P22509.2	rRNA 2'-O-methyltransferase fibrillarin	Q62908.3	Cysteine and glycine-rich protein 2
P62859.1	40S ribosomal protein S28	P35213.3	14-3-3 protein beta/alpha
P62752.1	60S ribosomal protein L23a	P61983.2	14-3-3 protein gamma
P12001.2	60S ribosomal protein L18	P68255.1	14-3-3 protein theta
Q6AY56.1	Tubulin alpha-8 chain	P68511.2	14-3-3 protein eta
Q07936.2	Annexin A2	P38983.3	40S ribosomal protein SA
Q6URK4.1	Heterogeneous nuclear ribonucleoprotein A3	P62755.1	40S ribosomal protein S6
P62268.3	40S ribosomal protein S23	P62804.2	Histone H4
P45352.1	Thymidylate synthase	P48721.3	Stress-70 protein_ mitochondrial
A7VJC2.1	Heterogeneous nuclear ribonucleoproteins A2/B1	P04256.3	Heterogeneous nuclear ribonucleoprotein A1
A9UMV8.1	Histone H2A.J	P04764.4	Alpha-enolase
P02262.2	Histone H2A type 1	P11980.3	Pyruvate kinase PKM
P0C169.2	Histone H2A type 1-C	Q62736.1	Non-muscle caldesmon
P0C170.2	Histone H2A type 1-E	O35763.3	Moesin
P0CC09.1	Histone H2A type 2-A	P85834.1	Elongation factor Tu_ mitochondrial
Q00728.2	Histone H2A type 4	P63036.1	DnaJ homolog subfamily A member 1
Q4FZT6.3	Histone H2A type 3	Q6MG61.1	Chloride intracellular channel protein 1
Q64598.3	Histone H2A type 1-F	P47198.2	60S ribosomal protein L22
Q5RJR8.1	Leucine-rich repeat-containing protein 59	Q9R1Z0.2	Voltage-dependent anion-selective channel protein 3
Q5RKI1.1	Eukaryotic initiation factor 4A-II	P62909.1	40S ribosomal protein S3
P09495.3	Tropomyosin alpha-4 chain	P40112.1	Proteasome subunit beta type-3

Construction and optimization of BioID vectors

Standard affinity purification MS (AP-MS) has many challenges, including its inability to capture transient interactions, as well as losing interactions during purification steps. To overcome these challenges, the BioID approach was developed. We employed the BioID approach by cloning MTMR2 into FLAG-BirA* vectors, constructing FLAG-BirA*-MTMR2^{WT}, the phosphodeficient mutant FLAG-BirA*-MTMR2^{S58A}, and the phosphomimetic mutant FLAG-BirA*-MTMR2^{S58E} (**Figure 4.5A**), from here 'BirA' refers to BirA*. Following optimization of transfection in various cell lines, we sought to optimize the biotin labeling time and the concentration of biotin to be added to the cells. Biotin amount was tested by supplementing cells with 50–250 µM of biotin and evaluated using immunoblotting probing for biotin using streptavidin-HRP. Signals were detected with all concentrations used, leading us to continue using 50 µM biotin (**Figure 4.5B**). The length biotin labeling time was next tested, in which cells expressing the BirA-MTMR2 construct were supplemented with 50 µM of biotin at 18, 2, or 0.5-hours prior to lysis, and analyzed using immunoblotting (**Figure 4.5C**). Biotin labeling was detected at all time points. A labeling time of 18 hours was used for all subsequent experiments.

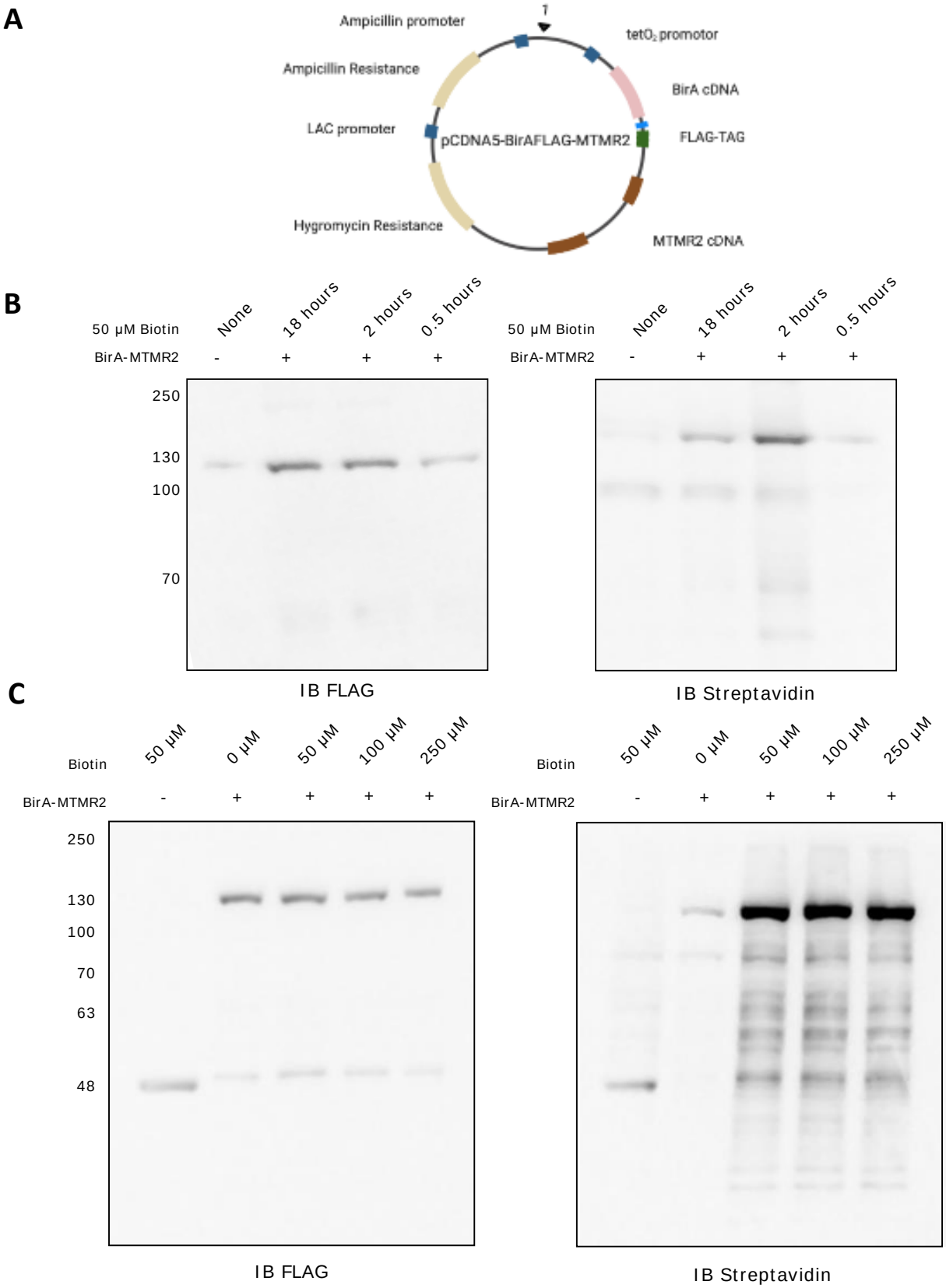

A
Ampicillin promoter
T
tetO2 promotor
Ampicillin Resistance
BirA cDNA
LAC promoter
pCDNA5-BirAFLAG-MTMR2
FLAG-TAG
Hygromycin Resistance
MTMR2 cDNA
B
50 µM Biotin
None
18 hours
2 hours
0.5 hours
BirA-MTMR2
- + + +
250
130
100
70
IB FLAG
IB Streptavidin
C
Biotin
50 µM
0 µM
50 µM
100 µM
250 µM
BirA-MTMR2
- + + + +
250
130
100
70
63
48
IB FLAG
IB Streptavidin

Figure 4.5. Optimization of FLAG-BirA-MTMR2 expression and biotin treatments.

In vivo labeling of proteins proximal to MTMR2 was achieved by creating BirA-MTMR2 fusion constructs.

(**A**) Schematic representation of BirA-FLAG-MTMR2 constructs. To achieve maximum labeling, biotin concentration and labeling time was performed by transfecting HEK293 cells with BirA-MTMR2^{WT}.

(**B**) Biotin-labeling time was evaluated using 50 µM for 0.5, 2, and 18 hours and

(**C**) biotin concentration, ranging from 50–250 µM. Biotin labeling was evaluated by immunoblotting with streptavidin antibody, with extracts from cells expressing BirA-EV or BirA-MTMR2^{WT} are indicated with "-" and "+", respectively. BirA represents BirA*.

Localization of BirA-MTMR2 fusion constructs

To ensure that the BirA-MTMR2 fusion construct functioned correctly, we evaluated its subcellular localization using immunofluorescence. The fusion construct was expressed in cells which were then fixed and probed with fluorescently conjugated antibodies. Fluorescence microscopy revealed that the fusion proteins localized similarly to FLAG-MTMR2 constructs (green),[45] with BirA-MTMR2^{S58A} strongly colocalizing with the endosomal marker Rab5 (red), fluorescing as yellow punctate structures (**Figure 4.6A**). BirA-MTMR2^{S58E} showed a more diffuse cytoplasmic pattern and a low amount of colocalization with Rab5 (**Figure 4.6B**). Cells expressing BirA-MTMR2^{WT} display some zones of colocalization with Rab5 (**Figure 4.6C**). Meanwhile, EV-BirA constructs were not visualized upon overexpression (**Figure 4.6D**). These results strongly suggested that the BirA-MTMR2 fusion constructs localized to the proper subcellular locations.

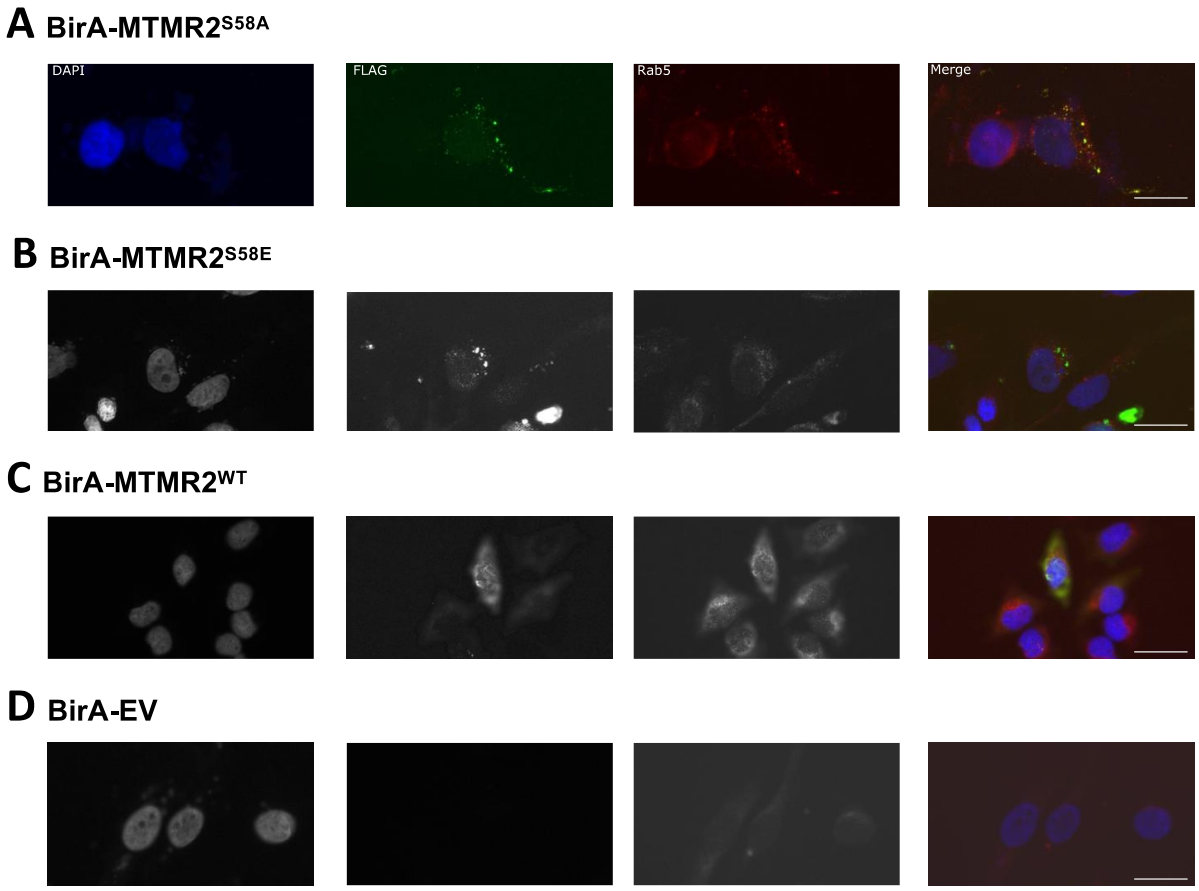


Figure 4.6. Subcellular localization studies of BirA-MTMR2 fusion proteins.

Validation of BioID construct localization was done by expressing the fusion constructs in a variety of cells and monitoring their cellular distribution alongside an early endosome marker, Rab5. Following expression, cells were fixed with paraformaldehyde and prepared for immunofluorescence analysis. BirA fused to (**A**) MTMR2^{S58A}, (**B**) MTMR2^{S58E}, (**C**) MTMR2^{WT}, and (**D**) EV-BirA were labeled with anti-FLAG and visualized Alexa Fluor® FITC (green) and rabbit anti-Rab5 probed with anti-rabbit AlexaFluor®568 (red) using epifluorescence microscopy. DNA was stained with Hoechst 33342 seen in blue. Scale bar represents 25 µm.

BioID: Identifying the MTMR2 interactome

Using our optimized conditions, we attempted to elucidate the interactome of MTMR2 by expressing several of our BioID vectors in Schwann cells. The BirA-MTMR2 fusion constructs were expressed, and cells were then supplemented with biotin to initiate the biotinylation reaction. Cells were then lysed, and lysates were precleared using FLAG beads, and biotin-labeled proteins were enriched using streptavidin. Captured proteins were subjected to in-solution trypsin digestion to generate peptides. The protein input prior to enrichment, as well as the precleared material, were subjected to immunoblotting (**Figures 4.7A–E**). FLAG preclearing was mainly composed of the bait proteins (**Figure 4.7B** and **C**), with streptavidin blots only illuminating the bait proteins (**Figure 4.7C**). Immunoblots following streptavidin binding also detect the bait as well as lower molecular weight proteins (**Figure 4.7E**). Our analysis provided identifications of peptides with/without biotin modification. Upon analysis of BirA-MTMR2^{S58A}, BirA-MTMR2^{WT}, and empty vector control an overlap of 42 proteins were found, those being involved in cytoskeletal structure (**Figure 4.8B**). A full list of protein identifications following enrichment from control and FLAG-MTMR2 overexpressing cells can be seen in **Table 4.2** and continued in **Appendix C**. The phosphomimetic mutant co-enriched with proteins involved in the UPS. One protein of interest that co-enriched with MTMR2^{S58A} was the tumour suppressing subtransferable fragment candidate gene 1 (TSSC1).

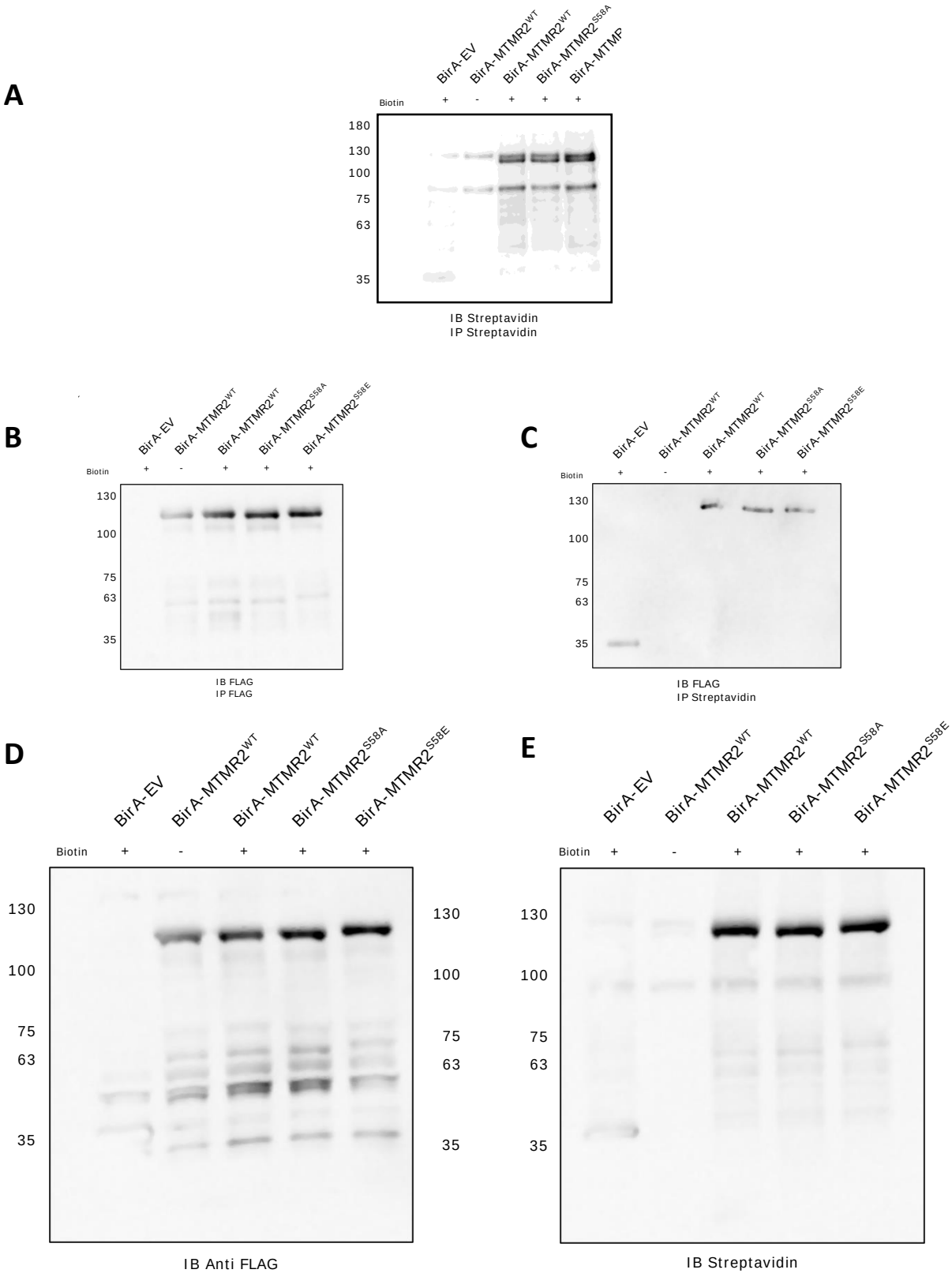

A
BirA-EV
BirA-MTMR2WT
BirA-MTMR2WT
BirA-MTMR2S58A
BirA-MTMP
Biotin + - + + +
180
130
100
75
63
35
IB Streptavidin
IP Streptavidin
B
BirA-EV
BirA-MTMR2WT
BirA-MTMR2WT
BirA-MTMR2S58A
BirA-MTMR2S58E
Biotin + - + + +
130
100
75
63
35
IB FLAG
IP FLAG
C
BirA-EV
BirA-MTMR2WT
BirA-MTMR2WT
BirA-MTMR2S58A
BirA-MTMR2S58E
Biotin + - + + +
130
100
75
63
35
IB FLAG
IP Streptavidin
D
BirA-EV
BirA-MTMR2WT
BirA-MTMR2WT
BirA-MTMR2S58A
BirA-MTMR2S58E
Biotin + - + + +
130
100
75
63
35
IB Anti FLAG
E
BirA-EV
BirA-MTMR2WT
BirA-MTMR2WT
BirA-MTMR2S58A
BirA-MTMR2S58E
Biotin + - + + +
130
100
75
63
35
IB Streptavidin

Figure 4.7. Expression of BirA-MTMR2 constructs in Schwann cells.

BioID constructs were expressed in Schwann cells to prepare for mass spectrometric analysis. Following expression, cells were supplemented with 50 µM biotin for 18 hours and lysed. Lysates underwent successive pulldowns: FLAG IP preclear followed by streptavidin enrichment. IPs and lysates were analyzed via immunoblotting using anti-FLAG and streptavidin-HRP. (**A**) Biotin enrichment probed with streptavidin-HRP, FLAG IP probed with (**B**) anti-FLAG and (**C**) streptavidin-HRP, respectively. Immunoblotting of lysates, probed with (**D**) anti-FLAG and (**E**) streptavidin-HRP, respectively.

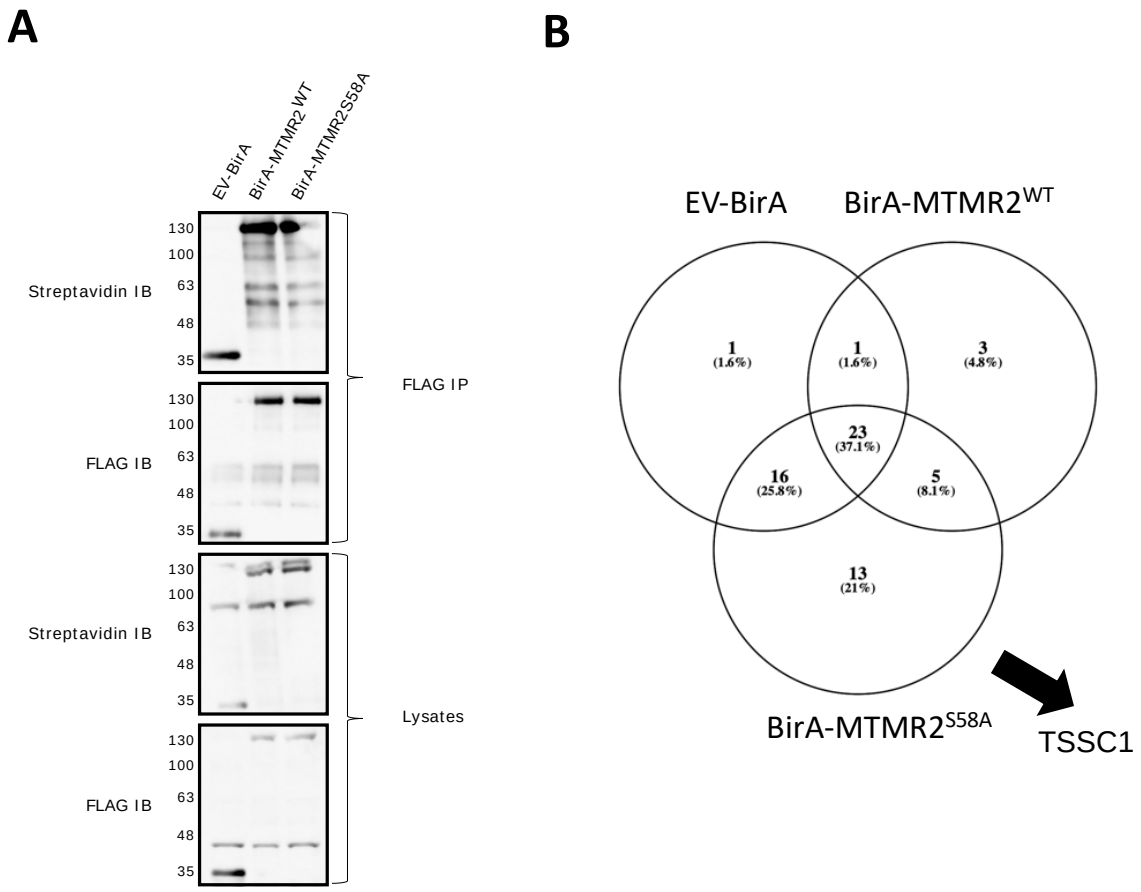


Figure 4.8. BioID protein identifications.

The interactome of MTMR2 at the endosome in Schwann cells was analyzed through large-scale expression of BirA-MTMR2 WT and SS58A followed by successive enrichments using anti-FLAG and streptavidin. Streptavidin-enriched fractions were then subjected to in-solution digestion. Tryptic peptides were submitted to LC-MS/MS analysis and processed using ProteinLynx Global Server. (**A**) Immunoblots of FLAG-precleared material and cell lysates, probed with anti-FLAG and streptavidin-HRP. (**B**) A Venn diagram of proteins identified in cells expressing EV-BirA, BirA-MTMR2^{WT} or BirA-MTMR2^{S58A} was created using Venny (https://bioinfogp.cnb.csic.es/tools/venny/). A list of identifications and their comparisons in samples can be seen in **Table 4.2**.

Table 4.2: MTMR2^{S58A} BioID protein identifications.

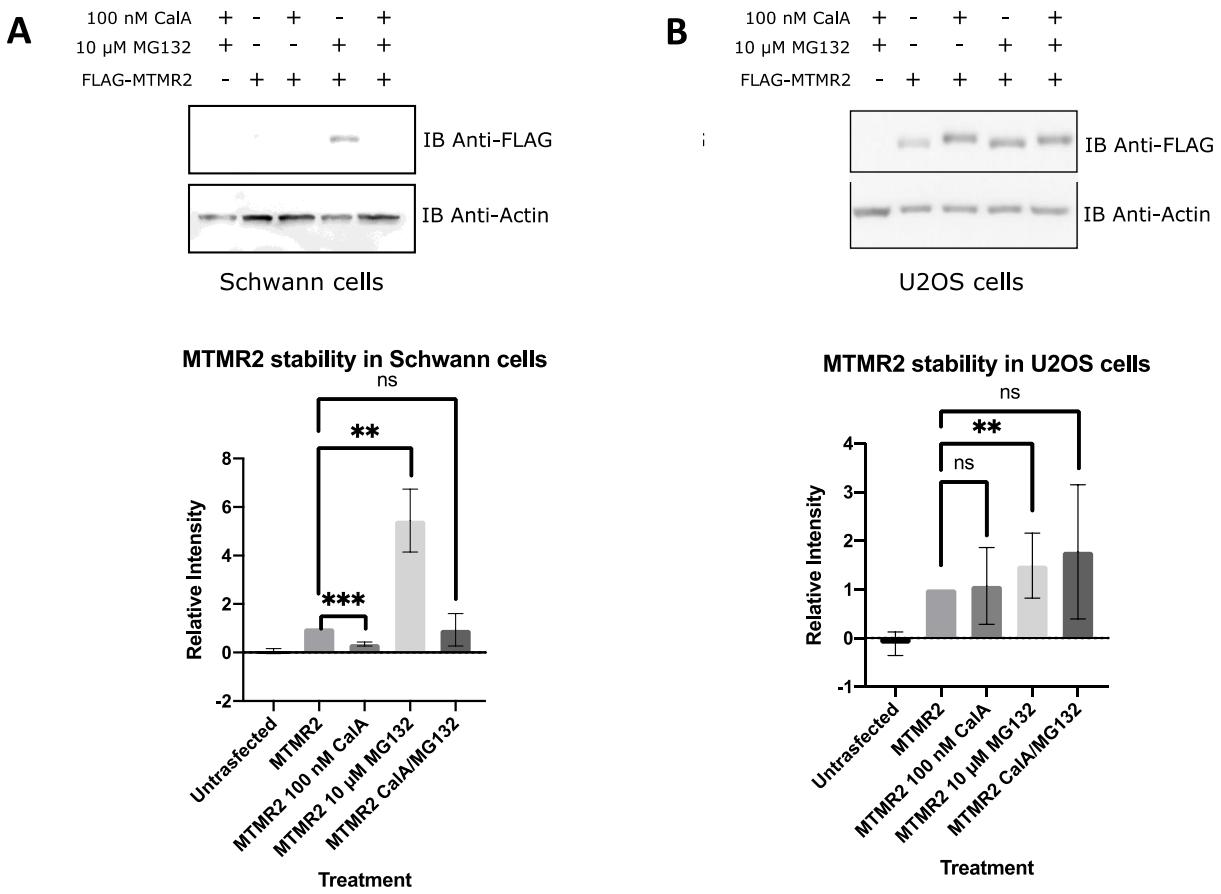
A full list of identifications can be seen in **Appendix C**.

Q62812.3	Myosin-9
Q64119.3	Myosin light polypeptide 6
P63259.1	Actin_ cytoplasmic 2
P60711.1	Actin_ cytoplasmic 1
P13832.2	Myosin regulatory light chain RLC-A
P18666.3	Myosin regulatory light chain 12B
Q64122.2	Myosin regulatory light polypeptide 9
P68035.1	Actin_ alpha cardiac muscle 1
P62738.1	Actin_ aortic smooth muscle
P63269.1	Actin_ gamma-enteric smooth muscle
P68136.1	Actin_ alpha skeletal muscle
P62864.1	40S ribosomal protein S30
Q9JLT0.1	Myosin-10
P04692.3	Tropomyosin alpha-1 chain
P58775.1	Tropomyosin beta chain
Q63862.3	Myosin-11
P62243.2	40S ribosomal protein S8
P09495.3	Tropomyosin alpha-4 chain
P83732.1	60S ribosomal protein L24
Q64598.3	Histone H2A type 1-F
P02262.2	Histone H2A type 1
P0C169.2	Histone H2A type 1-C
Q00728.2	Histone H2A type 4
Q4FZT6.3	Histone H2A type 3
A9UMV8.1	Histone H2A.J
P0C170.2	Histone H2A type 1-E
P0CC09.1	Histone H2A type 2-A
Q63610.2	Tropomyosin alpha-3 chain
P62752.1	60S ribosomal protein L23a
P16409.2	Myosin light chain 3
Q9R1T1.1	Barrier-to-autointegration factor
P31000.2	Vimentin

P15865.3	Histone H1.4
Q4KLF8.3	Actin-related protein 2/3 complex subunit 5
P05942.1	Protein S100-A4
Q4V7E8.1	Leucine-rich repeat flightless-interacting protein 2
D4A3K5.1	Histone H1.1
Q9WUL0.1	DNA topoisomerase 1
B2GUZ5.1	F-actin-capping protein subunit alpha-1
P62630.1	Elongation factor 1-alpha 1
Q3T1K5.1	F-actin-capping protein subunit alpha-2
Q62736.1	Non-muscle caldesmon
A1L1K4.2	CDC42 small effector protein 2
O88656.3	Actin-related protein 2/3 complex subunit 1B
P85970.1	Actin-related protein 2/3 complex subunit 2
B4F777.1	High mobility group nucleosome-binding domain-containing protein 5
P45592.3	Cofilin-1
P63018.1	Heat shock cognate 71 kDa protein
Q4V7C7.1	Actin-related protein 3
P41777.1	Nucleolar and coiled-body phosphoprotein 1
Q9QXQ0.2	Alpha-actinin-4
Q07266.3	Drebrin
Q9Z1P2.1	Alpha-actinin-1
P06349.2	Histone H1t
Q5RJR2.1	Twinfilin-1
Q5PPK9.1	**Protein TSSC1**
Q68FP1.1	Gelsolin

Degradation of MTMR2 in Schwann cells

While investigating reversible phosphorylation of MTMR2, we observed a decrease in MTMR2 protein levels in calyculin A- or okadaic acid-treated Schwann cells. As such, we investigated the potential route of phosphorylation-dependent degradation of MTMR2. We explored the possibility of ubiquitination facilitating degradation by expressing FLAG-MTMR2 in Schwann cells, followed by treatment with the proteasome MG132, with or without calyculin A. Cells were subsequently lysed and analyzed via immunoblotting using anti-ubiquitin. Even with MG132 treatment, no changes in the amount of ubiquitination could be detected nor was high molecular weight smearing in either cell line detected (**Figure 4.9C** and **D**). Upon analysis of lysates probed with anti-FLAG, we reproducibly observed a decrease in FLAG-MTMR2-expressing Schwann cells treated with calyculin A, including cells co-treated with the two inhibitors (**Figure 4.9A**, lanes 3 and 5). Interestingly, treatment using MG132 alone caused a large increase in the amount of FLAG-MTMR2 (**Figure 4.9**, lane 4). Calyculin A treatment did not reduce FLAG-MTMR2 levels in U2OS cells (**Figure 4.9B**), leading us to exploring proteasome-independent routes of MTMR2 degradation.


A
100 nM CalA + - + - +
10 μM MG132 + - - + +
FLAG-MTMR2 - + + + +
IB Anti-FLAG
IB Anti-Actin
Schwann cells
B
100 nM CalA + - + - +
10 μM MG132 + - - + +
FLAG-MTMR2 - + + + +
IB Anti-FLAG
IB Anti-Actin
U2OS cells
MTMR2 stability in Schwann cells
ns
**
***
Relative Intensity
8
6
4
2
0
-2
Untrasfected
MTMR2
MTMR2 100 nM CalA
MTMR2 10 μM MG132
MTMR2 CalA/MG132
Treatment
MTMR2 stability in U2OS cells
ns
**
ns
Relative Intensity
4
3
2
1
0
-1
Untrasfected
MTMR2
MTMR2 100 nM CalA
MTMR2 10 μM MG132
MTMR2 CalA/MG132
Treatment

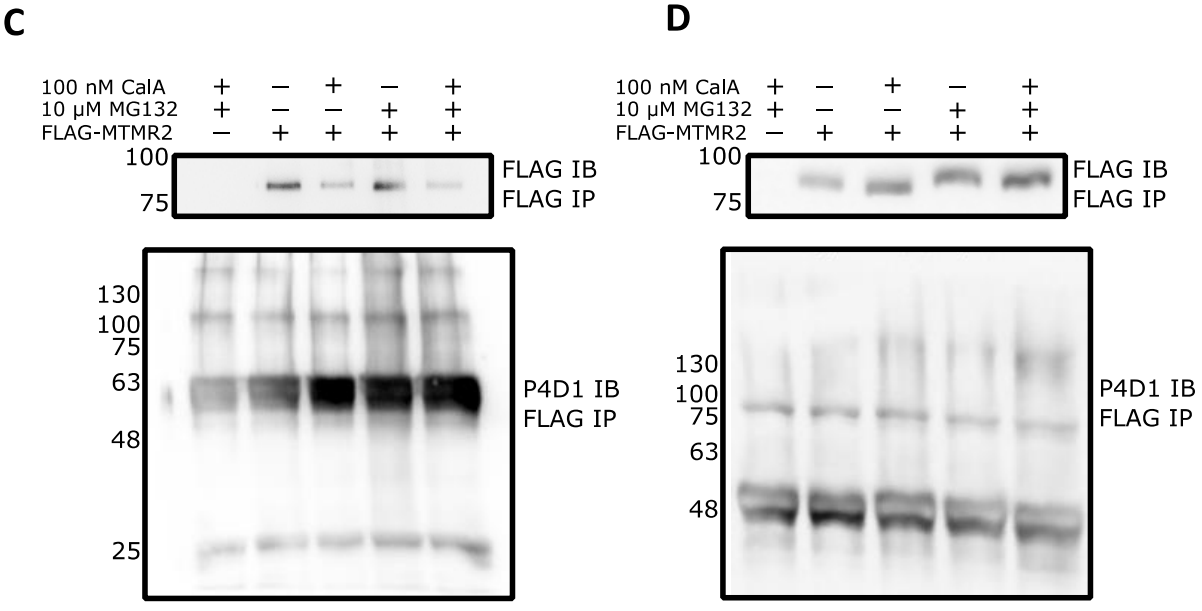

C
100 nM CalA + – + – +
10 μM MG132 + – – + +
FLAG-MTMR2 – + + + +
100
75
FLAG IB
FLAG IP
130
100
75
63
48
25
P4D1 IB
FLAG IP
D
100 nM CalA + – + – +
10 μM MG132 + – – + +
FLAG-MTMR2 – + + + +
100
75
FLAG IB
FLAG IP
130
100
75
63
48
P4D1 IB
FLAG IP

Figure 4.9. MTMR2 degradation in Schwann cells.

The concept of phosphorylation-dependent degradation of MTMR2 was examined in (**A**) Schwann and (**B**) U2OS cells. Cells overexpressing FLAG-MTMR2 were treated with either 5 nM CalA for 30 minutes, 10 μM MG132 for 6 hours, or both. Cells were then lysed, underwent FLAG immunoprecipitation, and analyzed via immunoblotting with anti-actin and anti-FLAG. Quantification of the effectiveness of the treatments are shown for Schwann cells and U2OS cells using a minimum of three independent transfections normalized to untreated FLAG-MTMR2^{WT}, plotting the mean ±S.D. ***P*<0.001, ****P*<0.0001 using an unpaired Student's *t* test. Ubiquitination of the immunoprecipitated proteins from (**C**) Schwann cells and (**D**) U2OS cells was evaluated using anti-ubiquitin (P4D1).

Lysosomal degradation and MTMR2

Proteasome inhibition did not rescue the levels of FLAG-MTMR2 upon phosphatase inhibitor treatment; therefore, we investigated the lysosomal-mediated degradation pathway for MTMR2 downregulation. Here, Schwann cells were treated with the autophagy/lysosome inhibitor chloroquine for 24 hours, calyculin A, or both (**Figure 4.10**). Cells treated with calyculin A, including those pretreated with chloroquine, still showed a reduction in FLAG-MTMR2 levels upon immunoblotting with anti-FLAG (**Figure 4.10**, lanes 3 and 5). Treatment with chloroquine alone or in combination with calyculin A did not increase FLAG-MTMR2 levels (**Figure 4.10**, lanes 4 and 5). These results are consistent with the MG132 treatments, suggesting that MTMR2 is not regulated by lysosomal degradation under the conditions utilized.

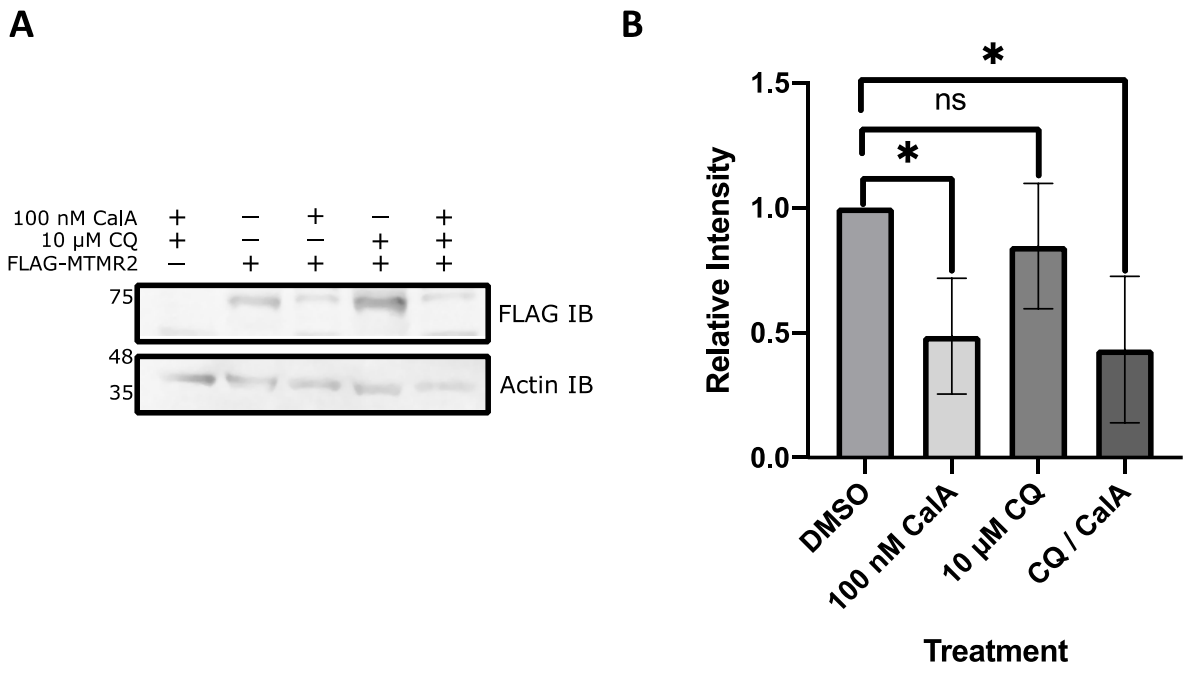


Figure 4.10. Assessing potential lysosomal degradation of MTMR2 in Schwann cells.

The lysosomal-mediated degradation of FLAG-MTMR2 following calyculin A treatment was tested in Schwann cells using the autophagy/lysosome inhibitor chloroquine (CQ). Cells overexpressing FLAG-MTMR2 were treated with 5 nM CalA for 30 mins, 10 µM CQ for 24 hours, or both.

(**A**) Lysates were probed with anti-FLAG and anti-actin.

(**B**) Statistics were performed using 3 independent transfections normalized to DMSO-treated FLAG-MTMR2^{WT} plotting the mean ± S.D, **P*<0.05, ns = not significant using an unpaired Student's *t* test.

Uncharacterized phosphorylation sites exist on MTMR2

The phosphatase inhibitor approach reproducibly showed that a phosphatase sensitive to calyculin A or okadaic acid has the potential to regulate MTMR2 serine/threonine phosphorylation, apparent through a mobility shift in a variety of cell lines. To determine if the mobility shift was caused by phosphorylation of a site that has been identified, we treated cells expressing FLAG-MTMR2 S58A and S631A with calyculin A, finding that they too displayed a mobility shift (**Figure 4.11A**). We also expressed the double phosphodeficient mutant S58A/S631A, and still observed a shift (**Figure 4.11**). The BioID constructs also exhibited this mobility shift, where BirA-MTMR2^{WT} treated with phosphatase inhibitor displayed a more retarded mobility compared to untreated cells, and BirA-MTMR2^{S58A} displayed a greater mobility than BirA-MTMR2^{S58E}, consistent with the idea of a phosphorylation-induced mobility shift (**Appendix C**).

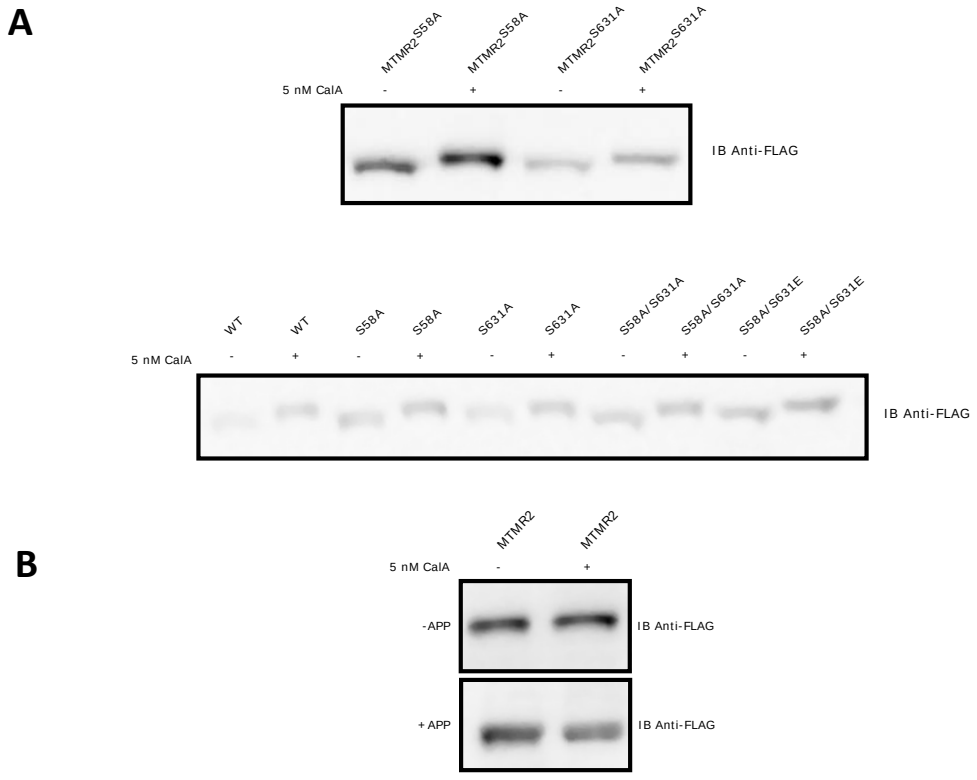


Figure 4.11. MTMR2 mobility shift is phosphorylation dependent.

The phosphorylation site responsible for the calyculin A (CalA)-induced mobility shift was examined by treating known MTMR2 phosphodeficient mutants with CalA. (**A**) $MTMR^{S58A}$ and $MTMR^{S631A}$ were expressed in HEK293 cells and treated with calyculin A. Following this, cells were transfected with double mutants (S58A/S631A or S58A/S631E) and treated with CalA or DMSO for 30 minutes. Lysates were analyzed via immunoblotting probed with anti-FLAG.

(**B**) The mobility of FLAG-MTMR2 treated with CalA followed by APP treatment was examined using immunoblotting.

DISCUSSION

MTMR2^{pSer58} is a critical regulator of MTMR2 localization, which is phosphorylated by ERK1/2,[35] localizing the protein away from its substrate on the endosome. A definitive link is yet to be established between ERK activation, MTMR2 phosphorylation, and the onset of CMT4B. One mechanism that has been investigated is the regulation of endocytosis, where dephosphorylated MTMR2 is able to deplete PI3P from early endosomes, leading to an increase in ERK1/2 activation.[35,34] As such, it is critical to elucidate the mechanisms and proteins responsible for MTMR2^{pSer58} dephosphorylation. We employed two complementary approaches to address these issues: first, we used phosphatase inhibitors in an attempt to observe an increase in MTMR2^{pSer58} levels, and second, we attempted to enrich for the phosphatase in question. In the cell, an overwhelming majority of phosphorylation occurs on serine residues, with over 400 serine/threonine kinases encoded in the human proteome.[46] Only ~30 serine/threonine phosphatases exist in the proteome, which gain substrate specificity through association with regulatory subunits.[46] We attempted to inhibit the activity of a variety of members of the serine/threonine phosphatase family using a panel of inhibitors. No dramatic increase in MTMR2^{pSer58} levels were observed upon treatment with the inhibitors, although we reproducibly observed a lack of MTMR2 mobility in cells treated with calyculin A or okadaic acid (**Figure 4.3**). These two inhibitors act on overlapping substrates with different affinity, with the major targets being PP2A and PP1.[47] Mobility shifts are indicative of increases in protein phosphorylation, where proteins in their phosphorylated forms exhibit a slower mobility compared to its unphosphorylated form.[48] The MTMR2 phosphomimetic mutants MTMR2^{S58E} displays this shift in comparison to untreated MTMR2, leading us to

conclude that these treatments influence MTMR2 phosphorylation, but not at Ser58, as pSer58 levels do not elevate upon treatment as shown by immunoblotting (**Figure 4.3**). This was also supported by the observations that treatment of cells expressing the MTMR2^{S58A} mutant with calyculin A also produced a mobility shift. This led us to examine the other known phosphorylation site of MTMR2, Ser^{631}, regulating the early endosome subtypes with which MTMR2 localizes.[35] However, treating cells expressing MTMR2^{S631A} and MTMR2$^{S58A/S631A}$ with calyculin A still resulted in a mobility shift as compared to DMSO-treated controls (**Figure 4.11**). Unfortunately, no MTMR2^{S631} phosphoantibody is currently available, further hindering our ability to determine whether this site is affected by our phosphatase inhibitor treatment. This leads us to believe an unmapped phosphorylation site of MTMR2 exists, which is regulated by a calyculin A/okadaic acid-sensitive phosphatase or signaling pathway. Further evidence suggesting this mobility shift is phosphorylation-dependent is observed when performing *in vitro* phosphatase reactions on lysates pretreated with calyculin A. Prior to APP treatment, calyculin A-treated lysates display a slowed mobility as compared to untreated (**Figure 4.11A**). Following APP treatment, calyculin A-treated lysates migrate the same distance as MTMR2 from untreated cells (**Figure 4.11B**). How this uncharacterized phosphorylation site is regulated requires further exploration, potentially influencing MTMR2 stability as discussed below.

We transitioned to using more disease-relevant cell lines, employing the myelinating cells of the PNS, Schwann cells. When translating our phosphatase inhibitor approach to these cells we observed the MTMR2 mobility shift seen upon treatment with calyculin A and okadaic acid consistent with HEK293 cells. An interesting phenomenon we observed was a decrease in FLAG-MTMR2 levels compared to DMSO control (**Figure**

4.9A). Phosphorylation-dependent degradation of proteins is a known mechanism, where the phosphate group(s) acts as a signal for degradative enzymes to recognize and bind the phosphorylated protein. The ubiquitin-proteasome system (UPS) is known to regulate proteins in this manner.[49] Here, substrate lysine residues are tagged with ubiquitin, allowing them to be recognized, ultimately dictating the cellular fate of the tagged proteins. One pathway synonymous with ubiquitination is degradation via the 26S proteasome, one pathway using the F-box protein β-TrCP to recognize phosphorylation-dependent destruction motifs on proteins, leading to their modification with ubiquitin and subsequent 26S proteasomal degradation.[50] To assess whether MTMR2 is subject to proteasome-mediated degradation, we expressed MTMR2 in Schwann cells, treated them with calyculin A, and pretreated cells with the proteasome inhibitor MG132 (carbobenzoxy-Leu-Leu-Leucinal) (**Figure 4.9**). MG132 treatment inhibits the proteolytic activity of the proteasome, allowing for accumulation of ubiquitinated proteins that would typically be degraded, often exhibiting as a smear on immunoblots. Co-treatment with calyculin A and MG132 did not rescue MTMR2 levels compared to DMSO or MG132-only treated controls, nor was an increase in ubiquitination observed upon treatment (**Figures 4.9A** and **B**, respectively). This led us to explore ubiquitin-independent routes of degradation, one being autophagy. Autophagy utilizes double-membrane vesicles called autophagosomes to engulf molecules, including organelles and proteins, and shuttle them to the lysosome for degradation in both a ubiquitin-dependent and independent path.[51] Components of the autophagosome are related to the endosomal system, such as PI3K, further enticing us to explore the concept of proteasome-independent degradative mechanism of MTMR2 regulation, by using the autophagy/lysosomal inhibitor chloroquine.[52] Chloroquine

treatment did not result in an increase in FLAG-MTMR2 levels as was observed MG132 treatment (**Figure 4.10**). A question that remains to be answered is whether calyculin A induces autophagy or lysosomal-mediated degradation. We attempted to address this by monitoring endogenous lysosome-associated membrane protein 1 (LAMP1) levels, a protein found on autophagosomes, but obtained inconsistent results (data not shown). Proteasome or lysosome inhibition did not rescue MTMR2 levels, suggesting MTMR2 may be regulated through one of the other ubiquitin-independent pathways of degradation. A route that has been explored in other neurological diseases is degradation through calpains,[53] which recognize PEST sequences or regions rich in Pro, Glu, Ser, and Thr residues within proteins,[54] as seen with the PI3P phosphatase PTEN and the PI(3,4)P2 phosphatase inositol polyphosphate 4-phosphatase.[55,56] Using EMBOSS explorer, MTMR2 possesses a putative PEST domain sequence between positions 18 and 32 (http://emboss.bioinformatics.nl), making this an intriguing route of degradation to examine. Our findings suggest that MTMR2 does not experience phosphorylation-dependent degradation but may be regulated by a calyculin A-specific function, possibly influencing the levels of other proteins that regulate MTMR2. Proteins that may be regulated in this fashion are the catalytically inactive interacting partners, MTMR5 and MTMR13, whose levels are dependent on MTMR2 abundance and vice versa.[57,58] It is possible that calyculin A treatment induces mechanisms—being proteasome or lysosome independent—that allows for MTMR2 to be downregulated with MTMR5/13.

While challenged to determine the routes of MTMR2 degradation we found MG132 treatment alone lead to a large increase in FLAG-MTMR2 levels, but we could not detect any increases in ubiquitin levels. The calyculin A-induced reduction of MTMR2 and the

MG132-induced increase in FLAG-MTMR2 levels appear to be a Schwann cell specific phenomenon. Alternatively, there is potential these treatments display rat cell-specific effects, where the human cell lines used, U2OS and HEK293, did not display those similar effects following treatment (**Figure 4.9B**). To decipher why this increase occurs will require analysis of multiple pathways and proteins therein, as these inhibitors can influence a wide variety of substrates. To date, the most widely affected calyculin A pathways are involved with calcium influx and MAPK signaling (where multiple steps are inhibited by activity of PP2A), whereas MG132 treatment has also been shown to activate JNK in neuronal cells.[59]

From these findings our working model is that MTMR2 is synthesized, localizes to endosomes where it dephosphorylates its substrate. MTMR2 is subsequently phosphorylated by ERK1/2, causing it to localize away from the endosome. Following this, MTMR2 is degraded via an unresolved route.

A second approach used to identify the phosphatase—as well as other proteins with potential to be critical in the endosomal program—utilized traditional AP-MS strategies with enrichment of FLAG-MTMR2 using FLAG-IP followed by MS analysis. Using this approach, no serine phosphatases unique to MTMR2 constructs compared to control were identified (**Figure 4.4**). Some proteins identified consistent with MTMR2 being involved in endocytosis are the GTPases Rab1 and Rab7, t-SNARE interacting protein, a PI4P 5-kinase (**Table 4.1**). The abundant protein phosphatase 1B was identified in control and FLAG-MTMR2-expressing cells, providing confidence in the feasibility of isolating a phosphatase as well as capturing other potential transient interactions in the cell. To gain a better understanding of the interactome of MTMR2 at the endosome, as well as identify

the Ser[58] phosphatase, we attempted to label the proteins in question *in vivo* using proximity labeling, theoretically allowing us to identify the direct and indirect interactors of MTMR2. This led to the production of BioID vectors for the expression of MTMR2 and its variants as fusions to the promiscuous biotin ligase, BirA*.[60] Following optimization of biotin-labeling (**Figure 4.5**) and assessment of proper biological functioning of the constructs through immunofluorescence assays (IFA) (**Figure 4.6**), BioID was performed. We used IFA to examine the degree of colocalization between our fusion constructs and the physiological substrate of MTMR2, PI3P, by using the GTPase Rab5 as a marker of EEs.[61] We found strong colocalization between BirA-MTMR2^{S58A} with Rab5. Meanwhile, EV and S58E qualitatively showed little to no colocalization with Rab5-positive endosomes. These results are consistent with the localization patterns observed for past MTMR2 Ser58 phosphomutants.[35] A large amount of MTMR2-staining was often observed, likely due to overexpression; to avoid this, future experiments can employ detergents to allow for non-membrane-bound leakage of material, such as with the use of saponins,[45] allowing for a clearer signal to be observed. The IFA results provided confidence in the proper localization of the BirA* fusion constructs, suggesting that the fusion constructs retained their biological function, making them feasible in examining the MTMR2 interactome.

BioID has provided interactome data for a variety of proteins in different organisms, by allowing the robust labeling of proteins proximal to a bait and subsequent enrichment and MS identification.[60,62,63] We expressed these BioID vectors in Schwann cells, the myelinating cells of the PNS, as these are the cells affected by MTMR2 mutations causing CMT4B.[2] These experiments allowed for identifications consistent with expected results, such as the identification of MTMR2 and protein S100, a factor important for

Schwann cell function (**Table 4.2**).[64] Other proteins typical of pulldown experiments (as reported in the contaminant repository for affinity purification mass spectrometry data such as actin, keratin, and ribosome components) were detected.[65] Gene ontology analysis of findings generated terms consistent with Schwann cells and MTMR2-specific functions, such as myelination and intracellular transport (**Figure 4.4C**). Comparison of the proteins identified from EV-BirA, BirA-MTMR2^{WT} and BirA-MTMR2^{S58A}, 38 proteins were found from in all, with 11 unique proteins being found in the S58A background. An interesting identification unique to BirA-MTMR2^{S58A} was the protein TSSC1 also known as EIPR1. This finding displays the importance of MTMR2 regulating PI3P-dependent proteins functions in the endosomal system as shown for other proteins.[4,19,45] TSSC1 was originally thought to be involved in tumour suppression based on its DNA sequence; however, further studies revealed its role in endosomal trafficking, where it interacts with the endosome-associated recycling protein (EARP), being referred to as EARP-interacting protein-1, EIPR-1.[66,67] EIPR associates with EARP and the closely related Golgi-associated retrograde protein (GARP) complex, recruiting them to transport cargo to recycling endosomes or the trans-Golgi network, respectively.[66] These tethering complexes are important for their respective roles in endocytic trafficking and regulated by interactions with EIPR-1. Silencing of EIPR-1 shows impairment of Shiga toxin B subunit and internalized transferrin receptor transport to the TGN and plasma membrane, respectively. For EIPR-1 to perform its role, we believe it uses one of its four predicted WD-40 domains to bind onto PI3P-containing endosomes,[67] like the WD-repeat protein interacting with phosphoinositides.[68] As MTMR2 is a PI3P phosphatase, we propose MTMR2 localizes to PI3P-rich endosomes where EIPR-1 is functioning associated with EARP or GARP,

allowing for the labeling and enrichment of EIPR-1, with MTMR2 serving as a temporal regulator of EIPR-1 through dephosphorylation of PI3P. To our knowledge, the mechanisms regulating EIPR-1 endosomal localization remains unknown. Studies of dense core vesicles suggest that EIPR-1 functions in the same pathway as Rab-2;[67] this, taken with its activity as a tethering complex, leads us to postulate that EIPR-1 regulation is similar to RME-8 and its dependence on MTMR2 regulating PI3P levels and is involved in regulated retrograde transport.[45] We have struggled to successfully express EIPR-1 in cells but have attempted to measure its interaction with MTMR2 via co-IP and IFA. The validation and understanding of the MTMR2/EIPR-1 interaction will provide insight into the consequence of MTMR2 in the cell.

While we were successful in identifying a novel protein interactor, our methods could be improved using catalytically inactive BioID vectors. These catalytically inactive mutants will act as substrate traps potentially allowing MTMR2 to remain at the endosome longer, allowing for labeling of proximal proteins. Tracking the *in vivo* labeling using IFA probing for biotin would complement this approach and highlight the areas to which MTMR2 interactors localize, which could be correlated with their localization patterns in CMT4B to further explain MTMR2 loss and disease phenotype. The interactome of MTMR2 can be deepened by also creating MTMR13 and MTMR5 BioID constructs, allowing for labeling of proteins that come into proximity to the heterocomplexes formed. This would provide a deeper interactome, allowing for labeling of more indirect interactors of MTMR2, helping in the understanding of the MTMR2 interactome landscape. We attempted to decipher the mechanisms behind MTMR2 degradation upon calyculin A treatment by coupling this inhibitor treatment to our BioID approach. Unfortunately, only

one replicate was performed which provided no unique findings, although a promising observation was the downregulation of BirA-MTMR2^{WT} upon calyculin A treatment, consistent with what was observed with our FLAG-MTMR2 expression studies (**Appendix C**). A unique step in our workflow is the use of a FLAG-preclear, allowing for the removal of FLAG-BirA-MTMR2 and its direct/indirect interacting proteins prior to biotin enrichment. This, theoretically, aids in biotin enrichment by reducing both sample complexity and interference from the highly abundant bait proteins, which is also strongly observed in immunoblots with streptavidin, validating the biotinylation of the bait (**Figure 4.7B**, left). Based on our immunoblotting, FLAG-precleared material did not contain many non-bait biotinylated proteins (**Figure 4.7B**, right), propelling us to measure only streptavidin-enriched material. To gain a better understanding of the stronger or more proximal interactions occurring on MTMR2, FLAG-precleared and streptavidin enrichments can both be analyzed. Comparing our traditional AP-MS datasets and those obtained from BioID studies does show different protein enrichments. This is somewhat expected as our FLAG-IPs theoretically enrich for more direct-interacting proteins, whereas the proximity labeling can theoretically tag any proteins possessing accessible primary amines within a 10-nm radius.[69] Furthermore, the use of biotinylation allows for the labeling and capture of transient interactions that may be too weak to survive the enrichment methods of standard IPs, such as the detergent-containing washing steps.

A flaw in our current approach is that we do not supplement the cells with 50 µM biotin until 6 hours following transfection, leaving time where the sole source of biotin is from serum, and in relatively low amounts.[70] Roux et al, report that biotinylation occurs maximally at 6–24 hours, which may only allow us to see maximal biotinylation in a 12-

hour window, possibly allowing potential interactors to go undetected. To overcome this challenge, we prepared a large-scale experiment where biotin was supplemented in the transfection media, allowing for biotinylation throughout the course of the transfection. Unfortunately, the peptides generated from this study have not been measured but we are hopeful that these adjustments will aid in labeling and identification of proteins, allowing us to make further inferences of MTMR2's role in cellular function.

CONCLUSION

The myotubularin-related protein 2 is an important regulator of the endosomal system and critical in Charcot-Marie-Tooth disease type 4B1. To gain a better understanding of MTMR2 and its endosomal regulation we investigated MTMR2 protein-protein interactions leading us to generate MTMR2-BioID vectors, which allowed for the labeling of proximal proteins. We identified a novel substrate, EIPR-1 (also known as TSSC1), which is an important regulator in the recycling of cargo to the TGN. The previously identified phosphorylation site of MTMR2, Ser58, is critical for the proteins' regulation. As such, we attempted to identify this phosphatase in question. Our phosphatase inhibitor treatment suggests the existence of an uncharacterized MTMR2 phosphorylation site, whose assignment and function needs to be determined. Our work suggests that MTMR2 may not be subjected to dephosphorylation, but rather be degraded via proteasome-independent mechanisms. By combining high-resolution mass spectrometry with molecular biology techniques, we have created tools to explore the MTMR2 interactome with potential to deepen the understanding of the protein in disease and its biological regulation.

REFERENCES

1. Pareyson, D. *et al.* A multicenter retrospective study of charcot-marie-tooth disease type 4B (CMT4B) associated with mutations in myotubularin-related proteins (MTMRs). *Ann. Neurol.* **86**, 55–67 (2019).
2. Bolis, A. *et al.* Loss of Mtmr2 phosphatase in Schwann cells but not in motor neurons causes Charcot-Marie-Tooth type 4B1 neuropathy with myelin outfoldings. *J. Neurosci.* **25**, 8567–8577 (2005).
3. Bolino, A., Muglia, M., Conforti, F.L., LeGuern, E., Salih, M.A., Georgiou, D.M., Christodoulou, K., Hausmanowa-Petrusewicz, I., Mandich, P., Schenone, A. and Gambardella, A. Charcot-Marie-Tooth type 4B is caused by mutations in the gene encoding myotubularin-related protein-2. *Nat. Genet.* **25**, 17–19 (2000).
4. Bolino, A. *et al.* Disruption of Mtmr2 CMT4B1-like neuropathy with myelin outfolding and impaired spermatogenesis. *J. Cell Biol.* **167**, 711–721 (2004).
5. Nouioua, S. *et al.* Novel mutations in the PRX and the MTMR2 genes are responsible for unusual Charcot-Marie-Tooth disease phenotypes. *Neuromuscul. Disord.* **21**, 543–550 (2011).
6. Senderek, J. *et al.* Mutation of the SBF2 gene, encoding a novel member of the myotubularin family, in Charcot-Marie-Tooth neuropathy tye 4B2/11p15. *Hum. Mol. Genet.* **12**, 349–356 (2003).
7. Robinson, D. C. *et al.* An In Vitro Model of Charcot-Marie-Tooth Disease Type 4B2 Provides Insight Into the Roles of MTMR13 and MTMR2 in Schwann Cell Myelination. *ASN Neuro* **10**, (2018).
8. Azzedine, H. *et al.* Mutations in MTMR13, a new pseudophosphatase homologue of MTMR2 and Sbf1, in two families with an autosomal recessive demyelinating form of Charcot-Marie-Tooth disease associated with early-onset glaucoma. *Am. J. Hum. Genet.* **72**, 1141–1153 (2003).
9. Kim, S. A., Vacratsis, P. O., Firesteint, R., Clearyt, M. L. & Dixon, J. E. Regulation of myotubularin-related (MTMR)2 phosphatidylinositol phosphatase by MTMR5, a catalytically inactive phosphatase. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 4492–4497 (2003).
10. Robinson, F. L. & Dixon, J. E. The phosphoinositide-3-phosphatase MTMR2 associates with MTMR13, a membrane-associated pseudophosphatase also mutated in type 4B Charcot-Marie-Tooth disease. *J. Biol. Chem.* **280**, 31699–31707 (2005).
11. Ben Othmane, K. *et al.* Identification of a new locus for autosomal recessive Charcot-Marie- Tooth disease with focally folded myelin on chromosome 11p15. *Genomics* **62**, 344–349 (1999).
12. Nakhro, K. *et al.* SET binding factor 1 (SBF1) mutation causes Charcot-Marie-Tooth disease type 4B3. *Neurology* **81**, 165–173 (2013).
13. Anas M Alazami 1 , Fatema Alzahrani, Saeed Bohlega, F. S. A. SET binding factor 1 (SBF1) mutation causes Charcot-Marie-Tooth disease type 4B3 (Editor's choice).

Neurology **82**, 1665–1666 (2014).

14. Bohlega, S. *et al.* A novel syndromic form of sensory-motor polyneuropathy is linked to chromosome 22q13.31-q13.33. *Clin. Genet.* **79**, 193–195 (2011).

15. Mammel, A.E., Delgado, K.C., Chin, A.L., Condon, A.F., Hill, J.Q., Aicher, S.A., Wang, Y., Fedorov, L.M. and Robinson, F.L. Distinct roles for the Charcot-Marie-Tooth disease-causing endosomal regulators Mtmr5 and Mtmr13 in axon radial sorting and Schwann cell myelination. *bioRxiv* **843219**, (2019).

16. Berger, P., Schaffitzel, C., Berger, I., Ban, N. & Suter, U. Membrane association of myotubularin-related protein 2 is mediated by a pleckstrin homology-GRAM domain and a coiled-coil dimerization module. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 12177–12182 (2003).

17. Begley, M. J. *et al.* Molecular basis for substrate recognition by MTMR2, a myotubularin family phosphoinositide phosphatase. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 927–932 (2006).

18. Begley, M. J. *et al.* Crystal Structure of a Phosphoinositide Phosphatase, MTMR2: Insights into Myotubular Myopathy and Charcot-Marie-Tooth Syndrome. *Mol. Cell* **12**, 1391–1402 (2003).

19. Lee, H. W., Kim, Y., Han, K., Kim, H. & Kim, E. The phosphoinositide 3-phosphatase MTMR2 interacts with PSD-95 and maintains excitatory synapses by modulating endosomal traffic. *J. Neurosci.* **30**, 5508–5518 (2010).

20. Aung, K. T. *et al.* The class II phosphoinositide 3-kinases PI3K-C2α and PI3K-C2β differentially regulate clathrin-dependent pinocytosis in human vascular endothelial cells. *J. Physiol. Sci.* **69**, 263–280 (2019).

21. Arumugam, S. & Kaur, A. The Lipids of the Early Endosomes: Making Multimodality Work. *ChemBioChem* **18**, 1053–1060 (2017).

22. Holroyd, C., Kistner, U., Annaert, W. & Jahn, R. Fusion of endosomes involved in synaptic vesicle recycling. *Mol. Biol. Cell* **10**, 3035–3044 (1999).

23. Simonsen, A., Lippe, R., Christoforidis, S., Gaullier, J.M., Brech, A., Callaghan, J., Toh, B.H., Murphy, C., Zerial, M. and Stenmark, H. EEA1 links PI(3)K function to Rab5 regulation of endosome fusion. *Nature* **394**, 492–498 (1998).

24. Wallroth, A. & Haucke, V. Phosphoinositide conversion in endocytosis and the endolysosomal system. *J. Biol. Chem.* **293**, 1526–1535 (2018).

25. Johnstone, R. M., Adam, M., Hammond, J. R., Orr, L. & Turbide, C. Vesicle formation during reticulocyte maturation. Association of plasma membrane activities with released vesicles (exosomes). *J. Biol. Chem.* **262**, 9412–9420 (1987).

26. Marsh, E. W., Leopold, P. L., Jones, N. L. & Maxfield, F. R. Oligomerized transferrin receptors are selectively retained by a lumenal sorting signal in a long-lived endocytic recycling compartment. *J. Cell Biol.* **129**, 1509–1522 (1995).

27. Meister, M. *et al.* Regulation of cargo transfer between ESCRT-0 and ESCRT-I complexes by flotillin-1 during endosomal sorting of ubiquitinated cargo.

Oncogenesis **6**, (2017).

28. Anne M. Logan, Anna E. Mammel, Danielle C. Robinson, Andrea L. Chin, Alec F. Condon, and F. L. R. Schwann cell-specific deletion of the endosomal PI 3-kinase Vps34 leads to delayed radial sorting of axons, arrested myelination, and abnormal ErbB2-ErbB3 tyrosine kinase signaling. *Glia* **65**, 1452–1470 (2017).

29. Ishii, A., Furusho, M., Dupree, J. L. & Bansal, R. Strength of ERK1/2 MAPK activation determines its effect on myelin and axonal integrity in the adult CNS. *J. Neurosci.* **36**, 6471–6487 (2016).

30. Ogata, T. *et al.* Opposing extracellular signal-regulated kinase and Akt pathways control schwann cell myelination. *J. Neurosci.* **24**, 6724–6732 (2004).

31. Jang, S. Y. *et al.* Schwann cell dedifferentiation-associated demyelination leads to exocytotic myelin clearance in inflammatory segmental demyelination. *Glia* **65**, 1848–1862 (2017).

32. Napoli, I. *et al.* A Central Role for the ERK-Signaling Pathway in Controlling Schwann Cell Plasticity and Peripheral Nerve Regeneration In Vivo. *Neuron* **73**, 729–742 (2012).

33. Harrisingh, M. C. *et al.* The Ras/Raf/ERK signalling pathway drives Schwann cell dedifferentiation. *EMBO J.* **23**, 3061–3071 (2004).

34. Franklin, N. E., Taylor, G. S. & Vacratsis, P. O. Endosomal targeting of the phosphoinositide 3-phosphatase MTMR2 is regulated by an N-terminal phosphorylation site. *J. Biol. Chem.* **286**, 15841–15853 (2011).

35. Franklin, N. E., Bonham, C. A., Xhabija, B. & Vacratsis, P. O. Differential phosphorylation of the phosphoinositide 3-phosphatase MTMR2 regulates its association with early endosomal subtypes. *J. Cell Sci.* **126**, 1333–1344 (2013).

36. Cervellini, I. *et al.* Sustained MAPK/ERK activation in adult schwann cells impairs nerve repair. *J. Neurosci.* **38**, 679–690 (2018).

37. Mori, S. *et al.* Crystal structure of the guanylate kinase domain from discs large homolog 1 (DLG1/SAP97). *Biochem. Biophys. Res. Commun.* **435**, 334–338 (2013).

38. Bolis, A. *et al.* Dlg1, Sec8, and Mtmr2 regulate membrane homeostasis in Schwann cell myelination. *J. Neurosci.* **29**, 8858–8870 (2009).

39. Previtali, S. C. *et al.* Myotubularin-related 2 protein phosphatase and neurofilament light chain protein, both mutated in CMT neuropathies, interact in peripheral nerve. *Hum. Mol. Genet.* **12**, 1713–1723 (2003).

40. Goryunov, D., Nightingale, A., Bornfleth, L., Leung, C. & Liem, R. K. H. Multiple disease-linked myotubularin mutations cause NFL assembly defects in cultured cells and disrupt myotubularin dimerization. *J. Neurochem.* **104**, 1536–1552 (2008).

41. Cao, C., Backer, J.M., Laporte, J., Bedrick, E.J. and Wandinger-Ness. Sequential Actions of Myotubularin Lipid Phosphatases Regulate Endosomal PI(3)P and Growth Factor Receptor Trafficking. *Mol. Biol. Cell* **19**, 3334–3346 (2008).

42. Kim, S. A., Taylor, G. S., Torgersen, K. M. & Dixon, J. E. Myotubularin and MTMR2, phosphatidylinositol 3-phosphatases mutated in myotubular myopathy and type 4B Charcot-Marie-Tooth disease. *J. Biol. Chem.* **277**, 4526–4531 (2002).

43. Taylor, G. S., Maehama, T. & Dixon, J. E. Myotubularin, a protein tyrosine phosphatase mutated in myotubular myopathy, dephosphorylates the lipid second messenger, phosphatidylinositol 3-phosphate. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 8910–8915 (2000).

44. Gillooly, D. J. *et al.* Localization of phosphatidylinositol 3-phosphate in yeast and mammalian cells David. **19**, 4577–4588 (2000).

45. Xhabija, B., Taylor, G. S., Fujibayashi, A., Sekiguchi, K. & Vacratsis, P. O. Receptor mediated endocytosis 8 is a novel PI(3)P binding protein regulated by myotubularin-related 2. *FEBS Lett.* **585**, 1722–1728 (2011).

46. Shi, Y. Serine/Threonine Phosphatases: Mechanism through Structure. *Cell* **139**, 468–484 (2009).

47. Clark, A. R. & Ohlmeyer, M. Protein phosphatase 2A as a therapeutic target in inflammation and neurodegeneration. *Pharmacol. Ther.* **201**, 181–201 (2019).

48. Lee, C., Park, Y., Min, H. & Kim, Y. Determination of protein phosphorylation by polyacrylamide gel electrophoresis. **57**, 93–100 (2019).

49. Petroski, M. D. & Deshaies, R. J. Function and regulation of cullin-RING ubiquitin ligases. *Nat. Rev. Mol. Cell Biol.* **6**, 9–20 (2005).

50. Nana Zhenga, Zhiwei Wanga, and Wenyi Weib. Ubiquitination-mediated degradation of cell cycle-related proteins by F-box proteins. *Int. J. Biochem. Cell Biol.* **73**, 99–110 (2016).

51. Emmerich, C. H. & Cohen, P. Optimising methods for the preservation, capture and identification of ubiquitin chains and ubiquitylated proteins by immunoblotting. *Biochem. Biophys. Res. Commun.* **466**, 1–14 (2015).

52. Mauthe, M. *et al.* Chloroquine inhibits autophagic flux by decreasing autophagosome-lysosome fusion. *Autophagy* **14**, 1435–1455 (2018).

53. Schaecher, K. E., Shields, D. C. & Banik, N. L. Mechanism of myelin breakdown in experimental demyelination: A putative role for calpain. *Neurochem. Res.* **26**, 731–737 (2001).

54. Enomoto, A. *et al.* Prevention of calpain-dependent degradation of STK38 by MEKK2-mediated phosphorylation. *Sci. Rep.* **9**, 1–10 (2019).

55. Briz, V. *et al.* Calpain-2-mediated PTEN degradation contributes to BDNF-induced stimulation of dendritic protein synthesis. *J. Neurosci.* **33**, 4317–4328 (2013).

56. Norris, F. A., Atkins, R. C. & Majerus, P. W. Inositol polyphosphate 4-phosphatase is inactivated by calpain-mediated proteolysis in stimulated human platelets. *J. Biol. Chem.* **272**, 10987–10989 (1997).

57. Ng, A. A., Logan, A. M., Schmidt, E. J. & Robinson, F. L. The CMT4B disease-causing phosphatases Mtmr2 and Mtmr13 localize to the Schwann cell cytoplasm

and endomembrane compartments, where they depend upon each other to achieve wild-type levels of protein expression. *Hum. Mol. Genet.* **22**, 1493–1506 (2013).

58. Robinson, D. C. *et al.* An In Vitro Model of Charcot-Marie-Tooth Disease Type 4B2 Provides Insight Into the Roles of MTMR13 and MTMR2 in Schwann Cell Myelination. *ASN Neuro* **10**, (2018).

59. Zhang, L. *et al.* Proteasome Inhibition Modulates Kinase Activation in. **87**, 3231–3238 (2010).

60. Roux, K. J., Kim, D. I., Raida, M. & Burke, B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. **196**, 801–810 (2012).

61. Nagano, M., Toshima, J. Y., Elisabeth Siekhaus, D. & Toshima, J. Rab5-mediated endosome formation is regulated at the trans-Golgi network. *Commun. Biol.* **2**, 1–12 (2019).

62. Khan, M., Youn, J. Y., Gingras, A. C., Subramaniam, R. & Desveaux, D. In planta proximity dependent biotin identification (BioID). *Sci. Rep.* **8**, 2–9 (2018).

63. Chen, A. L. *et al.* Novel components of the toxoplasma inner membrane complex revealed by BioID. *MBio* **6**, 1–12 (2015).

64. Liu, Z. *et al.* Specific marker expression and cell state of Schwann cells during culture in vitro. *PLoS One* **10**, 1–17 (2015).

65. Mellacheruvu, D. *et al.* The CRAPome: a Contaminant Repository for Affinity Purification Mass Spectrometry Data. *Nat. Methods* **10**, 730–736 (2013).

66. Gershlick, D. C., Schindler, C., Chen, Y., Bonifacino, J. S. & Gruenberg, J. E. TSSC1 is novel component of the endosomal retrieval machinery. *MBoC* **27**, 2867–2878 (2016).

67. Topalidou, I. *et al.* The EARP Complex and Its Interactor EIPR-1 Are Required for Cargo Sorting to Dense-Core Vesicles. *PLoS Genet.* **12**, 1–29 (2016).

68. Proikas-Cezanne, T. *et al.* WIPI-1α (WIPI49), a member of the novel 7-bladed WIPI protein family, is aberrantly expressed in human cancer and is linked to starvation-induced autophagy. *Oncogene* **23**, 9314–9325 (2004).

69. Kim, D. I. *et al.* Probing nuclear pore complex architecture with proximity-dependent biotinylation. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 2453–2461 (2014).

70. Collins, J. C., Paietta, E., Green, R., Morell, A. G. & Stockert, R. J. Biotin-dependent expression of the asialoglycoprotein receptor in HepG2. *J. Biol. Chem.* **263**, 11280–11283 (1988).

CHAPTER 5

ANALYSIS OF THE ARABIDOPSIS UBIQUITOME THROUGH DIGLYCINE SCANNING

INTRODUCTION

The model plant species *Arabidopsis thaliana*

The investigation of plant-specific pathways is critical for understanding the molecular biology of such critical functions as energy production, protection, and agricultural applications that help sustain our world. A prominent model plant species for studying these functions is *Arabidopsis thaliana* – a Cruficerous weed and member of the mustard family. As a model organism, Arabidopsis has enjoyed much attention, and was the first flowering plant to have its genome sequenced,[1] revealing over 25,000 genes distributed across 5 chromosomes. Arabidopsis presents numerous technical advantages for genetic and genomic studies including short generation time, the ability to self-fertilize, a considerable repertoire of genetic resources and a relatively small genome (147 Mbp) makes it an attractive model species for understanding plant biology. Nonetheless, shortcomings do exist, particularly from a proteomics perspective. Proteome-level studies in Arabidopsis have proven to be challenging due in part to the subset of proteins that are unique to plants, a paucity of commercially available antibodies and the presence of wide dynamic range of relative protein abundance across developmental time and space (e.g. ribulose 1,5-bisphosphate carboxylase oxygenase [RuBisCO]) that can complicate the *de novo* identification of proteins along with their isoforms and splice variants.[2,3] Furthermore, a low correlation exists between gene activity and cognate protein abundance, making protein predictions challenging.[4] The Arabidopsis Information Resource (TAIR) has compiled genetic and proteomic data that serves as a repository for data and tools to study this model plant. The most recent tabulation of proteins found on TAIR and other open-access sources, referred to as Araport11, contains 27,655 protein-coding genes.[5]

Ubiquitin, a prevalent post-translational modification with plant-specific effects

As sessile organisms being exposed to unique sudden and extreme evolutionary challenges, plants have evolved unique strategies to accommodate variations in environmental conditions. These conditions require the rapid changes to the plant's metabolism, patterning, and development. Unbiased genetic studies have identified components of the protein ubiquitination machinery as an important post-translational gene regulatory mechanism for the regulation of plant processes including, circadian rhythm, light morphogenic response, pathogen defense, and flowering, to name a few.[6,7,8] The importance of ubiquitination in the regulation of Arabidopsis patterning and development is apparent in the number of E3 Ub ligases, a number ~10 times larger than in humans through the combinatorial potential of the Ub machinery, where the Arabidopsis genome encodes for two E1, ~37 E2, and over 1,300 E3 enzymes,[6] which differ from humans having 8 E1, 42 E2, and ~600 predicted E3 enzymes.[9] The SCF family of E3 ligases is highly represented in Arabidopsis. The SCF complex is composed of ASK (SKP1 in non-plants), CUL1, RBX and an F-box protein. The Arabidopsis genome encodes for ~700 F-box proteins, suggesting that SCF complexes are major contributors to the Arabidopsis ubiquitome.[10] The combinatorial diversity of the 21 SKP and over 700 F-box proteins is estimated to account for >10,000 different E3's, reflecting Ub control in plants modifying a wide array of protein targets and Ub sites which has remained poorly characterized.[11]

Auxins are a class of prevalent plant growth regulatory compounds (PGRs) that are involved in almost every facet of Arabidopsis patterning and development, and are intimately tied with the ubiquitin proteasome system (UPS). Auxins regulate transcription

through two families of transcription factors, auxin/indole-3-acetic acid (Aux/IAA) and Auxin response factors (ARFs).[12] Auxins signal through interaction with various auxin binding proteins. Auxin can stimulate the destruction of Aux/IAA transcription factors by binding onto the transport inhibitor response 1/auxin signaling F-boxes (TIR1), a member of the E3 Ub ligase family.[10] Many other examples of plant-specific UPS targets have been reported, such as Ub regulating abscisic acid (ABA) signaling affecting salt and drought stress response and stomata opening.[13,14] Furthermore, UPS activity allows for photomorphogenesis to occur specifically during the day, through degradation of DELLA proteins activating gibberellin signaling.[15] Transcription factors critical for plant development and response to abiotic stressors have also been identified in ubiquitome approaches.[15] Universally, some ubiquitinated substrates are conserved throughout most eukaryotic organisms studied, with divergence occurring in early protists.[16] Many proteins involved in the UPS are ubiquitinated themselves, as well as proteins involved in endocytosis as mentioned previously.[17]

To date, most MS-driven ubiquitome characterization studies have utilized enrichment of the diglycine remnant left on Ub-modified lysine residues after digestion followed by data-dependent acquisition (DDA) modes, where the most abundant precursors are selected for fragmentation (**Figure 5.1**).[18] To avoid the bias of identifying only the most abundant precursors, data-independent acquisition (DIA) methods have been developed, which require no knowledge of precursor ions to record fragment ions.[19] In this method, all precursors derived from an isolation window are analyzed, allowing for the analysis of nearly every precursor, which has become a powerful acquisition technique for studying PTMs,[19] and potentially allows for the identification of more ubiquitination sites.

The largest study of the human ubiquitome identified 63,000 sites on 9,200 different proteins, employing both DDA and DIA approaches.[20] In contrast, the most comprehensive ubiquitome analysis in Arabidopsis has thus far yielded 3,000 sites located on 1,607 proteins.[21] This dataset was obtained using the labor-intensive combined fractional diagonal chromatography (COFRADIC) model and DDA MS. In this method, primary amines are blocked via acetylation and cell-free extracts are incubated with the DUB USP2, to form free primary amines (previously ubiquitinated), which can then be tagged allowing for the isolation of the previously Ub-modified proteins.[20] While these studies have served to expand the Arabidopsis ubiquitome, the number of identified proteins is still small in comparison to the human data set.

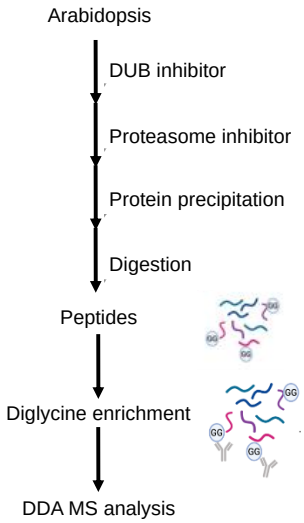


Figure 5.1. General diglycine enrichment workflow.

Arabidopsis was grown and typically treated with the proteasome inhibitor MG132 prior to protein extraction. Following extraction, proteins were precipitated and subsequently enzymatically digested, typically with trypsin, prior to enrichment with anti-diglycine antibodies. Following enrichment, peptides are submitted to MS analysis using the DDA approach.

Current state of the Arabidopsis ubiquitome

In all ubiquitome investigations, a recurring theme is a low conservation of ubiquitination sites relative to the identity of the ubiquitinated protein suggesting that the *site* of modification on the protein may not be as critical as the *act* of being ubiquitinated.[21,22] This finding has made the prediction of Ub sites difficult, further propelling the need to develop robust methods to identify this prevalent and critical modification. To expand the Arabidopsis ubiquitome, we have developed a protocol employing RuBisCO depletion followed by on-column trypsinization of plant extracts and diglycine-remnant enrichment using anti-diglycine antibodies. Samples were analyzed using a DIA method resulting in the identification of ~400 ubiquitination sites in 160 proteins with 398 ubiquitination sites and 49 proteins being novel. To our knowledge, this is the first DIA MS method used to study Arabidopsis ubiquitination employing a method without complex sample preparation and enrichment. Using this method to identify ubiquitinated proteins *in planta* can deepen our knowledge on how plant patterning and development is regulated in a post-translational manner.

MATERIALS AND METHODS

Arabidopsis growth conditions and treatments

Arabidopsis thaliana ecotype Col-0 seeds were sterilized with 70% ethanol, 50% bleach and 0.1% triton, and vernalized for 2 days at 4˚C in the dark. Seeds were then grown in liquid Murashige and Skoog (2.165 g/L MS salts, 1% sucrose, pH 5.7) media in a 21˚C growth chamber set to 24-hour light, with agitation, for 14 days. For MG132-treated plants, media was changed 18 hours prior to extraction and supplemented with 10 μM MG132.

Synthetic K48- and K6-linked polyubiquitin chain identification

Optimization of reaction conditions for large-scale diglycine profiling of Arabidopsis extracts began with the use of synthetic K48- and K63-linked polyUb chains (Enzo Life Sciences). To allow for more efficient and cost-effective analysis, we utilized MALDI-TOF MS/MS analysis to evaluate our initial optimization experiments. A critical parameter examined was the trypsin:protein ratio. Digestion of synthetic polyUb chains allowed for the identification of ions belonging to K48- and K63-linked polyUb chains. Diagnostic ions belonging to K48 (*m/z* 1460 (digly: modified; LIFAGK(GG)QLEDGR)), and human K63 (*m/z* 2244 (digly-modified; TLSDYNIQK(GG)ESTLHLVLR)) were observed and monitored during optimization methods.

Plant extracts were digested with the synthetic polyUb chains, using the diagnostic ions previously identified as guides to determine digestion efficiency. One format of digestion involved pre-mixing plant extract with synthetic peptides, followed by trypsin digestion using increasing protein:trypsin rations (1:100/50/25). The alternative method involved separately digesting extracts and synthetic polyUb chains using the same trypsin ratios and combining the peptides prior to MALDI-TOF analysis. Using a trypsin ratio of

1:50, we were able to identify diagnostic ions belonging to endogenous K63- and K48-linked polyUb from plant extracts. To increase the intensity of these ions and prepare for large scale experiments, we investigated the feasibility of employing anti-diglycine antibodies using this MALDI-TOF MS/MS method. Mouse monoclonal anti-diglycyl-lysl antibody (Lucerna, clone GX41) was used for these enrichment pilot studies and is described below. Enrichment showed a qualitative increase in diagnostic ions characteristic of trypsinized K48 (*m/z* 1460)—or K63 (*m/z* 2228)—polyUb linkages, propelling measurements of full-scale enrichments.

MALDI-TOF analysis trypsin digests

To prepare samples for MALDI-TOF measurements, peptides were combined 1:1 (va:v) with matrix (10 mg/mL alpha-cyano-4-hydroxycinnamic acid, 65% acetonitrile, 35% H2O, and 0.1% trifluoracetic acid), spotted on a MALDI target plate (Applied Biosciences), and allowed to crystalize before desorption.

A Voyager-DE pro matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF) mass spectrometer (Applied Biosystems) was used to evaluate the optimization experiments. The MS was operated under post source decay and was used to evaluate selected parent ions in reflectron mode for sequence determination, stitching together 8–10 segments of various mirror ratios and captured using Data Explorer software.

Alternate protocols

Protein extraction based on *Oryza sativa*

Plant protein extracts were prepared using a modified protocol developed for Rice (*Oryza sativa*; Xie et al., 2015).[23]Arabidopsis plants grown on solid media were ground in

liquid nitrogen and resuspended in a urea-based lysis buffer (20 mM HEPES, pH 8, 8 M urea, 1 mM sodium orthovanadate, 1 mM beta-glycerophosphate, and 2.5 mM sodium pyrophosphate). Lysates were then sonicated twice at 50 W for 10 seconds followed by 1-minute incubations on ice. Sonicated lysates were then centrifuged at 2000 *g* for 20 min at 4˚C. Cleared protein lysates concentration was determined via Bradford assay. 10 mg of total proteins were treated with 1.25 mM dithiothreitol (DTT) for 30 min at 55˚C followed by alkylation using 10 mM iodoacetamide (IAM) for 15 min at room temperature in the dark. Alkylated samples were digested overnight at 37˚C with sequencing grade trypsin (Promega) at a ratio of 1:50. Digestion was quenched with 0.1% trifluoracetic acid (TFA) and samples were desalted as described below and subsequently dried prior to diglycine enrichment.

Protein extraction based on Arabidopsis cold stress response,[24] by Waters®

Arabidopsis plants were removed from growth media and ground with liquid N_2 and resuspended in 10% TCA/0.07% β-mercaptoethanol in acetone. Precipitated proteins were washed in acetone and resuspended in extraction buffer (8 M urea, 2% CHAPS, 30 mM tris-HCl, pH 8.5). Extracts were diluted in water, 1:10, and digested with sequencing-grade trypsin (Promega) at a 1:100 trypsin:protein ratio overnight at 37˚C. Peptides were diluted and prepared for desalting and subsequent diglycine enrichment.

Finalized preparatory method

Optimized method: Protein extraction and RuBisCO depletion

14-day-old seedlings were removed from the growth media, dried, and weighed. 5 mg of plant material were flash frozen in liquid N_2, and ground into a fine powder using a chilled mortar and pestle. Cells were lysed using extraction buffer (0.5 M tris-HCl, 20 mM

MgCl2, 2% Igepal (Sigma-Aldrich) pH 8.3), and centrifuged at 12,000 *g* for 10 min at 4˚C. The supernatant was then collected and cleared through filter passage.

To deplete RuBisCO, 1% protamine sulfate (PS) was added to the clarified lysate to a final concentration of 0.02–0.12%, mixed and incubated on ice for 30 min. Samples were then pelleted at 12,000 *g* for 10 min at 4˚C. Protein concentration was determined by Bradford assay. Approximately 10 mg of plant material was used to generate 2.5 mg/mL of protein extract.

Optimized method: Filter-aided sample preparation (FASP) of Arabidopsis extracts

FASP has been commonly used to allow for the complete digestion of proteins in a sample and is compatible with highly denaturing reagents that are otherwise problematic for mass spectrometry analysis such as urea, in turn for efficient lysis of plant cells. 10 mL of lysate was mixed with 5 mL of UA buffer (8 M urea, 0.1 M Tris-HCl, pH 8.5) and loaded onto an Amicon ultra centrifugation 3 kDa filter (Millipore). Proteins were reduced on-column by addition of DTT to a final concentration of 5 mM for 45 min. Excess DTT was removed by centrifuging the filter at 4,000 *g* for ~15 min. Eluant from the column was removed and proteins subsequently alkylated by adding 10 mL UA buffer supplemented with 2 mL of 50 mM iodoacetamide for 30 min at room temperature in the dark. Samples were centrifuged at 4,000 *g* for ~15 min, and the eluant was discarded. The column was then washed three times with 10 mL UA buffer and prepared for protein digestion by twice conditioning with 10 mL of 50 mM ammonium bicarbonate (ABC).

Proteins were digested by addition of trypsin in 50 mM ABC added 1:100 (~2.5 µg) and incubated at 37˚C for 1 hour and shaking. An additional amount of trypsin was added to the sample and further incubated overnight. Peptides were then eluted by

centrifugation at 4,000 *g* for 15 min. ~1 mL of ABC was then added for a second elution. Peptides were then dried to near completion using a speed vacuum (Thermo).

Desalting of peptides

To remove salts that would interfere with MS analysis, trypsinized peptides were desalted by reverse phase chromatography. Peptides were resuspended in 200 µL 0.15% TFA (Sigma, T6508-1) prior to column addition. Oasis CLB cc100 (Waters) columns were activated with 100 µL acetonitrile (ACN, from Burdick and Jackson) and washed with 0.15% TFA prior to peptides addition to the column, then washed again with 0.15% TFA and eluted with 100 µL elution 1 (70% 0.15% TFA, 30% ACN), 100 µL elution 2 (50% 0.15% TFA, 50% ACN), 200 µL elution 3 (20% 0.15% TFA, 80% ACN). Cleaned peptides were concentrated via SpeedVac.

Diglycine enrichment

Diglycine-containing peptides were enriched using the PTMScan® ubiquitin remnant motif (K-*ε*-GG; Cell Signaling Technology). Dried peptides were reconstituted in 1.4 mL of 1X IAP buffer (50 mM 3-(N-morpholino) propanesulfonic acid, 10 mM sodium phosphate, 50 mM. NaCl) (supplied in kit, diluted with MilliQ water) and enriched overnight at 4˚C. Following enrichment, samples were washed as directed and peptides were eluted by addition of 0.15% TFA. Peptides were then desalted as described above.

A second antibody utilized for diglycine enrichment was the anti-diglycl-lysyl antibody GX41 (Lucerna). Peptides were resuspended in cold IAP buffer followed by addition of 5 µg of antibody for 4 hours at 4˚C. Antibodies were captured using 20 µL of 50% protein G resin (GenScript, L00209) for 2 hours at 4˚C. Supernatants were removed following a 5-min centrifugation at 2,000 *g*, and washed three times with 1 mL PBS-T for

5 min, followed by 2 washes with PBS (Hyclone). Peptides were eluted by addition of 100 µL 0.15% TFA incubated at room temperature for 15 min. Eluted peptides were desalted as described above.

LC-MS configurations

Peptides were loaded into glass vials and injected into a 2D nanoAcquity UPLC system (Waters) using a 1.8-µm HSS T3 75 µm X 150 mm reverse phase column. Aqueous mobile phase A (0.1% formic acid) and organic phase B (acetonitrile with 0.1% formic acid) were used to equilibrate the column and load generated peptides at 97:3 (A/B). For early runs, peptides were eluted using a 60-min acetonitrile gradient (3–10% B for 5 min, 10–25% B for 28 min, 25–50% B for 8 min, 50–85% B for 3 min). Samples displaying broad peaks during the 60-min gradient utilized a 120-min acetonitrile gradient was used (3–10% B for 5 min, 10–25% B for 47 min, 25–50% B for 40 min, 50–85% B for 3 min, and 2 min at 85% B). Separated peptides were directly injected into a SYNAPT G2-Si mass spectrometer (Waters) using a 30 V cone voltage and 3.8 kV capillary voltage. The instrument was run using the data-independent acquisition mode with the ion mobility chamber activated. Measurements acquired without ion mobility performed low energy scans at 4 eV while high energy scans were acquired from a 15–45 eV ramp in positive resolution mode scanning from 50-2,000 *m/z* with a scan rate of 0.6 s. External calibrations were performed using 50 fmol/µL [Glu1]- fibrinopeptide B in the lockmass channel.

Peptide identification from database

Raw data from the mass spectrometer was processed using Progenesis QI (Nonlinear Dynamics) or Protein Lynx Global Server (PLGS V2.02 and V3.03) to align chromatography, normalize findings, and perform peptide identification using the TAIR10

Arabidopsis protein lists (35,386 proteins, downloaded May 2019) or TAIR10 trimmed (27,416 proteins, made December 4, 2019). Various parameters were used: a low energy noise reduction threshold of 135, a high energy threshold of 30, and a 750-intensity threshold. Protein identification parameters included: a minimum of three fragment ions per peptide, minimum of 1 fragment ion match per protein, maximum of two missed cleavages using trypsin, and a false discovery rate of 4%. A fixed modification of carbamidomethylation (+57.0214) and variable modifications including methionine oxidation (+15.9949), lysine diglycine modification (+114.0429) and lysine LR-diglcyine (+383.2281) were employed. Data was lock mass corrected post-acquisition.

RESULTS

MALDI optimization of trypsin digestion

Our early investigations employed digestion of synthetic polyUb (K48- and K63-linked chains) to determine the optimal amount of trypsin and antibody to use for peptide preparation and enrichment. As with endogenous polyUb chains, digestion of synthetic polyUb chains results in the formation of peptides bearing the diglycine remnant and are amenable to enrichment with anti-diglycine antibodies. The optimization experiments involved supplementing protein preparations with synthetic polyUb chains, either before or after digest with varying quantities of trypsin. The resulting peptide preparations were enriched for diglycine peptides using the GX41 anti-diglycine antibody (Lucerna Technologies) and measured using MALDI-TOF MS. Detection of K48- and K63-modified ubiquitin peptides were strongly detected using a trypsin-to-protein ratio of 1:50 (wt:wt) (**Figures 5.2A** and **B**) and used all subsequent experiments. We then enriched for endogenous K48- and K63-modified ubiquitin peptides from Arabidopsis extracts. We were successful in detecting the diagnostic K48-modified (*m/z* 1460) and K63-modified (*m/z* 2227) peptides following enrichment (**Figure 5.2C**), allowing us to perform larger-scale experiments.

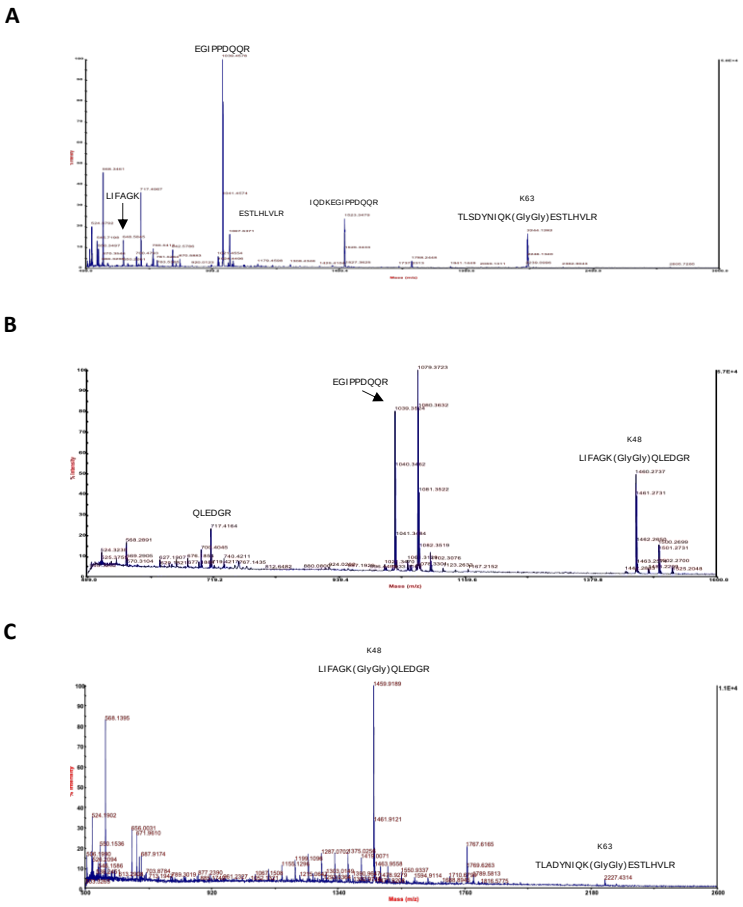


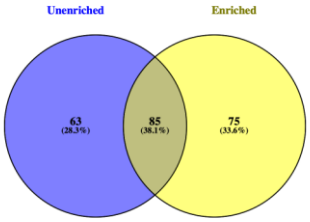
Figure 5.2. Optimization of trypsin digestion via MALDI-TOF analysis.

MALDI-TOF was used to evaluate the efficiency of trypsin digestion of synthetic K48- and K63-linked polyubiquitin. A trypsin-to-protein ratio of 1:50 (wt:wt) was found to be optimal for digestion of (**A**) human K63-linked polyubiquitin, modified peptide TLSDYNIQ**K**ESTLHLVLR (*m/z* 2245) and (**B**) K48-linked modified peptide LIFAG**K**QEDGR. (**C**) MALDI-TOF analysis of trypsin-digested Arabidopsis extracts. The feasibility of enriching diglycine bearing peptides was evaluated through trypsin digestion of MG132-treated plant extracts followed by diglycine diglycine scanning (Lucerna™, clone GX41), allowing for detection K48- (*m/z* 1459.91) and K63-linked (*m/z* 2227.14) ubiquitin. Bolded sequences represent diglycine-modified lysine residues.

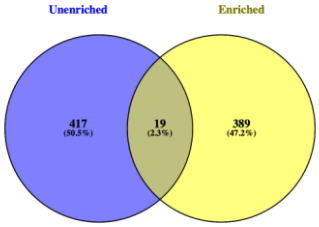
Ubiquitome analysis of plant extracts

We developed a protocol combining previously described methods including RuBisCO depletion and on-column trypsin digestion using FASP.[25,26] Following digestion, aliquots of the generated peptides were stored for analysis prior to diglycine enrichment. To date, an overwhelming majority of ubiquitome studies utilize traditional data-dependent acquisition modes. We employed a data-independent (DIA) method in an attempt to expand the Arabidopsis ubiquitome catalogue. Examining these peptides yielded 146 unique proteins with 436 unique peptides, with 48 proteins and 422 peptides being novel, and modified in at least 2 biological replicates (**Appendix D.1**). These analyses also ensured peptides were present in our samples providing confidence in analyzing our enriched samples. The use of successive enrichments using two different anti-diglycine antibodies allowed for enrichment of 160 proteins with 408 unique sites based on proteins found in at least 2 biological replicates (**Appendix D.2**). An overlap of 81 proteins was found between enriched and unenriched samples (**Figure 5.3A**), with only 19 modified peptides being present in both, belonging to Ub, GTP-binding elongation factor Tu family protein, HSP70, PDI, S-adenosylmethionine synthase, methionine synthase, serine transhydroxymethyltransferase 1, phosphoglucomutase, glutamate-1-semialdehyde-2 1-aminomutase and photosystem II subunit. Analysis of our dataset showed that over ~35% of diglycine-enriched proteins possessed two peptides that were modified, with a majority of peptides containing 20 residues or less (**Figure 5.3C** and **E**). Meanwhile, modifications sites identified in unenriched extracts identified 25% of proteins having six or more modified peptides (**Figure 5.3 D** and **F**), showing a slight difference in the pooled peptides upon enrichment.

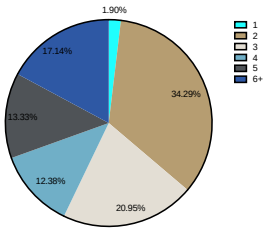
A



B

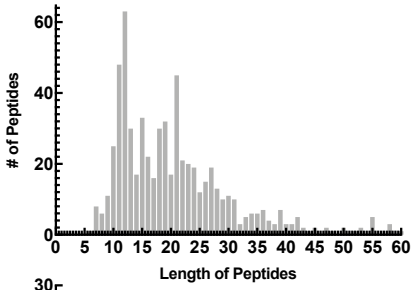


C

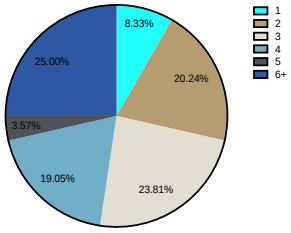


419 Ub sites in 160 proteins, 408 peptides

D



E



450 Ub sites in 146 proteins, 436 peptides

F

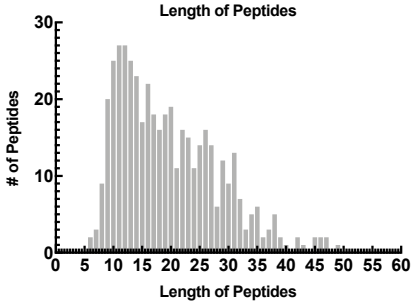


Figure 5.3. Summary of the Ubiquitome of Arabidopsis extracts using LC-MS^{E}.

Plant extracts prepared using protamine sulfate and FASP were measured using a Synapt G2Si under a DIA method. Likewise, peptides enriched using successive diglycine enrichments were evaluated. Measurements from three independent biological replicates were analyzed and data outputted using PLGSv3.03 (**Appendix D.1–3**). (**A**) Comparison of protein identifications between enriched and unenriched peptide measurements, finding an overlap of 81 proteins. (**B**) The unique peptides identified using diglycine enriched and unenriched samples were analyzed displaying an overlap of 13 peptides. Pie charts displaying the composition of the peptide datasets correlating the number of diglycine modified peptides detected per protein, for (**C**) enriched and (**E**) unenriched preparations. The abundance of peptides of in accordance to their lengths for both the (**D**) enriched and (**F**) unenriched fractions was evaluated. Venn diagram was produced using Venny (https://bioinfogp.cnb.csic.es/tools/venny/index.html).

RuBisCO depletion

RuBisCO is the most abundant protein in the biosphere and comprises a large proportion of soluble proteins in the plant cell. Owing to their high abundance in photosynthetic tissues (typically 10% wt:wt) RuBisCO-derived peptides can obscure enrichment and detection of other diglycine-modified peptides, which became apparent in early experiments in which RuBisCO was not otherwise depleted. Protocols have been developed to deplete RuBisCO in Arabidopsis lysates based on previous studies. For this study, we employed a protamine sulfate (PS)-mediated RuBisCO precipitation method previously described.[25,27] This involved adding PS to Arabidopsis lysates, and using the soluble (RuBisCO-depleted) fraction for further processing. CBB stains were performed to qualitatively assess the reduction in the heavy chain of RuBisCO (**Figure 5.4A**). RuBisCO was not fully depleted in the sample, and modified RuBisCO peptides were still identified in our MS analysis (**Figure 5.4B**).

Validation of PEPC ubiquitination

Phosphoenolpyruvate carboxylase1/2 (PEPC1/2) was found in each of the sequential enrichments, with the peptide EYGV**K**(GlyGly)LTMFHGR consistently found to be oxidized and diglycine-modified at Lys^{628}. Previous conventional biochemical studies identified this peptide to be similarly modified in other plant species, notably in in *Ricinus communis* (common Castor Bean). An amino acid sequence alignment was performed to show the high evolutionary conservation with other plant species such as Castor as well as Rice (**Figure 5.5C**). Validation of our MS findings was undertaken using PEPC immunopurification followed by immunoblotting using anti-ubiquitin (clone P4D1)

(**Figure 5.5A**). A spectrum showing the modified peptide containing Lys^{628} of PEPC was extracted from the raw data validating the identification (**Figure 5.5B**).

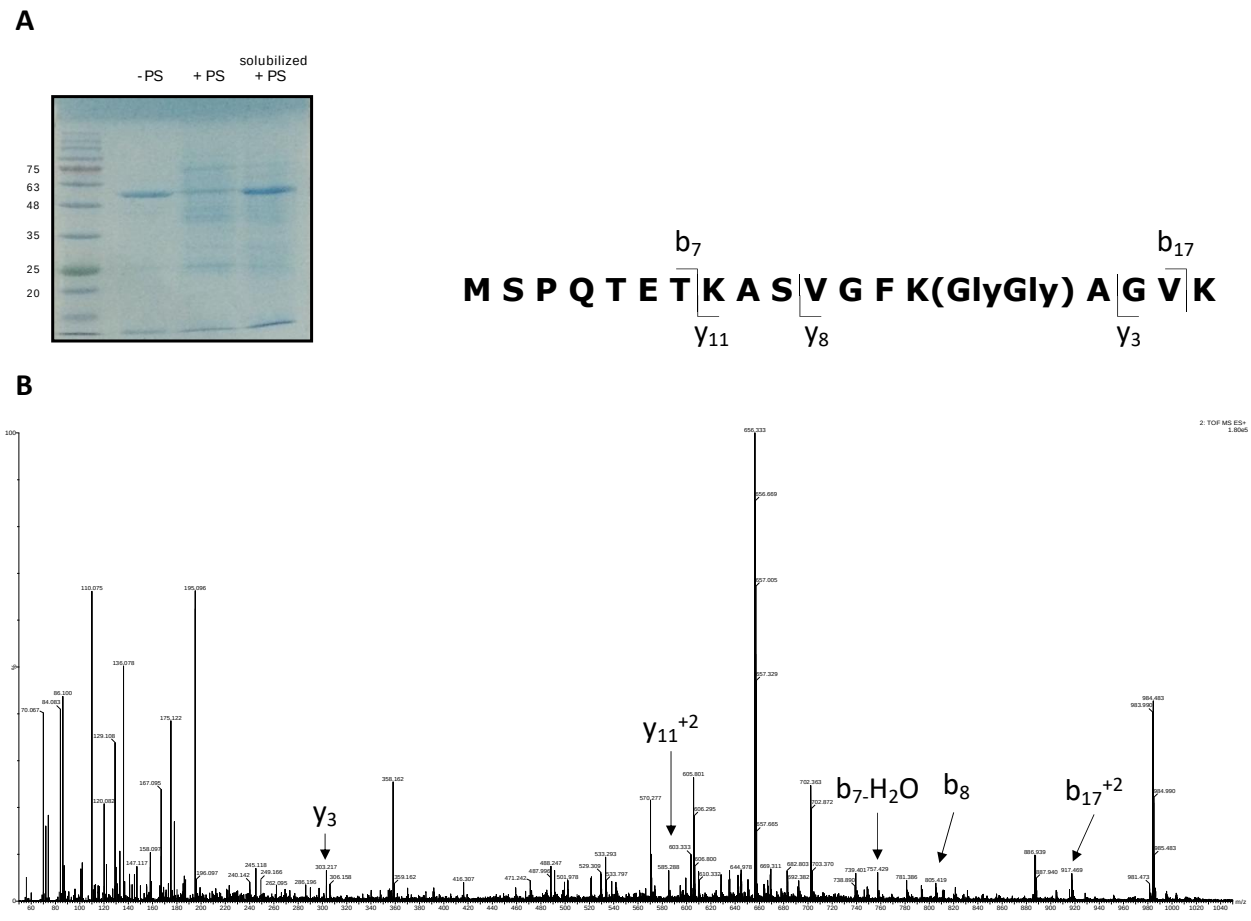


Figure 5.4. Sample RuBisCO depletion.

Protamine sulfate was added to plant extracts prior to trypsin digestion to produce RuBisCO-depleted extracts.

(**A**) Coomassie Brilliant blue (CBB) staining was used to analyze extracts prior to depletion (-PS), depleted plant extracts (+PS), and the resulting pellet which was resolubilized for CBB analysis.

(**B**) High energy MS^E scan of diglycine-modified RuBisCO peptide, MSPQTET**K**(GlyGly)ASVGFKAGVK identified.

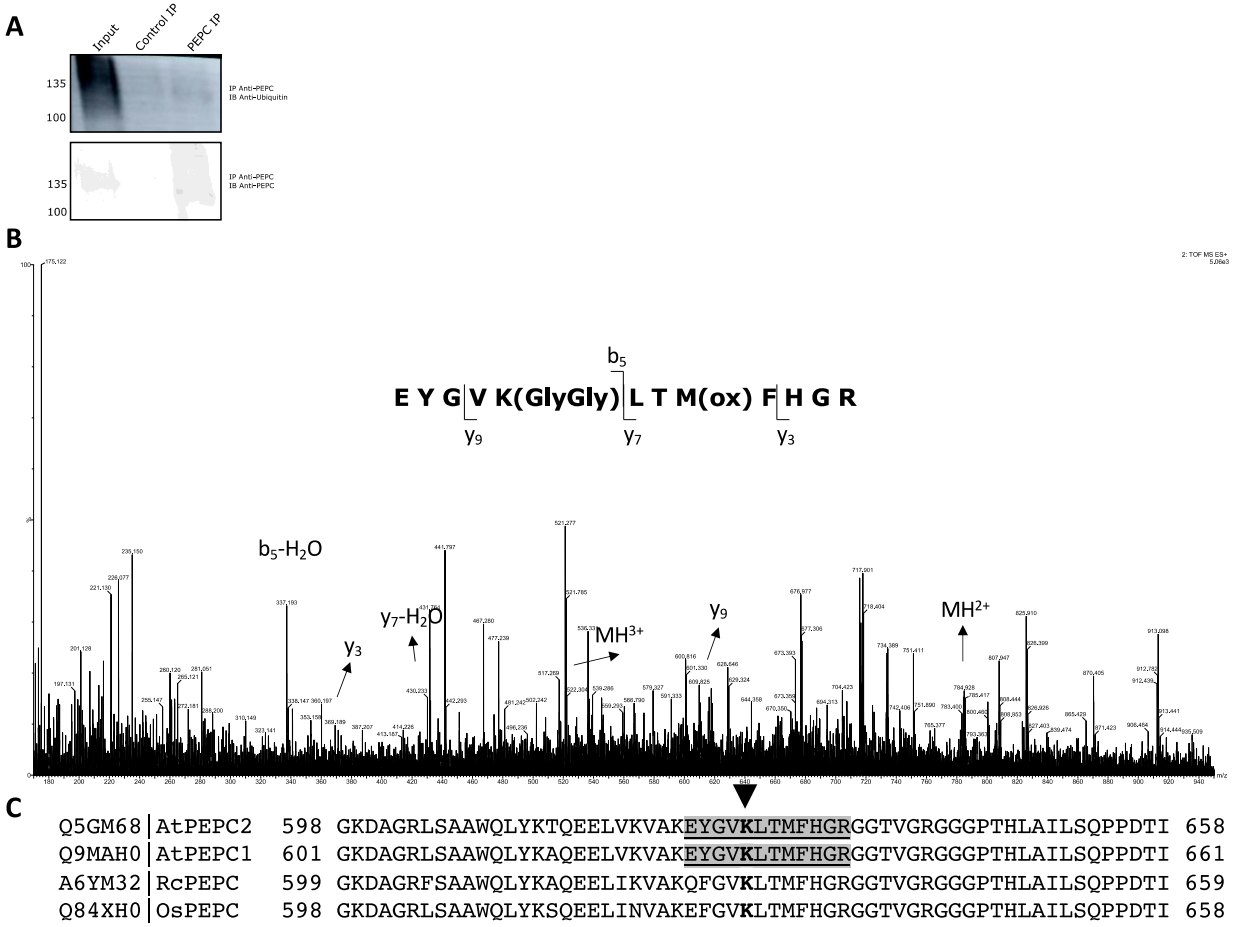


Figure 5.5. Ubiquitination of Phosphoenolpyruvate carboxylase.

(A) Immunoblot detection of PEPC protein and ubiquitination using anti-ubiquitin antibody (P4D1).

(**B**) High energy scan of identified modified PEPC Lys^{628} (EYGV**K**(GlyGly)LTM(ox)FHGR).

(**C**) Sequence alignment of identified Arabidopsis (At)PEPC ubiquitination site to *Ricinus communis* (Rc) and *Oryza sativa* (Os) with UniProt identifiers. The conserved modified lysine is in bold and indicated with an arrow. The identified peptide is highlighted in grey.

Comparison to previously reported plant ubiquitomes

To assess the novelty of our results we compared the repertoire of peptides identified in our studies to other Arabidopsis studies (Kim et al., 2013 & Walton et al., 2016) as well as other plant species such as rice (Chen et al., 2018), tea leaves (Xie et al., 2019), rose petal (Lu et al., 2019), and wheat (Zhang et al., 2017).[28,18,29,30,31] Peptides identified in these landmark studies and ours belonged to Ub (K11-, K48-, and K63-linked chains), PEPC, and other abundant proteins (**Figure 5.6A**; **Appendix D.4**), suggesting a very low conservation of ubiquitination sites in proteomes. We evaluated the protein identification overlap with peptides identified using the COFRADIC method (Walton et al., 2016), where a low overlap of protein identification was similarly reported (**Figure 5.6B, Appendix D.4**). To determine if this overlap was unique to our dataset, we compared modified peptides identified in two of the largest Arabidopsis studies and rose petal, where we found a low level of overlap (**Figure 5.6C, Appendix D.5**).

Ubiquitination motifs

To further investigate the variability in ubiquitination sites seen between datasets, we searched for putative conserved motifs surrounding the modified lysine residues observed, with particular interest in the C-terminal lysine residues where ~40% of our identifications contained C-terminal diglycine modifications. A ubiquitinated lysine residue is generally resistant to trypsin cleavage; therefore, theoretically, no peptides should terminate with a C-terminal digly-modified lysine. Interestingly, our motif analyses show that where a peptide terminates in a C-terminal digly-modified lysine, a glycine residue can be found ~25–30 residues N-terminal to that C-terminal modified lysine

(**Figures 5.7A–D**). No conserved amino acid sequence surrounding the ubiquitinated lysine residue was readily observed.

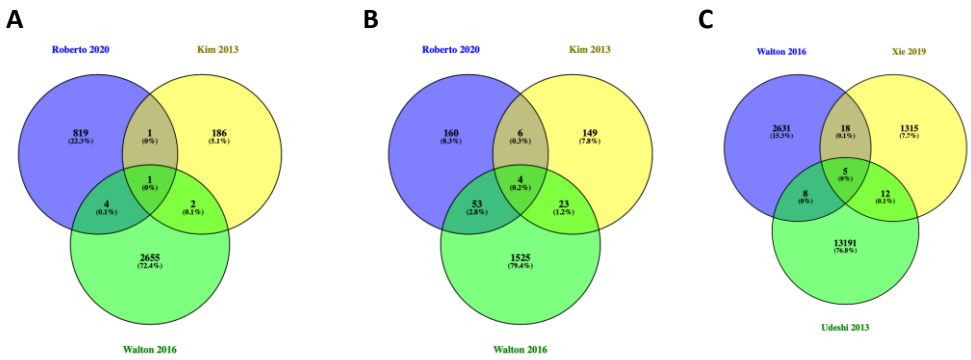


Figure 5.6. Comparison of plant ubiquitome datasets.

(**A**) To determine the contribution of this study to the previously known Arabidopsis ubiquitome, peptide identifications arising from this study to the two largest reported Arabidopsis ubiquitome studies (Kim et al., 2013, and Walton et al., 2016).

(**B**) Comparing the list of proteins identified in our study, Kim et al., 2013, and Walton et al., 2016.

(**C**) To determine the conservation of ubiquitination sites *in planta*, we compared peptides identified in the largest plant ubiquitome studies (Kim et al., 2013 and Xie et al., 2019 and Walton et al., 2016). A full list of comparisons is included in **Appendix D.4-5**. Venn diagrams were generated using Venny (https://bioinfogp.cnb.csic.es/tools/venny/index.html).

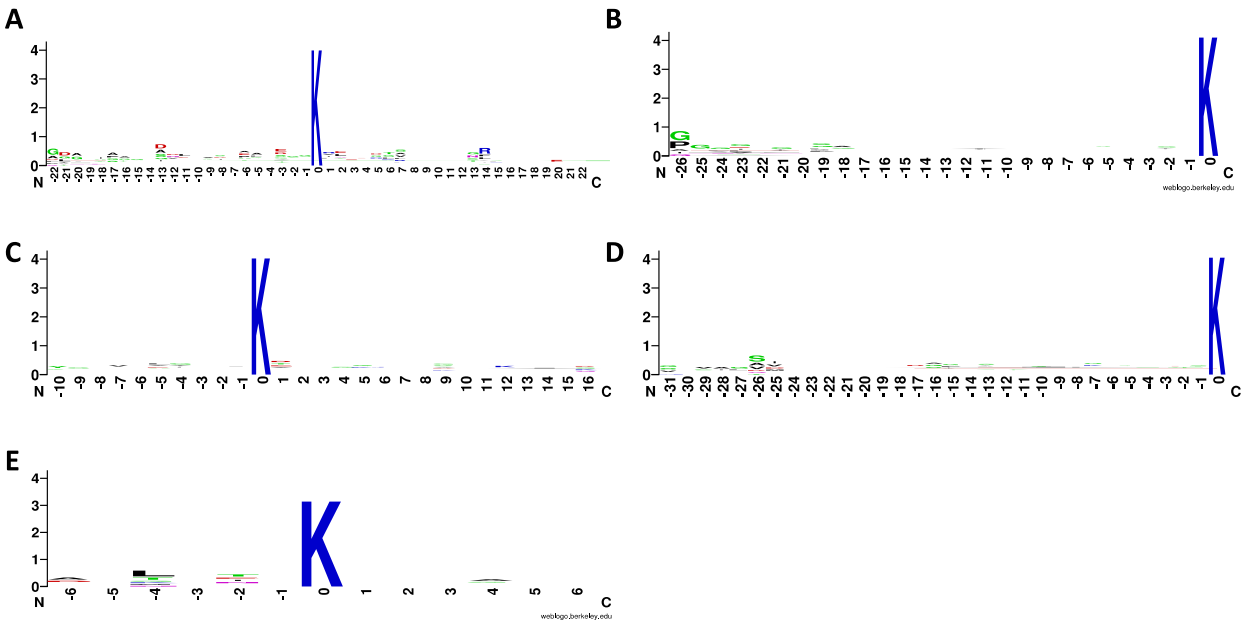
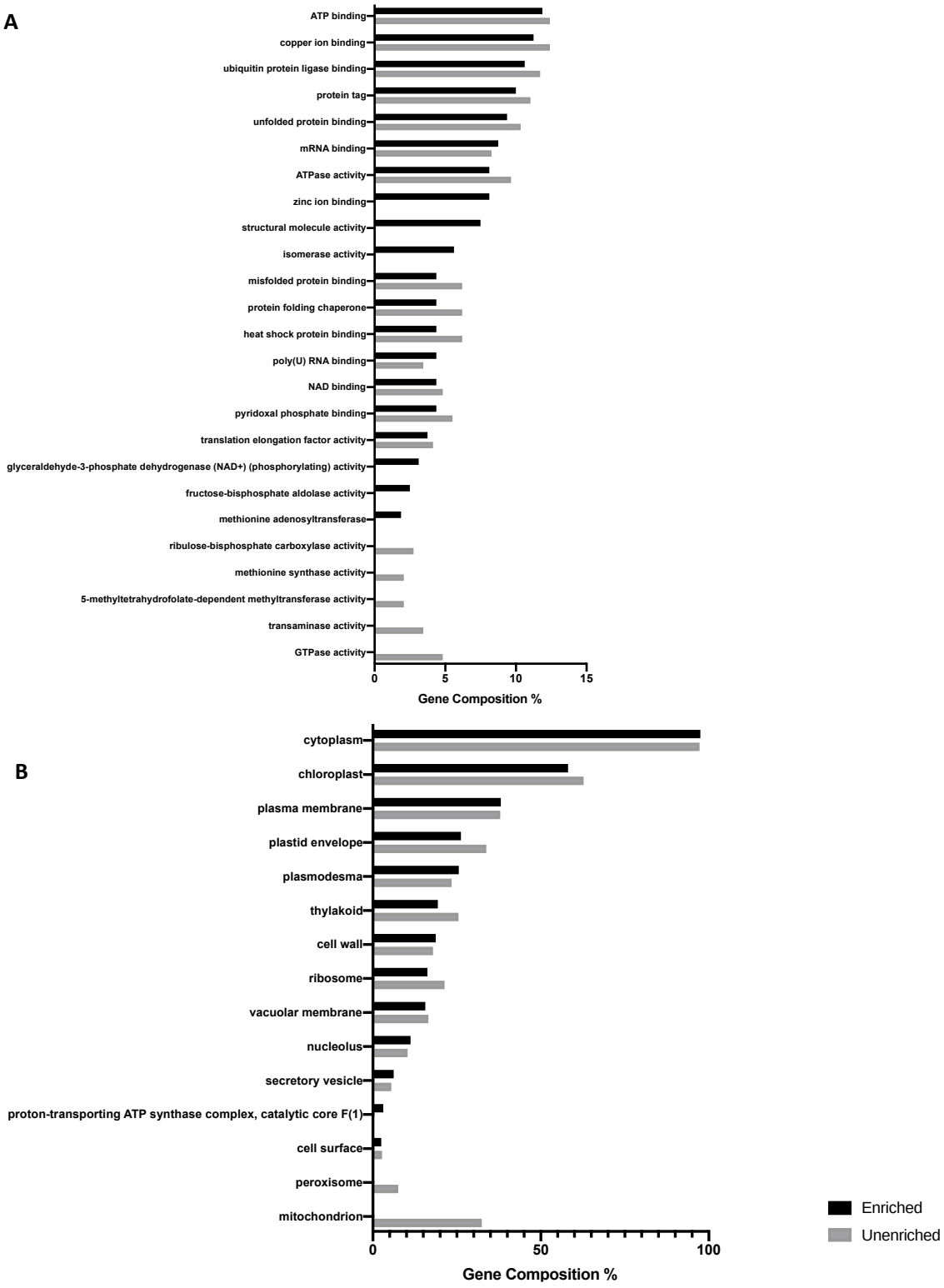


Figure 5.7. Peptide motif analysis.

Summary of ubiquitination motifs for peptides containing (**A**) a single internal lysine residue, central K at $x = 0$, and (**B**) a single C-terminal lysine, right-most K at $x = 0$. Motif analysis of peptides containing two lysine residues was also evaluated for (**C**) internal modified lysine residues, central K at $x = 0$, and (**D**) internal modified lysine residues ending with non-lysine residues, central K at $x = 0$. (**E**) Ubiquitination motif of peptides containing two lysine residues with detected C-terminal-modified lysine residues, last K at $x = 0$. Logo images were produced using WebLogo version 2.8.2 (https://weblogo.berkeley.edu).

Gene ontology

Diglycine-modified proteins identified in our scans were submitted to the PANTHER classification system for gene ontology (GO) analysis. The cellular component, biological process, and molecular functions of the proteins were analyzed (**Figure 5.8**). The molecular function of proteins found in unenriched and enriched fractions share the function of ATP-, copper ion-, and Ub protein ligase-binding (**Figure 5.8A**). The cellular component of the proteins identified were enriched in plant-specific structures, such as the chloroplast, plastid envelope, plasmodesma, thylakoids, vacuolar membrane, and the cell wall (**Figure 5.8B**). Several differences in the biological processes assigned to the proteins found using whole cell extracts in comparison to the diglycine enrichment (**Figure 5.8C**) were observed. Among the highest enriched terms using our protocol was 'response to cadmium ion,' whereas no enrichment had a large number of genes associated with 'response to light stimulus.' Proteins involved in response to cold and heat, as well as protein ubiquitination, were highly represented in both datasets. The GO terms generated represent many functions of Arabidopsis, with identifications found throughout the organism, showing low clustering of ubiquitinated peptides to known biological functions of Arabidopsis, further suggesting ubiquitination is widespread regulatory modification of Arabidopsis.


A
ATP binding
copper ion binding
ubiquitin protein ligase binding
protein tag
unfolded protein binding
mRNA binding
ATPase activity
zinc ion binding
structural molecule activity
isomerase activity
misfolded protein binding
protein folding chaperone
heat shock protein binding
poly(U) RNA binding
NAD binding
pyridoxal phosphate binding
translation elongation factor activity
glyceraldehyde-3-phosphate dehydrogenase (NAD+) (phosphorylating) activity
fructose-bisphosphate aldolase activity
methionine adenosyltransferase
ribulose-bisphosphate carboxylase activity
methionine synthase activity
5-methyltetrahydrofolate-dependent methyltransferase activity
transaminase activity
GTPase activity
0
5
10
15
Gene Composition %
B
cytoplasm
chloroplast
plasma membrane
plastid envelope
plasmodesma
thylakoid
cell wall
ribosome
vacuolar membrane
nucleolus
secretory vesicle
proton-transporting ATP synthase complex, catalytic core F(1)
cell surface
peroxisome
mitochondrion
0
50
100
Gene Composition %
Enriched
Unenriched

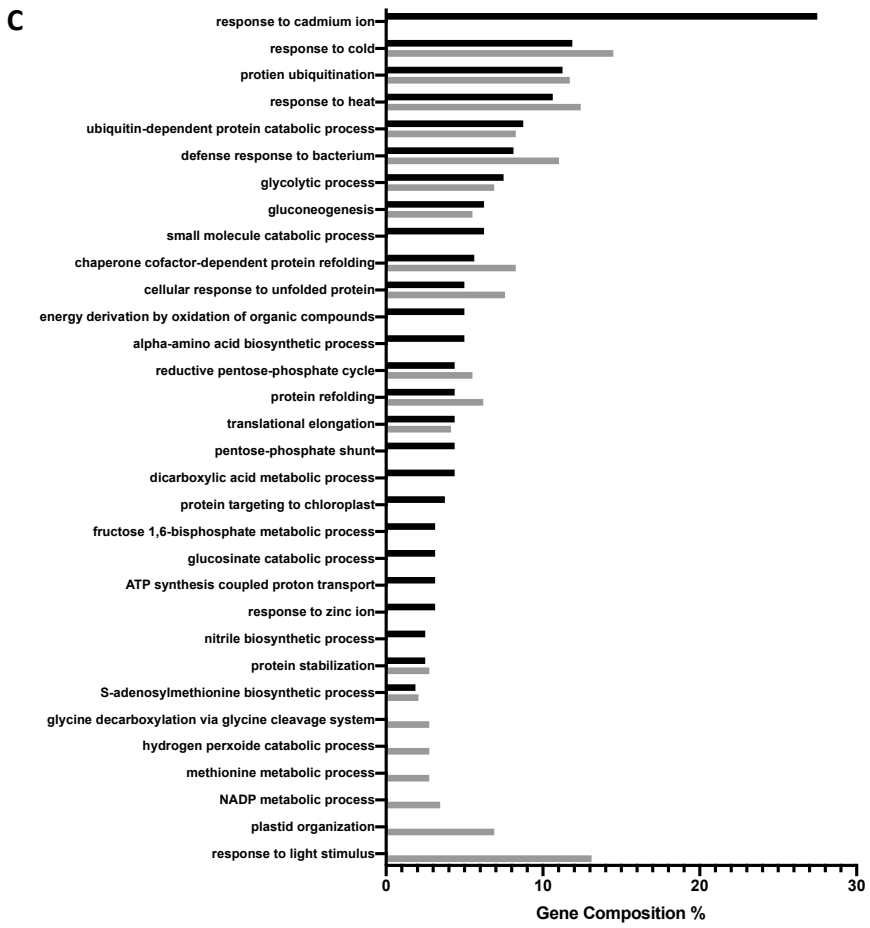


Figure 5.8. Gene ontology enrichment analysis of identifications.

Protein identifications were submitted to the PANTHER classification system (http://pantherdb.org) to generate gene ontology (GO) terms. Select terms were converted to gene composition percentages and plotted. The terms are shown for proteins found following enrichment (black bars) and prior to enrichment (grey bars) for (**A**) molecular function, (**B**) cellular component, and (**C**) biological process.

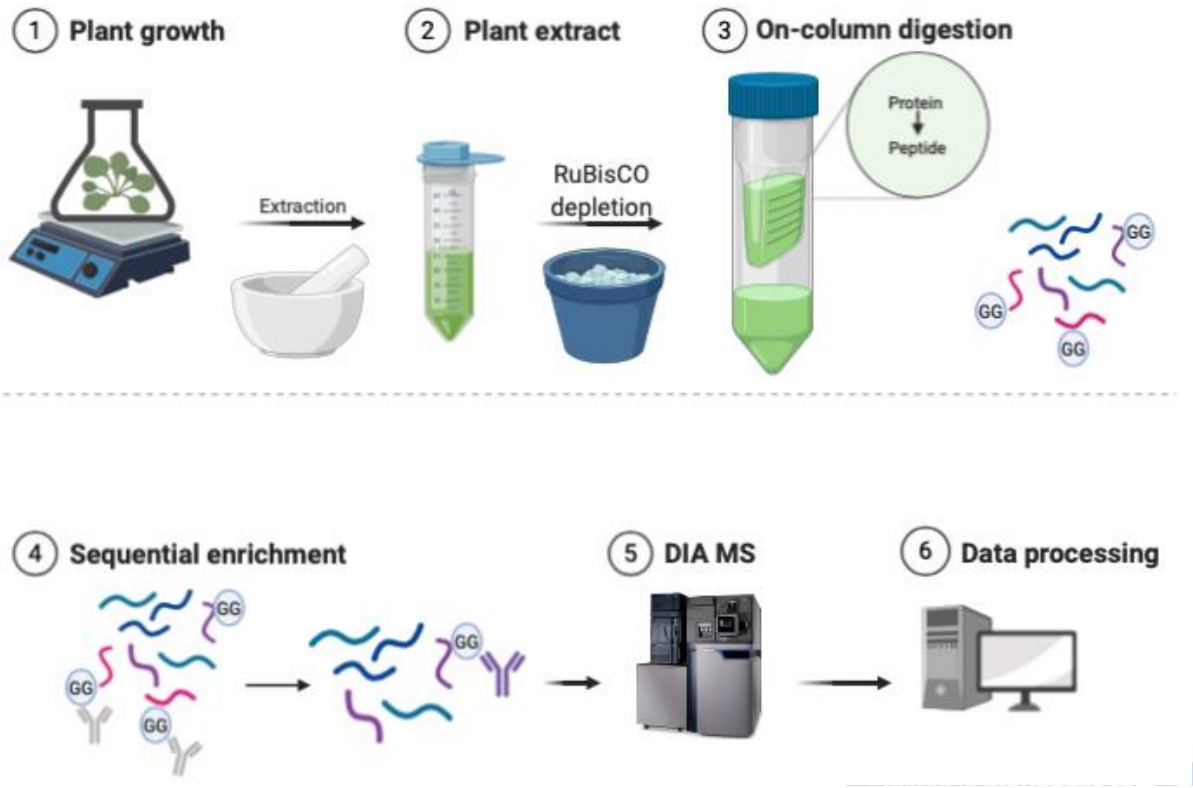


Figure 5.12. Diglycine enrichment workflow using DIA MS^E methodology.

2-week-old Col-0 Arabidopsis were imbibed with 10 µM MG132 for 24-hours prior to extraction. Extracts were subsequently depleted of RuBisCO and peptides are generated by FASP, followed by successive enrichments with different diglycine enrichments followed by MS^E analysis and subsequent data processing to identify ubiquitinated peptides. Image was generated using BioRender (https://biorender.com).

DISCUSSION

Analysis of Arabidopsis ubiquitome through diglycine scanning using DIA MS

Analysis of the Arabidopsis genome strongly suggests that Ub-mediated pathways are critical regulators in plant development and evolutionary success. In this study, we developed a DIA-MS method to aid in Arabidopsis proteomics analysis, targeted specifically to the characterization of the Ub-modified proteome in axenically-grown developing seedlings. A total of 160 proteins were identified with over 400 ubiquitination sites identified, with 81 proteins being and 398 novel peptides identified, using our method. The method developed utilizes RuBisCO-depletion using protamine sulphate followed by FASP.[3,26] Our CBB stains suggest that RuBisCO levels were successfully reduced, though not entirely depleted, given that peptides belonging to RuBisCO were identified in our MS analyses. Early protocols employing TCA precipitation resulted in the identification of only ~20 modified proteins using two different precipitation methods,[23,24] likely due to protein loss associated with precipitation.[32] This is consistent with our early trials using previously published protocols (data not shown). Implementation of FASP allowed us to avoid precipitation while using highly denaturing conditions (8M urea, 2% Igepal) for protein solubilization without interference with trypsin digestion. We performed successive enrichments of plant extracts, first using the CST® anti-diglycl-lysine antibody, followed by submitting the unbound material through a second enrichment (referred to as 'second pass') with anti-diglycine (Lucerna Technologies), which was captured using protein G following binding with the precleared material. A greater number of diglycine-containing peptides was routinely identified in first pass measurements (>600 peptides), while the second pass identified over 200 peptides, suggesting the majority of digly-

modified peptides were captured in the first pass enrichment. The use of second pass enrichment isolated peptides that were unique from the peptides enriched in the first pass, deepening modification identification. This result suggests that not all digly-containing peptides are captured in the first enrichment - possibly owed to saturation of the antibody.

Unenriched Arabidopsis extracts were also investigated, revealing >140 modified proteins (450 unique sites in 436 peptides, **Appendix D.1**). Querying the unenriched dataset against the enriched dataset showed an overlap of 85 proteins found in both analyses, including Ub and chaperone proteins. A 1.3% overlap in peptides between enriched and unenriched samples was detected with many belonging to Ub, heat shock binding proteins, GTP-elongation factor Tu, and single peptides belonging to photosystem II subunit, phosphoglucomutase, S-adenosylmethionine synthase, serine transhydroxymethyl-transferase 1, and methionine synthase 2. This low overlap in ubiquitinated peptides identified between nonenriched and enriched samples is expected since enrichments serve to fractionate the peptides available for measurements. The 38% of proteins identified in both unenriched and enriched belong to Ub family members, heat shock protein family members, and metabolic proteins like transketolase or ATP synthase, and a chaperone calreticulin. Tricarboxylic acid cycle protein phosphoenolpyruvate carboxylase (PEPC) was identified in our investigation consistent with previously reported findings where PEPC was identified as being ubiquitinated at Lys^{628} in other plant species (**Figure 5.5**).[23,33] This Lys^{628} site—which has been determined to be a monoubiquitination site—is conserved across different plants, and is critical in the regulation of PEPC.[33]

To further explore the variability in modified peptides isolated, we compared our peptide and protein identifications to previously published ubiquitomes from Arabidopsis

and other plant model species. A recurring theme in ubiquitome investigations is the lack of conserved motifs, hindering prediction of modification sites; only a handful of ubiquitination sites identified in our study have been reported by others. Examination of the Arabidopsis datasets obtained through COFRADIC or diglycine scanning were compared to rose petal diglycine enrichment, where <40 peptides matched from among thousands of identifications.[21,31,34] This is further supported with Logo analysis (**Figure 5.7**), which suggests there is no sequence conservation around the modified lysine. This finding suggests that many ubiquitination sites may not be conserved; rather, what may be conserved is the identity of the protein itself being ubiquitinated (regardless of site), allowing the protein to be cleared by the proteasome or otherwise impart its cellular function in the modified state. This is seen in our dataset by less than 50 of our 160 protein identifications being novel, showing a greater overlap in the modified proteins identified compared to peptides (398 unique). Nonetheless, conserved ubiquitination sites may be critical for protein function, like the mono-Ub modification of PEPC at Lys^{628}. The Ub protein target specificity we may be observing may be due to the conserved signals which the Ub machinery can recognize.

About 50% of peptides identified throughout our study contain the expected internal lysine residues, 44% contain C-terminal lysine residues, and ~6% possess unassigned modification sites. Peptides containing C-terminal digly-modified lysine residues are somewhat anomalous since it is believed that lysine ubiquitination prevents trypsin cleavage. A significant number of C-terminal diglycine-modifications were identified in our study, and this finding has been echoed in other studies, which identified 14 (~3%) unique peptides with C-terminal diglycine-modified lysine (found predominantly

on Histone H2B and 60S acidic ribosomal protein P1).[35] It was hypothesized that the positive charge from the diglycine tag's amino group can be recognized by trypsin, similar to the ε-amine of lysine. Similarly, rat brain ubiquitome studies found C-terminally diglycine-modified proteins. The authors conducted a deep analysis on these peptides, manually reassigning a number of the hits. This study concluded that the software wrongly assigned C-terminal ubiquitination in peptides with more than one lysine residue, with software biasing modification assignment to more C-terminal positions requiring them to reassign sites, allowing more matches to the literature.[36] However, there was very little overlap between the peptides identified (regardless of modification assignment) with those reported in literature, hindering our ability to assess the accuracy of the C-terminal diglycine modifications reported. Some 63% of our enriched peptides contain two or more lysine residues potentiating misassignment, with over 50% of these peptides terminating in lysine. Peptides with multiple lysine residues should be further evaluated to ensure correct assignment.

Protein identifications

Gene ontology (GO) terms annotating different datasets highlight the similarities and differences found between proteins identified before and after diglycine enrichment using our method. The cellular component of our identifications readily identified genes associated with plant-specific organelles like the chloroplast, as well as more broad areas like the cytoplasm and plasma membrane (**Figure 5.7A**), suggesting that proteins localizing to these structures and compartments are ubiquitinated, or localize to these regions following ubiquitination. Two areas represented in GO enrichment found in our unenriched dataset are the peroxisome and mitochondria. A variety of

catabolic/metabolomic processes were identified, as well as biological processes involving ATP production, protein folding, and bacterium defense (**Figure 5.7B**), supporting the notion that Ub is broadly important in the regulation of diverse facets of plant metabolism. Like all plants, Arabidopsis is sessile and must adjust *in situ* to precipitous changes in their environment such as light (both quantity and quality), water availability, temperature and pathogen exposure to name a few. Temperature response is regulated in part in a Ub-dependent manner as reflected by the identification of putative ubiquitinated cold and heat response proteins, consistent with findings in the literature.[37] Enrichment of proteins involved in the response to heavy metals such as cadmium ion are strongly represented in this dataset (**Figure 5.7C)**. The heavy metal cadmium is toxic to plants, mainly through production of reactive oxygen species.[38] Studies have shown that cadmium triggers ethylene and jasmonic acid signaling which has been shown to downregulate proteins, supporting our identification of this type of regulation in our scan for ubiquitination.[39] A second ion response reported in our GO terms was zinc, which has shown to alter the Arabidopsis transcriptome.[40] Correlating this work with proteomic investigations can deepen the knowledge of plant metabolic regulation as a whole using Arabidopsis as a specific instantiation. Consistent with previous research, the GO terms obtained suggests that ubiquitination occurs on proteins involved in a variety of functions involved in the regulation of growth, development, and response to abiotic and biotic stressors, thus emphasizing the importance of ubiquitination as a post-translational gene expression regulatory strategy in Arabidopsis.[25]

From a technical standpoint, evaluating the peptides generated prior to enrichment provided confirmation that our peptide preparations contained diglycine-modified

peptides. An unexpected observation was the detection of numerically more modified peptides in scans of unenriched preparations as compared to enriched samples. We hypothesize that more abundant diglycine-containing peptides are identified in our unenriched samples, resulting in a disproportionate number of proteins categorized as, for example, possessing 'ribulose bisphosphate carboxylase activity,' to be exclusively identified in this dataset. Since these proteins/peptides are not abundantly represented in our scans following enrichment, it is possible that these peptides have less affinity for the anti-diglycine antibodies, excluding them from detection following enrichment. Furthermore, 25% of proteins identified in the unenriched samples were identified as having six or more diglycine-modified peptides, in contrast to only ~17% identified in enriched preparations, suggesting that more abundant proteins and their corresponding peptides are found in this fraction. Taken together, the evidence illustrates the importance of analyzing samples before and after enrichment from a technical and biological perspective, allowing for a more accurate representation of the ubiquitome.

Nonspecific binding

While we were successful in identifying novel ubiquitination sites, a large number of unmodified peptides were also enriched (~900 shared in biological replicates, ~2000 per measurement), consistent with previous studies that reported 3,000–5,000 unmodified proteins to be isolated alongside 5,000 modified proteins from 1 mg of input.[34] Of the unmodified peptides found in our study, <15% of peptides contain two glycine residues, whether in tandem or dispersed within a sequence, suggesting that nonspecific binding might contribute to the high abundance of these peptides, as they are generated from the

digestion of highly abundant proteins such as RuBisCO and heat shock proteins, providing confidence in the identifications reported.

Method development/improvement

The simplicity of a protocol combining RuBisCO depletion, FASP, successive enrichments with different antibodies, coupled with DIA-MS contributes to its utility in different user-specific applications and further combination with other methods. Incorporating front-end fractionation techniques may be used to reduce the amount of peptides derived from K48- and K63-linked polyUb chains found in every sample, which might otherwise obscure isolation and detection of other low abundant peptides. A challenge in identifications is the head-to-tail repeats of Ub or the fusion of Ub with ribosomal proteins L40 and S27A,[41,42] which further add to the Ub pool. We utilized proteasome inhibition with MG132 in an attempt to reduce the degradation of ubiquitinated substrates via the 26S proteosome, with previous work reporting a ~40% decrease in identifications without this treatment.[34] This suggests that large shifts in the dataset can be observed under different protocols and conditions. DUB inhibitors, such as 2,6-diaminopyradine-3,5-bis(thiocynate) (PR-619), can also be used either alone or in conjunction with MG132. Ultimately, in order to provide the biological context within which a substrate is ubiquitinated, it is desirable to also determine the class of Ub modification (i.e., mono- or polyubiquitination, and linkage type) as well as the Ub ligase and Ub conjugating enzyme responsible for such modifications. This biological context is lacking for a large majority of ubiquitomes across species. The chain length of ubiquitinated proteins has been measured using the polyUb chain protection from trypsin (Ub-ProT) system, preserving the ubiquitin moieties on substrates through the addition of

ubiquitin binding domains to mask sites during proteolysis.[43] To further address the success and specificity of the RuBisCO depletion method implemented,[25,27] the precipitated material formed following PS addition should be resolubilized and prepared for MS analysis to identify proteins removed in this fraction. This will ensure the proteins available for enrichment are unbiased, and the analyses performed sampled an accurate snapshot of the Arabidopsis proteome as a whole.

A factor left unaddressed in our investigation was the existence of UBL modifications. Previous reports have shown that modifications such as NEDDylation accounts for a low percentage of diglycine modifications[44]. As a consequence, we refrained from alterations to our protocol to distinguish between Ub and UBL modifications. An alternative approach employs different proteases (such as LysC) to generate a UBL remnant distinct from the Ub remnant, and enriches for modified peptides using a different antibody, such as the UbiSite antibody, raised against the last 13 amino acids of ubiquitin.[20] In order to determine Ub's role in plant function, the Arabidopsis ubiquitome can be assessed under various stress conditions, such as cold/heat stress, drought, and different phytohormone stimuli, and different genetic backgrounds, where Arabidopsis mutants are available,[45] linking the ubiquitination machinery to the modified site and indicate sets of ligases for a particular target, determining its influence on the plant as a whole. The method described here can be employed to study changes to the Arabidopsis ubiquitome under these various conditions or loss of function mutants of the SCF class of E3 Ub ligases. Further, the interplay of ubiquitin and other PTMs may serve as an important regulator of ubiquitination; therefore, future work should examine how these PTMs are affected

alongside with ubiquitination, perhaps beginning with acetylation as a known effector with overlapping substrate specificity as ubiquitin.[46,47]

Development of a unique TAIR10 database

Following data acquisition and processing, we observed a large number of duplicate identifications due to the redundancies inherent to the TAIR10 database; therefore, we removed 200 duplicates from the database lessening it from 27,416 to 27,216 proteins. More improvements, including protein grouping, can be performed to further trim the database while still representing all of the peptides belonging to the proteome. This is especially important for this type of dataset, where ~40% of proteins identified only have one or two diglycine-modified peptides identified (**Figure 5.3**). Since only a few peptides per protein were identified, it can be difficult to ascribe the protein from which a particular peptide was derived, with protein paralogs and splice variants sharing high sequence identity. Utilizing such a trimmed database would serve to decrease processing time, reduce the number of duplicates observed and create a more streamlined workflow when analyzing data. Adjustments to the curated Arabidopsis proteomes must continue to ensure that global comparison of datasets between research groups is consistent.

From the perspective of MS measurements, our protocol can be used in conjunction with DDA methods, including our custom TAIR10 database. Analyses using both DIA and DDA methods will allow for monitoring ubiquitination events of specific targets, linking the ubiquitination machinery to the modified site and determining its functional contribution to plant form and function.

CONCLUSION

Arabidopsis thaliana is an important model plant species for understanding the genetic and molecular basis of plant form and function, with parallel implications and impacts to species important in agriculture. Genetically, the large abundance of proteins involved in the UPS suggests that a significant component of the Arabidopsis proteome may be regulated via post-translational modification involving ubiquitination. However, in comparison to other model organisms, only a modest set of ubiquitinated proteins have thus far been identified. Most proteomic endeavors use data-dependent forms of data acquisition. To expand the ubiquitome of Arabidopsis, and to provide new tools to investigate it, we have applied techniques for RuBisCO depletion and efficient trypsin digestion, followed by data-independent acquisition mass spectrometry methods, revealing ~100 novel ubiquitinated proteins and over 400 novel ubiqutination sites in Arabidopsis. This toolset can be added and applied to Arabidopsis ubiquitome research, ultimately leading to greater insight into the role of ubiquitination to plant metabolism, patterning and development.

REFERENCES

1. Initiative, T. A. G. Analysis of the genome sequence of Arabidopsis thaliana. *Nature* **408**, 796–815 (2000).
2. Vaneechoutte, D., Estrada, A. R., Lin, Y. C., Loraine, A. E. & Vandepoele, K. Genome-wide characterization of differential transcript usage in Arabidopsis thaliana. *Plant J.* **92**, 1218–1231 (2017).
3. Gupta, R. *et al.* Time to dig deep into the plant proteome: A hunt for low-abundance proteins. *Front. Plant Sci.* **6**, 1–3 (2015).
4. Zhang, H. *et al.* Arabidopsis proteome and the mass spectral assay library. *Sci. Data* **6**, 278 (2019).
5. Cheng, C. Y. *et al.* Araport11: a complete reannotation of the Arabidopsis thaliana reference genome. *Plant J.* **89**, 789–804 (2017).
6. Gagne, J. M., Downes, B. P., Shiu, S. H., Durski, A. M. & Vierstra, R. D. The F-box subunit of the SCF E3 complex is encoded by a diverse superfamily of genes in Arabidopsis. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 11519–11524 (2002).
7. Feng, B., Li, L., Zhou, X., Stanley, B. & Ma, H. Analysis of the Arabidopsis floral proteome: Detection of over 2000 proteins and evidence for posttranslational modifications. *J. Integr. Plant Biol.* **51**, 207–223 (2009).
8. Vierstra, R. D. The ubiquitin-26S proteasome system at the nexus of plant biology. *Nat. Rev. Mol. Cell Biol.* **10**, 385–397 (2009).
9. Wei Li., Mario H. Bengtson, Axel Ulbrich, Akio Matsuda, Venkateshwar A. Reddy, Anthony Orth, Sumit K. Chanda, Serge Batalov, C. A. P. J. Genome-Wide and Functional Annotation of Human E3 Ubiquitin Ligases Identifies MULAN, a Mitochondrial E3 that Regulates the Organelle's Dynamics and Signaling. *PLoS One* e1487 (2008) doi:10.1371/journal.-pone.0001487.
10. Del Pozo, J. C. & Manzano, C. Auxin and the ubiquitin pathway. Two players-one target: The cell cycle in action. *J. Exp. Bot.* **65**, 2617–2632 (2014).
11. Schumann, N., Navarro-Quezada, A., Ullrich, K., Kuhl, C. & Quint, M. Molecular evolution and selection patterns of plant f-box proteins with C-terminal kelch repeats. *Plant Physiol.* **155**, 835–850 (2011).
12. Santner, A., Calderon-Villalobos, L. I. A. & Estelle, M. Plant hormones are versatile chemical regulators of plant growth. *Nat. Chem. Biol.* **5**, 301–307 (2009).
13. Sharma, B., Joshi, D., Yadav, P. K., Gupta, A. K. & Bhatt, T. K. Role of ubiquitin-mediated degradation system in plant biology. *Front. Plant Sci.* **7**, 1–8 (2016).
14. Park, C., Lim, C. W., Baek, W. & Lee, S. C. RING Type E3 Ligase CaAIR1 in Pepper Acts in the Regulation of ABA Signaling and Drought Stress Response. *Plant Cell Physiol.* **56**, 1808–1819 (2014).
15. Achard, P. *et al.* DELLAs contribute to plant photomorphogenesis. *Plant Physiol.* **143**, 1163–1172 (2007).

16. Zuin, A., Isasa, M. & Crosas, B. Ubiquitin Signaling: Extreme Conservation as a Source of Diversity. *Cells* **3**, 690–701 (2014).

17. Meister, M. *et al.* Regulation of cargo transfer between ESCRT-0 and ESCRT-I complexes by flotillin-1 during endosomal sorting of ubiquitinated cargo. *Oncogenesis* **6**, (2017).

18. Xie, H. *et al.* Global Ubiquitome Profiling Revealed the Roles of Ubiquitinated Proteins in Metabolic Pathways of Tea Leaves in Responding to Drought Stress. *Sci. Rep.* **9**, 1–12 (2019).

19. Gillet, L. C. *et al.* Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: A new concept for consistent and accurate proteome analysis. *Mol. Cell. Proteomics* **11**, 1–17 (2012).

20. Akimov, V. *et al.* Ubisite approach for comprehensive mapping of lysine and n-terminal ubiquitination sites. *Nat. Struct. Mol. Biol.* **25**, (2018).

21. Walton, A. *et al.* It's time for some "site"-seeing: Novel tools to monitor the ubiquitin landscape in Arabidopsis thaliana. *Plant Cell* **28**, 6–16 (2016).

22. Park, D., Goh, C. J., Kim, H., Lee, J. S. & Hahn, Y. Loss of conserved ubiquitylation sites in conserved proteins during human evolution. *Int. J. Mol. Med.* **42**, 1827–1836 (2018).

23. Xie, X., Kang, H., Liu, W. & Wang, G. L. Comprehensive profiling of the rice ubiquitome reveals the significance of lysine ubiquitination in young leaves. *J. Proteome Res.* **14**, 2017–2025 (2015).

24. Ritchie, M., Langridge, J., Mckenna, T. & Mock, S. A. H. Proteome analysis of cold stress response in Arabidopsis thaliana :a label-free qunatitative proteomic study. *Waters Corp. Poster* (2005).

25. Kim, Y. J. *et al.* Depletion of abundant plant RuBisCO protein using the protamine sulfate precipitation method. *Proteomics* **13**, 2176–2179 (2013).

26. Wiśniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nat. Methods* **6**, 359–362 (2009).

27. Kim, Y. J. *et al.* Protamine sulfate precipitation method depletes abundant plant seed-storage proteins: A case study on legume plants. *Proteomics* **15**, 1760–1764 (2015).

28. Kim, D. Y., Scalf, M., Smith, L. M. & Vierstra, R. D. Advanced proteomic analyses yield a deep catalog of ubiquitylation targets in Arabidopsis. *Plant Cell* **25**, 1523–1540 (2013).

29. Walton, A. *et al.* It's time for some "site"-seeing: Novel tools to monitor the ubiquitin landscape in Arabidopsis thaliana. *Plant Cell* **28**, 6–16 (2016).

30. Chen, X. L. *et al.* Proteomic analysis of ubiquitinated proteins in rice (Oryza sativa) after treatment with pathogen-associated molecular pattern (PAMP) elicitors. *Front. Plant Sci.* **9**, 1–16 (2018).

31. Lu, J. *et al.* Proteome and ubiquitome changes during rose petal senescence. *Int. J.*

Mol. Sci. **20**, 8–10 (2019).

32. Nickerson, J. L. & Doucette, A. A. Rapid and Quantitative Protein Precipitation for Proteome Analysis by Mass Spectrometry. *J. Proteome Res.* **19**, 2035–2042 (2020).

33. Uhrig, R. G., She, Y. M., Leach, C. A. & Plaxton, W. C. Regulatory monoubiquitination of phosphoenolpyruvate carboxylase in germinating castor oil seeds. *J. Biol. Chem.* **283**, 29650–29657 (2008).

34. Udeshi, N. D. *et al.* Refined preparation and use of anti-diglycine remnant (k-ε-gg) antibody enables routine quantification of 10,000s of ubiquitination sites in single proteomics experiments. *Mol. Cell. Proteomics* **12**, 825–831 (2013).

35. Xu, G., Paige, J. S. & Jaffrey, S. R. Global analysis of lysine ubiquitination by ubiquitin remnant immunoaffinity profiling. *Nat. Biotechnol.* **28**, 868–873 (2010).

36. Na, C. H. *et al.* Synaptic protein ubiquitination in rat brain revealed by antibody-based ubiquitome analysis. *J. Proteome Res.* **11**, 4722–4732 (2012).

37. Jacob, P., Hirt, H. & Bendahmane, A. The heat-shock protein/chaperone network and multiple stress resistance. *Plant Biotechnol. J.* **15**, 405–414 (2017).

38. Dalcorso, G., Farinati, S. & Furini, A. Regulatory networks of cadmium stress in plants. *Plant Signal. Behav.* **5**, 663–667 (2010).

39. Wang, T. *et al.* Mechanism enhancing arabidopsis resistance to cadmium: The role of NRT1.5 and proton pump. *Front. Plant Sci.* **871**, 1–12 (2018).

40. Landa, P. *et al.* The Transcriptomic Response of Arabidopsis thaliana to Zinc Oxide: A Comparison of the Impact of Nanoparticle, Bulk, and Ionic Zinc. *Environ. Sci. Technol.* **49**, 14537–14545 (2015).

41. Archibald, J. M., Teh, E. M. & Keeling, P. J. Novel ubiquitin fusion proteins: Ribosomal protein P1 and actin. *J. Mol. Biol.* **328**, 771–778 (2003).

42. Callis, J., Raasch, J. A. & Vierstra, R. D. Ubiquitin Extension Proteins of Arabidopsis thaliana. *J. Biol. Chem.* **265**, 12486–12493 (1990).

43. Tsuchiya, H. *et al.* Ub-ProT reveals global length and composition of protein ubiquitylation in cells. *Nat. Commun.* **9**, (2018).

44. Kim, W. *et al.* Systematic and quantitative assessment of the ubiquitin-modified proteome. *Mol. Cell* **44**, 325–340 (2011).

45. Malley, R. C. O., Barragan, C. C. & Ecker, J. R. A User's guide to the Arabidopsis T-DNA insertional mutant collections. *Methods Mol. Biol.* **1284**, 323–342 (2015).

46. Danielsen, J. M. R. *et al.* Mass spectrometric analysis of lysine ubiquitylation reveals promiscuity at site level. *Mol. Cell. Proteomics* **10**, 1–12 (2011).

47. Swaney, D.L., Beltrao, P., Starita, L., Guo, A., Rush, J., Fields, S., Krogan, N.J. and Villén, J.Global analysis of phosphorylation and ubiquitylation crosstalk in protein degradation. *Nat. Methods* **10**, 1–20 (2013).

CHAPTER 6

CONCLUSION AND FUTURE DIRECTIONS

Identification and characterization of TAZ phosphorylation

Our investigation of TAZ using both untargeted and untargeted-based MS methods identified numerous novel phosphorylation sites. Implementation of a targeted approach to monitor the phosphorylation status of a new site, Ser^{93}, a site proximal to the critical 14-3-3 binding site Ser^{89} allowed for the accurate assignment of positional phosphoisomers. Furthermore, we found that different serum conditions influenced $pSer^{93}$ levels, which hinders TAZ–14-3-3 binding, potentiating the nuclear localization of TAZ. This work shows that phosphorylation sites proximal to 14-3-3 binding sites can regulate the binding of 14-3-3 to its targets, adding another layer of phosphorylation-dependent regulation to these proteins, as is the case with the TAZ paralog YAP, where phosphorylation of Ser^{128} hinders access of 14-3-3 to its binding site on YAP.[1] Our method also allows for the accurate identification of positional phosphoisomers on peptides, which can be applied to other numerous proteins of interest that are regulated by 14-3-3, aiding in the understanding of their processes and regulation.

The kinase(s) and phosphatase(s) responsible for all the novel sites identified, especially Ser^{93}, should be elucidated, which would add to our understanding of the biological cues leading to these phosphorylation events and how $pSer^{93}$ interferes with protein activation through 14-3-3 binding. TAZ activation through these phosphorylation events may be involved in the signals allowing for TAZ-associated diseases to progress. The phosphorylation sites identified in this investigation are located throughout the protein, possibly regulating the protein's structure and altering protein-protein interactions and localization. The monitoring of TAZ phosphorylation in different cellular events (like the epithelial-to-mesenchymal transition (EMT)) and cell migration, coupled with

transcription factor binding/gene activation and the upstream signals inducing differential phosphorylation, can offer a full picture of the cellular processes influenced by different TAZ phosphoforms. Such monitoring can be achieved through expression of TAZ in various cell lines appropriate for EMT or cellular process TAZ is involved in, followed by immunoprecipitation. The co-enriched proteins captured can be identified by MS, and the phosphorylation status of the novel TAZ phosphorylation sites can be determined using our targeted our approach. TAZ localization has been reported to be regulated through a nuclear export sequence in the TEAD binding domain and a nuclear localization signal in the C-terminal portion of the protein.[2] We have identified three novel phosphorylation sites in TAZ in the TEAD binding domain and one near the extreme C-terminus. Mutagenesis of these sites, followed by immunofluorescence studies, should be performed to examine the potential link between these phosphorylation events and TAZ localization.

The ubiquitin proteasome system (UPS) is known to down regulate YAP/TAZ through phosphorylation-dependent protein degradation.[3] The ubiquitination site, however, has not been reported. Investigating the ubiquitination site can be achieved by using diglycine scanning, as performed in our *Arabidopsis thaliana* ubiquitome work. The enrichment and identification of TAZ-derived diglycine-modified peptides would pinpoint the location of the ubiquitination site(s), which may even be conserved with YAP. The redundancy between YAP/TAZ has been explored, with the two proteins sharing 60% similarity.[4] A difference observed between the two is the lower stability of the TAZ protein in U2OS cells.[5,6] Unlike YAP, TAZ has been reported to have both an N- and C-terminal phosphodegron,[3,7] perhaps accounting for the TAZ protein's reduced stability. Like TAZ, many proteomic efforts have examined the phosphorylation of YAP, reporting multiple

phosphorylation sites spread throughout the protein.[8,9] Comparing the phosphorylation landscapes of YAP and TAZ will likely provide mechanistic insights into the redundant functions of these two proteins and their participation in regulating various diseases.

The power of high-resolution mass spectrometry in DIA coupled with DDA modes has expanded the depth of knowledge of the phosphorylation-dependent regulation of TAZ and has provided tools for future investigation of the cell regulation through the transcriptional co-activator.

Quantitative analysis of EV/exosomal proteins derived from SMA affected cells

Our investigation of exosomes from SMA patient-derived cell lines yielded identification and quantitation of ~680 proteins. From these identifications, ~100 were found to be statistically significant, with 18 proteins showing an increase and the remaining proteins showing a decrease in SMA compared to age-matched controls. A majority of our dataset identified exosomal markers, as well as previously identified markers for SMA, such as Plastin-3 (PLS3). A majority of our significant identifications have not been reported in any SMA biomarker repositories. One protein identified, the insulin-like growth factor binding protein 3 (IGFBP3), was validated in patient samples. Further work will monitor the levels of our identified candidate biomarkers in patient blood to determine whether our observations remain consistent, and whether these candidate biomarkers can be used as *bona fide* biomarkers for SMA. The dataset identified multiple T-complex proteins (TCP), which serve as chaperones for protein folding. Interestingly, TCP1 has been shown to be important in the localization of the antisense oligonucleotide treatment Spinraza®.[10,11] TCP1 levels should continually be monitored throughout Spinraza® treatment to assess treatment efficacy.

Our proteomic analysis is lacking a thorough investigation of the post-translational modifications to which the isolated proteins can be subjected. We scanned our dataset for the modifications of phosphorylation and ubiquitination finding a small number of modification sites, many belonging to highly abundant proteins (such as actin, tubulin, and keratin) and IGFBP3, one of the most abundant proteins found in our study. We identified phosphorylation of IGFBP3 Ser^{147}, consistent with the literature, which has been found to be catalysed by the kinase FAM20C.[12] We also identified novel phosphorylation sites in fewer biological replicates, requiring validation and further investigation into the function of these sites. These novel sites can be examined by applying our targeted MS approach utilized on TAZ to determine whether IGFBP3 phosphorylation status varies in SMA or control cell-derived exosomes. Other PTMs have been associated with exosomal proteins, such as acetylation and glycosylation.[13] The interplay between these modifications should be investigated using modification-specific enrichments followed by MS analysis.

Performing our method on SMA patients with modifier gene mutations, such as mutations in the ubiquitin-activating enzyme 1 (UBA1), can help us determine whether mutation-specific programs of disease onset exist.[14] Furthermore, the techniques described here can be implemented to investigate SMA in gender-specific backgrounds; researchers showed that women with SMA experienced greater severity in their inability to walk as compared to men.[15] PLS3 has been postulated as an SMA gender-specific biomarker, with cell lines derived from asymptomatic females from a SMA-afflicted family showing an increase in PLS3 levels.[16,17] Our methods can also be extended to study different types of SMA and to identify modifiers specific to male and females, as well as those specific to ethnic traits. The proteomic composition of exosomes derived from the different SMA

types (I–IV) may also allow for SMA type-specific biomarker discovery, aiding in early disease detection. A significant effort in the field has focussed on characterizing the RNA and lipid content of exosomes,[18] isolating the RNA, and analyzing the lipid content of these exosomes; alongside our quantitative proteomics analyses, these efforts would provide a complete representation of the disease, using existing protocols which have investigated the blood transcriptome from SMA patients.[19]

As exosomes are able to retain the molecular characteristics of the cells from which they are derived, exosomes are a suitable platform to identify biomarkers for other neurological diseases, such as amyotrophic lateral sclerosis, which currently lacks robust biomarkers.[20] Investigating the proteome of exosomes derived from different multiple biological fluids (e.g., blood, saliva, and urine) and identifying fluid-specific alterations in SMA backgrounds would greatly complement our work. We expect that there would be alterations in the proteins identified given the fluids' exposure to different cell types.[21] Coupled with the therapeutic interventions currently available, use of this robust and non-invasive method of protein identification and quantitation can serve as a method of disease detection and prognosis serving as a beacon of hope for those afflicted with the disease.

Analysis of the *Arabidopsis thaliana* ubiquitome

Analysis of the Arabidopsis ubiquitome undertaken in this work has expanded the list of candidate proteins and the respective ubiquitination sites, identifying 398 novel ubiquitination sites. Even with this expanded ubiquitome catalogue a discrepancy exists between the large number of components of the UPS encoded in the Arabidopsis genome, and the current list of known or candidate ubiquitinated proteins. Ultimately, the details of each ubiquitination event are critical in understanding the role of specific ubiquitination

events in plant patterning and development. The work presented offers a rapid and affordable method to study ubiquitomes, with its strength lying in its ability to be coupled with numerous other techniques based on user requirements. Many studies have utilized sequential protein digestions using GluC or LysC, in combination with trypsin, creating different subset of peptides. One specific advantage of using LysC is that distinct signatures are left on peptides generated, allowing one to distinguish between ubiquitination and ubiquitin-like modifications (which leaves ESTLHLVLRLRGG on modified peptides) distinguishing ubiquitin from other ubiquitin-like modifications.[22,23] Surfactants, such as RapiGest™, have also been used to help further denature proteins, allowing for more efficient digestion.[24,25] Employing fractionation methods before digestion and enrichment has also been shown to reduce sample complexity by aiding in digestion. Although this added method is time consuming and increases sample handling, it is amenable to our protocol,[24,26,27] and could prove to be beneficial when honing in on specific targets which can be masked in areas rich in highly abundant proteins/peptides. In attempts to preserve protein ubiquitination plants were treated with MG132 prior to extraction. Groups have found alterations in the ubiquitomes studied using different proteasome inhibitors, such as bortezomib, which would be an interesting stimulus.[28,29] A component lacking in our analyses is the use of deubiquitinase inhibitors, which would help prevent the removal of ubiquitin modifications that might otherwise result in the loss of substrates identified.

Proteins, nor their modifications, do not work in isolation. An overlap between acetylation and ubiquitination has been shown, with the two modifications capable of occurring on the same lysine residue, effectively acting to block one another.[30] Phosphorylation is a prevalent modification in Arabidopsis, with over 14,000 proteins

being shown to be phosphorylated.[31] A subset of E3-ubiquitin ligases utilizes phosphorylation as a signal for ubiquitination, such as the F-box proteins, demonstrating that crosstalk between these two PTMs is an important form of Arabidopsis regulation, as is the case with CONSTANS transcription factor which regulates flowering. Under conditions of low light, CONSTANS is phosphorylated allowing the E3 ligase complex CONSTITUTIVE PHOTOMORPHOGENIC 1 to recognize and ubiquitinate phosphorylated CONSTANS, resulting in its degradation.[32] Many proteomics endeavours exclude undesired PTMs due to the increased complexity and added processing times; however, a more accurate depiction of the proteomes in question would be strengthened by these measurements, particularly when the plant is exposed to different conditions.

A majority of our dataset was unique, possibly owed to the use of DIA acquisition methods, differing from most previous ubiquitome studies. Analyzing preparations using our protocol, followed by DDA workflows, will help determine whether the uniqueness in our hits is inherent to the protocol or the MS methodologies applied. As DDA and DIA methods have shown their own separate strengths and weaknesses, a more comprehensive approach may require measurements performed in both modes, where it is expected that DDA methods would identify the more abundant peptides/proteins and DIA methods finding the less abundant species. DDA methods could also be applied to investigate the unassigned ubiquitination sites identified in our study, given that DDA methods are known to provide more accurate site assignments than DIA methods, which has shown strength in identifying modified peptides.[33]

While characterizing the ubiquitome of Arabidopsis alone is a daunting task, providing the biological context within which such ubiquitination events occurs would help

explain the overrepresentation of genes involved in the UPS in plant genomes. Investigating the Arabidopsis ubiquitome at different ages or under different stressors will be a critical facet to investigate as the UPS allows for rapid proteomic alteration, ultimately influencing the fate of the organism as a whole.[34] The diglycine-scanning method created can also be applied to numerous transgenic mutagenic Arabidopsis lines that have been produced.[35] Comparison of the ubiquitome datasets obtained from mutant and wild-type genetic backgrounds can implicate the involvement of an E3 in the modification of a specific target or target set. As an example, mutant backgrounds of the phytohormone auxin, which is critical for nearly every facet of plant growth and regulation and tied to the UPS, should be investigated. These mutants display critical plant-specific alterations like leaf and flowering morphology as a result of mutagenesis of auxin resistance genes. How the Ub landscape and activity of SCF^{TIR1} complexes are shifted by these mutations can be evaluated by the methods presented in this work.[36]

Future work may focus less on discovery and more on targeting specific ubiquitination events and cascades. Pathways that should be heavily investigated include auxin and jasmonate signaling, critical for regulation of root development,[37] as well as flowering, which shows gross changes to protein expression in order to produce reproductive organs.[38] An aspect of understanding the ubiquitome requires knowledge of the linkage types and possible branching of ubiquitin moieties or solving the Ub code. The ubiquitin protection from trypsin (Ub-ProT) method maintains the Ub linkage on modified peptides, followed by SDS-PAGE separation for analysis.[23] The protocol should be modified to allow for global MS analysis with these larger modifications, preserving ubiquitin not just diglycine. Understanding individual ubiquitination events may require

elucidating the E3 ligases responsible for such modifications. This appears to be important as 5% of the Arabidopsis genome encodes for E3 proteins.[39] Identifying a substrate's cognate E3 Ub ligase(s) is paramount and is a challenge that is not unique to plant proteomics. From our dataset, a very interesting E3 to identify would be the E3 that catalyzes ubiquitination of Lys^{628} of phosphoenolpyruvate carboxylase (PEPC), whose ubiquitination is conserved throughout plants and which regulates the protein's catalytic activity to allow for oxaloacetate production.[40,41] While this ubiquitination site has been known for years, the E3 responsible remains unknown. Strategies to investigate E3–substrate interactions may require *in vivo* labeling followed by MS analysis, such as the BioID approach implemented with MTMR2; however, BioID work possesses inherent challenges in plant systems,[42] requiring method development, such as the preclearing of bait proteins through FLAG-IP, as we employed in our MTMR2 BioID interactome study. Through the use of modern proteomics techniques in conjunction with the methods we have developed here, the Arabidopsis ubiquitome can be solved, which will shed light on the contribution of protein ubiquitination in plant function, patterning, and development.

Reversible phosphorylation of the lipid phosphatase MTMR2

In order to better understand the interactome of the lipid phosphatase MTMR2, and its link to myelination, we developed methods that led us to discover new avenues of MTMR2 regulation. The calyculin A-induced degradation of MTMR2 in Schwann cells suggests a role of phosphorylation in regulating MTMR2 stability, either through activation or inactivation of MTMR2 itself or its effectors. The details of this mechanism still require further to properly conclude this aspect of this work. One way to understand this phenomenon is to analyze the stability of endogenous MTMR2 protein in Schwann cells.

Ascertaining the half-life and stability of MTMR2 is crucial to understanding how the protein is cleared. We attempted to perform cycloheximide experiments to determine MTMR2 stability; however, we encountered difficulties with detecting MTMR2 using our methods, likely arising from using too few cells in comparison to the number of cells used in our overexpression studies (data not shown). Our phosphatase inhibitor approach can be coupled with immunoblotting to examining the levels of endogenous MTMR2, which will confirm the validity of our results, ensuring that what we observed occurs endogenously, and is not an artifact of overexpression. While investigating this MTMR2 downregulation, we explored the idea of proteasome-dependent degradation of MTMR2, for which we used the proteasome inhibitor MG132. While we did not observe a rescue of MTMR2 levels when cotreated with calyculin A and MG132, we observed an upregulation of MTMR2 upon MG132 treatment alone. MG132 is known to display non-proteasomal effects, such as activating p38 MAPKs and caspases,[43] and upregulating levels of certain proteins.[44] It is possible that MG132-induced activation of these pathways leads to an increase in transcription factor activity, in turn leading to an increase in MTMR2 protein levels. The stability of the inactive myotubularins that interact with MTMR2, MTMR5 and MTMR13, need to be monitored following calyculin A treatment using immunoblotting. This will help determine whether the calyculin A-induced downregulation of MTMR2 influences levels of the inactive phosphatase, leading to a decrease in MTMR2. Likewise, MTMR5/13 levels should be measured following MG132 treatment to assess whether MTMR2 stability following MG132 treatment is dependent on these inactive binding partners.

We postulate the mobility shift we reproducibly observe during calyculin A treatment is due to increases in phosphorylation, suggesting MTMR2 is phosphorylated at

a site that has not yet been reported, based on treatment of Ser^{58}/Ser^{631} phosphodeficient mutants. The purpose of this unidentified site, as well as the identity of the kinase(s) and phosphatase(s) which influence it, is important, with modifications potentially occurring on one or more of the Ser^{65} or Thr^{29} residues present in MTMR2. These phosphorylation events need to be examined for their influence on MTMR2 function, stability, and structure, which may all play a role in MTMR2 protein-protein interactions. Future work can employ phosphopeptide enrichment or targeted MS approaches to measure the phosphorylation of MTMR2 as a whole and map novel phosphosites given that more sites have not been identified in the past. Our group struggled in measuring the phosphorylation status of Ser^{631} due to low detection of the tryptic fragment containing this site. Employing targeted MS analysis against this peptide, as well as other potential sites, will allow for the detection of new sites; peptides with the predicted *m/z* of the peptides housing these sites can be selected for in a targeted MS approach, a form of parallel reaction monitoring that allows for stringent phosphosite mapping as well as quantitation, which was used in our TAZ analyses.[45] By coupling this targeted MS approach with calyculin A-treated MTMR2 digests, we may be able to identify the site responsible for the mobility shift and its suspected degradation.

The phosphatase responsible for Ser^{58} dephosphorylation remains elusive following our AP-MS and BioID strategies. Expressing tagged serine/threonine phosphatases can be used; however, compensation between PP1 and PP2 has been shown.[46] Alternatively, substrate trapping has been used to study PTP-substrate interactions. Since PSPs function through a one-step method without intermediate formation, PSP-substrate trapping requires a method different than the traditional form employed for PTPs.[47] PSP-substrate trapping

involves forming hypoactive PSP holoenzymes using a mutant form of the catalytic subunit, which is inactive but retains its ability to bind with its regulatory subunit. This hypoactive holoenzyme can associate with phosphosubstrates allowing for its co-enrichment and identification.[48] This may allow for the identification of the specific phosphatase responsible for MTMR2 dephosphorylation. While our BioID scans did not yield a Ser[58] phosphatase, we enriched EIPR-1/TSSC1, a protein important for regulating retrograde transport. We postulate that MTMR2 regulates EIPR-1 in a manner similar to MTMR2 regulation of RME-8, which is by regulating the PI3P levels on which EIPR-1 may be dependent. It would be interesting to monitor the activity and levels of EIPR-1 in CMT models. If altered, this would suggest that MTMR2 plays a large role in EIPR-1 cargo shuttling in CMT. This method can be used to investigate MTMR2's interactome throughout the myelination process in Schwann cell cocultures to observe whether different endocytic effectors are enriched through the myelination program.

Merging our BioID vectors with established mouse models lacking MTMR2 (MTMR2$^{-/-}$) or MTMR13 (*Mtmr13*$^{-/-}$ mouse) SC-Doral root ganglion (DRG) explants, which show a decrease in MTMR2 levels.[49] These explants display myelin outfoldings, consistent with CMT pathology. Expressing our BioID vectors in these MTMR2$^{-/-}$ mice models may identify interactors that participate in CMT onset and myelination. Previous studies have shown that myelin outfoldings of Mtmr2-null mice could be rescued by MTMR2 replacement, supporting the feasibility of our BioID vectors in such studies.[50] A specific pathway that should be monitored is the ERK1/2 pathway responsible for MTMR2 phosphorylation and is involved in the dedifferentiation of Schwann cells.[51,52,53,54]

Clinically, gene therapy for CMT patients involves the use of adeno-associated virus to transport neurotrophic factor for CMT subtypes 1A, 1X, and 4C2.[55] These gene therapies have commonly addressed neurological diseases caused by gene duplications, as in CMT1A. The suitability of employing antisense oligonucleotides should be examined, which could potentially restore proteins in CMT types characterized by decreases in specific proteins, as is the case with Spinraza® and its treatment of SMA by allowing production of full-length SMN protein.[56] Using these oligonucleotides as a way to correct for common *mtmr2* mutations in CMT4B may provide a future treatment plan. MTMR2 itself has also been use as a treatment for some subtypes of CMT in which MTMR2 is delivered intravenously to compensate for defects in MTM or other MTMR proteins; such clinical relevance highlights the importance of studying this protein.[57,58] Our findings and toolset created may serve as a platform to further guide MTMR2 studies, which should focus on the reversible phosphorylation of the protein, its stability, and its interaction with other endocytic proteins, all of which may have wide implications for CMT disease onset and the myelin program.

In conclusion, the use of high-resolution MS will remain a mainstay in the all fields of science. Future developments in equipment, methodology, and preparative techniques, coupled with deepened molecular biology, will allow for the users to address their scientific needs while contributing novel and critical data to the field.

REFERENCES

1. Moon, S. *et al.* Phosphorylation by NLK inhibits YAP -14-3-3-interactions and induces its nuclear localization. *EMBO Rep.* **18**, 61–71 (2017).
2. Kofler, M. *et al.* Mediated nuclear import and export of TAZ and the underlying molecular requirements. *Nat. Commun.* **9**, (2018).
3. Huang, W. *et al.* The N-terminal phosphodegron targets TAZ/WWTR1 protein for SCF β-TrCP-dependent degradation in response to phosphatidylinositol 3-kinase inhibition. *J. Biol. Chem.* **287**, 26245–26253 (2012).
4. Plouffe, S. W. *et al.* The Hippo pathway effector proteins YAP and TAZ have both distinct and overlapping functions in the cell. *J. Biol. Chem.* **293**, 11230–11240 (2018).
5. Finch-Edmondson, M. L. *et al.* TAZ protein accumulation is negatively regulated by YAP abundance in mammalian cells. *J. Biol. Chem.* **290**, 27928–27938 (2015).
6. Vigneron, A. M., Ludwig, R. L. & Vousden, K. H. Cytoplasmic ASPP1 inhibits apoptosis through the control of YAP. *Genes Dev.* **24**, 2430–2439 (2010).
7. Liu, C. Y. *et al.* The hippo tumor pathway promotes TAZ degradation by phosphorylating a phosphodegron and recruiting the SCFβ-TrCP E3 ligase. *J. Biol. Chem.* **285**, 37159–37169 (2010).
8. Hu, J. K. H. *et al.* An FAK-YAP-mTOR Signaling Axis Regulates Stem Cell-Based Tissue Renewal in Mice. *Cell Stem Cell* **21**, 91-106.e6 (2017).
9. Zhou, H. *et al.* Toward a comprehensive characterization of a human cancer cell phosphoproteome. *J. Proteome Res.* **12**, 260–271 (2013).
10. Li, Q. Nusinersen as a therapeutic agent for spinal muscular atrophy. *Yonsei Med. J.* **61**, 273–283 (2020).
11. Liang, X. H., Shen, W., Sun, H., Prakash, T. P. & Crooke, S. T. TCP1 complex proteins interact with phosphorothioate oligonucleotides and can co-localize in oligonucleotide-induced nuclear bodies in mammalian cells. *Nucleic Acids Res.* **42**, 7819–7832 (2014).
12. Tagliabracci, V. S. *et al.* A single kinase generates the majority of the secreted phosphoproteome. *Cell.* **161**, 1619–1632 (2015).
13. Moreno-Gonzalo, O., Villarroya-Beltri, C. & Sánchez-Madrid, F. Post-translational modifications of exosomal proteins. *Front. Immunol.* **5**, 1–7 (2014).
14. Groen, E. J. N. & Gillingwater, T. H. UBA1: At the Crossroads of Ubiquitin Homeostasis and Neurodegeneration. *Trends Mol. Med.* **21**, 622–632 (2015).
15. Ar Rochmah, M. *et al.* Gender effects on the clinical phenotype in Japanese patients with spinal muscular atrophy. *Kobe J. Med. Sci.* **63**, E41–E44 (2017).
16. Oprea, Gabriela E., Sandra Kröber, Michelle L. McWhorter, Wilfried Rossoll, Stefan Müller, Michael Krawczak, Gary J. Bassell, Christine E. Beattie, and B. W. Plastin 3 Is a Protective Modifier of Autosomal Recessive Spinal Muscular Atrophy.

Science. **320**, 524–527 (2008).

17. Nishio, H. PLS3 expression and SMA phenotype: A commentary on correlation of PLS3 expression with disease severity in children with spinal muscular atrophy. *J. Hum. Genet.* **59**, 64–65 (2014).

18. Burns, D. T. *et al.* Variants in EXOSC9 Disrupt the RNA Exosome and Result in Cerebellar Atrophy with Spinal Motor Neuronopathy. *Am. J. Hum. Genet.* **102**, 858–873 (2018).

19. Siranosian, J. J. *et al.* Whole-blood dysregulation of actin-cytoskeleton pathway in adult spinal muscular atrophy patients. *Ann. Clin. Transl. Neurol.* 1–8 (2020) doi:10.1002/acn3.51092.

20. Verber, N. S. *et al.* Biomarkers in motor neuron disease: A state of the art review. *Front. Neurol.* **10**, 1–28 (2019).

21. Zabeo, D. *et al.* Exosomes purified from a single cell type have diverse morphology. *J. Extracell. Vesicles* **6**, (2017).

22. Jeram, S.M., Srikumar, T., Zhang, X.D., Anne Eisenhauer, H., Rogers, R., Pedrioli, P.G., Matunis, M. and Raught, B. An improved SUMmOn-based methodology for the identification of ubiquitin and ubiquitin-like protein (Ubl) conjugation sites identifies novel Ubl chain linkages. *Proteomics* **10**, 254–265 (2010).

23. Tsuchiya, H. *et al.* Ub-ProT reveals global length and composition of protein ubiquitylation in cells. *Nat. Commun.* **9**, (2018).

24. Rose, C. M. *et al.* Highly Multiplexed Quantitative Mass Spectrometry Analysis of Ubiquitylomes. *Cell Syst.* **3**, 395-403.e4 (2016).

25. Ohtake, F., Tsuchiya, H., Saeki, Y. & Tanaka, K. K63 ubiquitylation triggers proteasomal degradation by seeding branched ubiquitin chains. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E1401–E1408 (2018).

26. van der Wal, L. *et al.* Improvement of ubiquitylation site detection by Orbitrap mass spectrometry. *J. Proteomics* **172**, 49–56 (2018).

27. Zhang, N. *et al.* Comprehensive profiling of lysine ubiquitome reveals diverse functions of lysine ubiquitination in common wheat. *Sci. Rep.* **7**, 1–14 (2017).

28. Xu, Q., Farah, M., Webster, J. M. & Wojcikiewicz, R. J. H. Bortezomib rapidly suppresses ubiquitin thiolesterification to ubiquitin-conjugating enzymes and inhibits ubiquitination of histones and type I inositol 1,4,5-trisphosphate receptor. *Mol. Cancer Ther.* **3**, 1263–1269 (2004).

29. Gladman, N. P., Marshall, R. S., Lee, K. H. & Vierstra, R. D. The proteasome stress regulon is controlled by a pair of NAC transcription factors in arabidopsis. *Plant Cell* **28**, 1279–1296 (2016).

30. Danielsen, J. M. R. *et al.* Mass spectrometric analysis of lysine ubiquitylation reveals promiscuity at site level. *Mol. Cell. Proteomics* **10**, 1–12 (2011).

31. Mergner, J. *et al.* Mass-spectrometry-based draft of the Arabidopsis proteome. *Nature* **579**, 409–414 (2020).

32. Sarid-Krebs, L. *et al.* Phosphorylation of CONSTANS and its COP1-dependent degradation during photoperiodic flowering of Arabidopsis. *Plant J.* **84**, 451–463 (2015).

33. Bekker-Jensen, D. B. *et al.* Rapid and site-specific deep phosphoproteome profiling by data-independent acquisition without the need for spectral libraries. *Nat. Commun.* **11**, (2020).

34. Chiu, R. S., Pan, S., Zhao, R. & Gazzarrini, S. ABA-dependent inhibition of the ubiquitin proteasome system during germination at high temperature in Arabidopsis. *Plant J.* **88**, 749–761 (2016).

35. Malley, R. C. O., Barragan, C. C. & Ecker, J. R. A User's guide to the Arabidopsis T-DNA insertional mutant collections. *Methods Mol. Biol.* **1284**, 323–342 (2015).

36. Lincoln, C., Britton, J. H. & Estelle, M. Growth and development of the axr1 mutants of arabidopsis. *Plant Cell* **2**, 1071–1080 (1990).

37. Xu, P. *et al.* Integration of Jasmonic Acid and Ethylene Into Auxin Signaling in Root Development. *Front. Plant Sci.* **11**, 1–6 (2020).

38. Anne M Feke, Jinh, Hong, WEi Lui, Joshua, M. G. A decoy library uncovers U-box E3 ubiquitin ligases that regulate flowering time in Arabidopsis. *Orphanet J. Rare Dis.* **21**, 1–9 (2020).

39. Mazzucotelli, E. *et al.* The E3 Ubiquitin Ligase Gene Family in Plants: Regulation by Degradation. *Curr. Genomics* **7**, 509–522 (2006).

40. Uhrig, R. G., She, Y. M., Leach, C. A. & Plaxton, W. C. Regulatory monoubiquitination of phosphoenolpyruvate carboxylase in germinating castor oil seeds. *J. Biol. Chem.* **283**, 29650–29657 (2008).

41. Ruiz-Ballesta, I. *et al.* In vivo monoubiquitination of anaplerotic phosphoenolpyruvate carboxylase occurs at Lys624 in germinating sorghum seeds. *J. Exp. Bot.* **65**, 443–451 (2014).

42. Khan, M., Youn, J. Y., Gingras, A. C., Subramaniam, R. & Desveaux, D. In planta proximity dependent biotin identification (BioID). *Sci. Rep.* **8**, 2–9 (2018).

43. Chen, J.-J., Chou, C.-W., Chang, Y.-F. & Chen, C.-C. Proteasome Inhibitors Enhance TRAIL-Induced Apoptosis through the Intronic Regulation of DR5: Involvement of NF-κB and Reactive Oxygen Species-Mediated p53 Activation. *J. Immunol.* **180**, 8030–8039 (2008).

44. Luo, D. *et al.* MG132 selectively upregulates MICB through the DNA damage response pathway in A549 cells. *Mol. Med. Rep.* **19**, 213–220 (2019).

45. Gillet, L. C. *et al.* Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: A new concept for consistent and accurate proteome analysis. *Mol. Cell. Proteomics* **11**, 1–17 (2012).

46. Ellen R. Lubbers, P. J. M. Roles and regulation of protein phosphatase 2A (PP2A) in the heart Ellen. *J. Mol. Cell. Cardiol.* 127–133 (2016).

47. Flint, A. J., Tiganis, T., Barford, D. & Tonks, N. K. Development of 'substrate-

trapping' mutants to identify physiological substrates of protein tyrosine phosphatases. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 1680–1685 (1997).

48. Wu, D. *et al.* A substrate-trapping strategy for protein phosphatase PP1 holoenzymes using hypoactive subunit fusions. **1**, (2018).

49. Robinson, D. C. *et al.* An In Vitro Model of Charcot-Marie-Tooth Disease Type 4B2 Provides Insight Into the Roles of MTMR13 and MTMR2 in Schwann Cell Myelination. *ASN Neuro* **10**, (2018).

50. Bolis, A. *et al.* Dlg1, Sec8, and Mtmr2 regulate membrane homeostasis in Schwann cell myelination. *J. Neurosci.* **29**, 8858–8870 (2009).

51. Franklin, N. E., Bonham, C. A., Xhabija, B. & Vacratsis, P. O. Differential phosphorylation of the phosphoinositide 3-phosphatase MTMR2 regulates its association with early endosomal subtypes. *J. Cell Sci.* **126**, 1333–1344 (2013).

52. Harrisingh, M. C. *et al.* The Ras/Raf/ERK signalling pathway drives Schwann cell dedifferentiation. *EMBO J.* **23**, 3061–3071 (2004).

53. Cervellini, I. *et al.* Sustained MAPK/ERK activation in adult schwann cells impairs nerve repair. *J. Neurosci.* **38**, 679–690 (2018).

54. Berger, P., Tersar, K., Ballmer-Hofer, K. & Suter, U. The CMT4B disease-causing proteins MTMR2 and MTMR13/SBF2 regulate AKT signalling. *J. Cell. Mol. Med.* **15**, 307–315 (2011).

55. Sahenk, Z. & Ozes, B. Gene therapy to promote regeneration in Charcot-Marie-Tooth disease. *Brain Res.* **1727**, 146533 (2020).

56. Finkel, R. S. *et al.* Treatment of infantile-onset spinal muscular atrophy with nusinersen: a phase 2, open-label, dose-escalation study. *Lancet* **388**, 3017–3026 (2016).

57. Raess, M. A. *et al.* Expression of the neuropathy-associated MTMR2 gene rescues MTM1-associated myopathy. *Hum. Mol. Genet.* **26**, 3736–3748 (2017).

58. Danièle, N. *et al.* Intravenous administration of a MTMR2-encoding AAV vector ameliorates the phenotype of myotubular myopathy in mice. *J. Neuropathol. Exp. Neurol.* **77**, 282–295 (2018).

APPENDICES

APPENDIX A

Please see attached Appendix A TAZ. TAZ peptide identifications gathered using targeted and untargeted approaches and presented in **Appendix A.1**.

APPENDIX B

Please see attached Appendix B SMA EV, for protein identifications following our quantitative MS preparation of EVs. Protein outputs generated using Progenesis QI and abundance and significance calculations are presented in **Appendix B.1**. The results of investigation into the phosphorylation of IGFBP3 is presented in **Appendix B.2**.

APPENDIX C

Please see the attached file Appendix C MTMR2. **Appendix C.1** contains MS data obtained from traditional AP-MS, using FLAG immunoprecipitation. **Appendix C.1.1** shows protein identifications from control cells. Identifications from FLAG-MTMR2 expressing cells is shown in **Appendix C.1.2**. A comparison of protein identifications from control and FLAG-MTMR2 expressing cells is seen in **C.1.3**, presenting the overlap and unique identifications between the datasets. BioID was also used following traditional AP-MS measurements. **Appendix C.2** shows protein identifications from these experiments. EV-BirA, BirA-MTMR2^{WT}, and BirA-MTMR2^{S58A}, identifications and output are seen in **Appendix C.2.1**, **C.2.2**, and **C.2.3**, respectively.

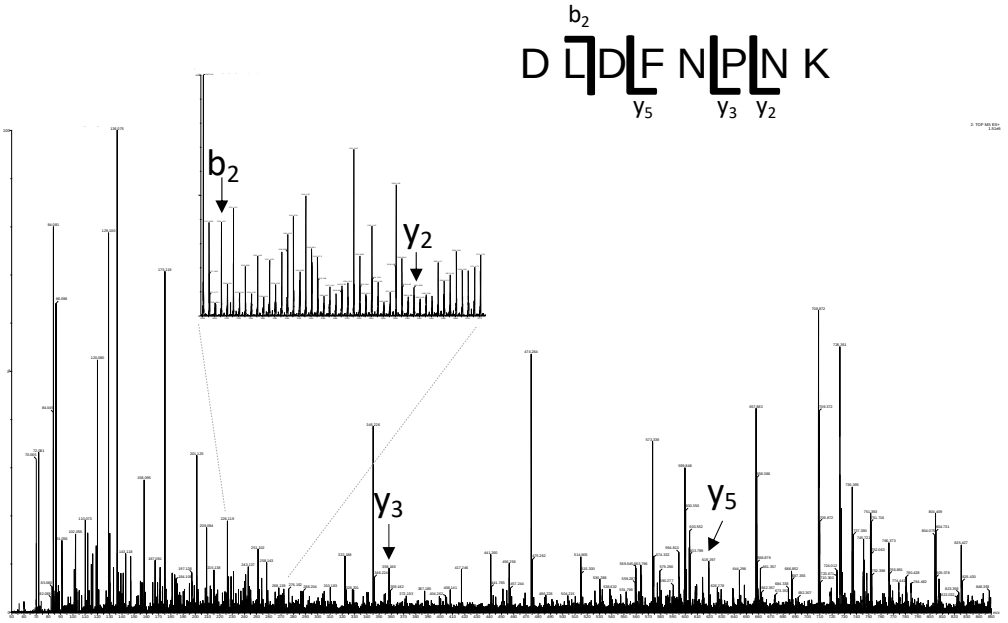


Figure C.1. Extracted spectra of TSSC1 identified peptide.

Through our BioID investigation of MTMR2^{Ser58A}, five peptides belonging to TSSC1 were reported. An annotated spectrum of the DLDNPNK peptide from high energy scans is shown, highlighting the identified y and b-ions.

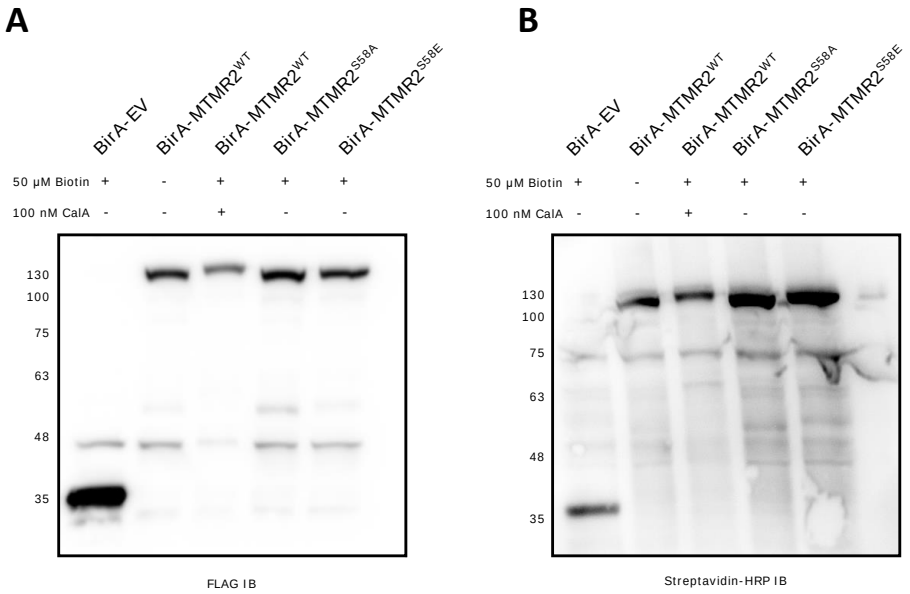


Figure C.2. MTMR2-BirA fusions treated with calyculin A are also downregulated. Schwann cells expressing BirA-MTMR2 constructs were treated with calyculin A and immunoblotted to see the changes in protein levels and biotin labeling. A mobility shift is observed upon phosphatase inhibitor treatment. Blots were probed with (**A**) anti-FLAG and (**B**) streptavidin-HRP.

APPENDIX D

Please see the attached file Appendix D Arabidopsis Ubiquitome. **Appendix D.1** contains identifications of modified proteins from plant extracts. **Appendix D.2** contains MS identifications following diglycine enrichment. **Appendix D.3** shows comparisons of proteins and peptides identified using our diglycine enrichment approach, and peptides analyzed without enrichment. A comparison of our findings to previously published Arabidopsis ubiquitome studies is shown in **Appendix D.4**. A comparison of peptides identified throughout different plant species is presented in **Appendix D.5**.

VITA AUCTORIS

NAME:	Justin Roberto
PLACE OF BIRTH:	Windsor, ON
YEAR OF BIRTH:	1992
EDUCATION:	Catholic Central High School, Windsor, ON, 2010
	University of Windsor, B.Sc., Windsor, ON, 2014