THE DYNAMICS OF POST-TRANSLATIONAL MODIFICATIONS AND PROTEIN-PROTEIN INTERACTIONS IN ESSENTIAL SIGNALING NETWORKS FOR STRESS RESPONSE AND SURVIVAL

Anthony A. Iannetta

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Chemistry in the School of Arts and Sciences.

Chapel Hill
2021

Approved by:

Leslie M. Hicks

Matthew R. Lockett

J. Michael Ramsey

Brandie M. Ehrmann

Gary L. Glish

ABSTRACT

Anthony A. Iannetta: The dynamics of post-translational modifications and protein-protein interactions in essential signaling networks for stress response and survival
(Under the direction of Leslie M. Hicks)

Organism stress causes perturbations to cellular homeostasis, which is required for proper growth and survival. Defense against stress depends on signaling networks regulated by post-translational modifications (PTMs) and protein-protein interactions (PPIs) that induce changes in protein expression and activity. Proteolysis is an irreversible PTM that helps maintain proteostasis, but can also result in the release of bioactive signaling peptides. Reversible PTMs, such as protein phosphorylation, modulate protein function and serve as a molecular switch to relay rapid signals through the cell. Physical interactions between proteins can dictate modifications and alter protein structure. Structural changes influence subsequent interactions with adjacent proteins and influence cellular signaling. The aim of this dissertation is to highlight method development and research characterizing PTMs and PPIs that are essential to stress signaling.

First, workflows for quantitative peptidomics and *in vitro* enzyme assays were optimized to determine potential thimet oligopeptidase substrates and confirm direct peptide substrates in the model plant species *Arabidopsis thaliana* (Chapter 2). Then, the early stages of the immune response in locally infected tissue of *Arabidopsis* were characterized to understand thimet oligopeptidase-mediated proteostasis (Chapter 3). Modulation in protein phosphorylation following antimicrobial treatment in the bacteria *Enterococcus faecalis* was investigated using quantitative phosphoproteomics (Chapter 4). From this, phosphorylation dependence on a kinase

known to be involved with *E. faecalis* antimicrobial resistance was revealed. Chapter 5 covers standard operating procedures for LC-MS/MS methods that can be employed to discover novel PPIs are discussed, including crosslinking- and immunoprecipitation-mass spectrometry. Then, the global protein interactome and novel protein structures were elucidated in the model alga species *Chlamydomonas reinhardtii* via crosslinking-mass spectrometry (Chapter 6).

Together, the work described in this dissertation highlights the utility of LC-MS/MS-based methods to analyze PTMs and PPIs in a discovery-based, high throughput manner. The insights gained from these experiments can be used to guide mechanistic studies that provide a better understanding into how changes in protein structure, function, or activity affect stress response signaling networks.

ACKNOWLEDGMENTS

I would first like to express gratitude to my advisor, Professor Leslie M. Hicks. Thank you for your mentorship over the years. I appreciate you allowing me to take ownership of my research and giving me the space to make critical decisions within my projects, while always being available to provide assistance if needed. Under your guidance and through the opportunities you provided, I have been able to develop skills across multiple subject areas that have enabled me to become a well-rounded researcher. To lab alumni, thank you for laying the foundation for the Hicks Laboratory and assisting me throughout my academic journey, especially upon joining the lab. To current lab members, I cherished our conversations and collaborations, thank you for continuously providing feedback.

To my family, especially my parents, thank you for your constant support and encouragement. Whenever I am going through a tough period, I know that our weekly, hour-long phone conversations will put me at ease and help me push through. Your confidence in me and my abilities is greatly appreciated. To my lover and friend Emily, thank you for being my number one cheerleader. You are always there to listen and provide motivation when I need it most. I am excited to see you flourish in graduate school and your future career. To my friends at Chapel Hill, I enjoyed our time together including weekend nights at Franklin St., Bert St., and Justice St. Thank you for providing respite from long hours in the lab. I hope that we can stay in touch.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND SYMBOLS

ACN	acetonitrile
AGC	automatic gain control
ANOVA	analysis of variance
AU	absorbance units
BH	Benjamini and Hochberg
CAA	chloroacetamide
CDD	conserved domain database
ClpP	caseinolytic protease P
CSMs	crosslink-spectrum matches
CV	coefficient of variation
CyOP	cytosolic oligopeptidase
DAMPs	damage‐associated molecular patterns
DDA	data-dependent acquisition
Deg	degradation of periplasmic proteins
DIA	data-independent acquisition
DMF	dimethylformamide
DMSO	dimethyl sulfoxide
Dnp	2,4-dinitrophenol
DSP	dithiobis(succinimidyl propionate)
DSSO	disuccinimidyl sulfoxide
DTT	dithiothreitol

ETI	effector-triggered immunity
FC	fold change
FDR	false discovery rate
GAPDH	glyceraldehyde-3-phosphate dehydrogenase
GI	gastrointestinal
GO	Gene Ontology
HA	hemagglutinin antigen
HADDOCK	High Ambiguity Driven protein-protein Docking
HCD	higher-energy C-trap dissociation
HPLC	high-performance liquid chromatography
HR	hypersensitive response
IAM	iodoacetamide
IMAC	immobilized metal affinity chromatography
IP	immunoprecipitation
IPTG	isopropyl β-D-1-thiogalactopyranoside
IreK	intrinsic resistance of enterococci kinase
I-TASSER	iterative threading assembly refinement
LC	liquid chromatography
LFQ	label-free quantification
LRR	leucine-rich repeat
LST8	lethal with SEC13 protein 8
Mca	7-methoxycoumarin-4-acetyl
MCAP	Mg-chelatase associated protein

MOAC	metal oxide affinity chromatography
MS	mass spectrometry
MS/MS	tandem mass spectrometry
MSP-MS	multiplex substrate profiling by mass spectrometry
NHS	N-hydroxysuccinimide
OD	optical density
OLLAS	Escherichia coli OmpF Linker and mouse Langerin fusion sequence
OOP	organellar oligopeptidase
PAGE	polyacrylamide gel electrophoresis
PAMPs	pathogen-associated molecular patterns
PAR	photosynthetically active radiation
PASTA	penicillin-binding protein and serine/threonine kinase associated
PCD	programmed cell death
PDB	Protein Data Bank
PPI	protein-protein interaction
PreP	presequence protease
PRK	phosphoribulokinase
PRM	parallel reaction monitoring
PRRs	pattern recognition receptors
Pst	Pseudomonas syringae pv. tomato pathovar DC3000
PTM	post-translational modification
PVDF	polyvinylidene fluoride
R	quantitative ratio

R2	coefficient of determination
REP27	photosystem II repair protein
ROS	reactive oxygen species
RPL11	chloroplast ribosomal protein L11
rpm	rotations per minute
RT	room temperature
SA	salicylic acid
SAR	systemic acquired resistance
SCX	strong cation exchange
SDS	sodium dodecyl sulfate
SOP	standard operating procedure
SPE	solid-phase extraction
SPP	stromal processing peptidase
TAILS	terminal amine isotopic labeling of substrates
TAP	Tris-acetate-phosphate
TBS	Tris-buffered saline
TBST	0.1% Tween 20 in Tris-buffered saline
TCA	trichloroacetic acid
TCEP	Tris(2-carboxyethyl)phosphine
TCS	two-component system
TFA	trifluoroacetic acid
TIC	total ion chromatogram
TOP	thimet oligopeptidase

UPLC	ultra performance liquid chromatography
XL	crosslinking

CHAPTER 1: Introduction

1.1 Stress Response

Organisms have mechanisms in place to monitor and regulate their internal environment (*e.g.*, temperature, ion concentration, pH). Cellular homeostasis preserves stability by maintaining acceptable levels of these variables, which is required for proper growth and survival. For example, auxin, a plant hormone involved in cell division, elongation, and differentiation, is regulated through various biosynthesis, conjugation, degradation, and transportation mechanisms[1–3].

Biotic and abiotic stressors cause disruptions to essential biological systems through events such as DNA damage and protein denaturation[4,5]. Organisms rely on complex biochemical signaling pathways to sense and defend against pathogen- or environment-induced stress that perturbs cellular homeostasis. When harmful stimuli are not resolved through adaptive responses, pathways leading to programmed cell death are triggered to eliminate these cells from the organism or population[6]. Defense responses are dependent on the type of stress, but typically include changes in protein expression and activity through modulation of gene transcription/translation and post-translational modifications (PTMs)[4,7].

In eukaryotes, one of the main survival pathways is the heat shock response, which counteracts misfolded proteins and protein aggregation due to heat, oxidative, and heavy metal stresses[8]. General protein translation is halted to prevent additional protein misfolding, while heat shock factors enhance the expression of heat shock protein genes, producing molecular

chaperones that help to refold misfolded proteins and reduce protein aggregation, while inhibiting apoptosis signaling pathways[9–11]. The activities of some of these heat shock proteins are regulated by PTMs, with stress-induced phosphorylation promoting the monomeric form and stabilizing its interaction with different partners[12]. In bacteria, two-component systems (TCSs) act as key regulators of signal transduction[13,14]. TCSs typically contain a histidine kinase and its cognate response regulator, where upon a signal (*e.g.*, cell wall stress) the kinase autophosphorylates at a conserved histidine residue and transfers the phosphoryl group to a conserved aspartate on the response regulator. The phosphorylation of this response regulator can lead to transcriptional/translational control and protein-protein interactions for downstream signaling[15–18]. Gene expression and PTM control are conserved across all domains of life as a means to correct cellular homeostasis in response to stress signals.

1.2 Post-Translational Modifications

While genomes across organisms of the same species are fairly identical, even throughout their lifespans, the proteome varies significantly within an individual during its lifecycle and between organisms. Contributing to this complexity is post-translational control of the proteome, where proteins from the same gene can contain various PTMs resulting in different variations, known as proteoforms[19]. PTMs occur either through covalent chemical modification or hydrolytic peptide bond cleavage. These modifications facilitate rapid changes to protein structure and function, which can alter protein stability, signaling, interactions, and/or localization[20–25]. More than 400 types of PTMs have been experimentally discovered; however, only a small subset are common and have been extensively examined including phosphorylation, reversible oxidation, ubiquitination, and glycosylation[26,27]. The transient nature of certain PTMs make them suitable as dynamic switches for relaying rapid signals through the cell; for example,

kinase cascades are turned on and off by reversible phosphorylation[28]. Irreversible PTMs can signify loss of protein function and degradation during the stress response (*e.g.*, some forms of oxidation), but proteolytic cleavage of pro-proteins can produce signaling proteins like cytokines that are involved in regulating immune responses[29,30].

Liquid chromatography-tandem mass spectrometry (LC-MS/MS) has become the predominant analytical tool for investigating PTMs due to high sensitivity and the ability to perform large scale PTM analyses across an entire proteome[31,32]. However, certain PTMs are transient and present at substoichiometric levels, often requiring specific enrichment during sample preparation to obtain modified peptides within the limit of detection and dynamic range of the mass spectrometer for thorough depth of PTM coverage[33]. Developed LC-MS/MS methods can be applied to study changes in PTMs following organism stress to identify proteins and specific modification sites involved in defense signaling.

1.3 Proteostasis and Proteolysis

Protein homeostasis (proteostasis) is imperative to maintain a stable proteome for proper cellular function and efficient responses to stress[34]. Organisms have dedicated quality control mechanisms to fold and assemble defective proteins impaired by stress, but damaged proteins that fail to refold are eliminated through proteolysis[34,35]. In eukaryotes, the ubiquitin-proteasome system is the major pathway for targeting terminally misfolded proteins and degrading these substrates, but cells contain other quality control proteases and peptidases as well[36]. While proteolysis is heavily associated with proteostasis, stress-induced proteolytic processing of a protein can result in the production of signaling peptides[37,38]. In animals, this controlled proteolysis produces peptides that are presented by the major histocompatibility complex for recognition by T cell receptors and immune response signaling to eliminate pathogens[39,40].

Controlled proteolysis for stress response and defense signaling is less understood in plants, but it is continuously being explored. For example, plant elicitor peptides released by metacaspase-mediated proteolysis of PROPEPs activate salicylic acid and jasmonic acid-mediated stress responses and enhance resistance to bacterial pathogens[41–43]. Additional investigation of controlled proteolysis in plants will provide an increased understanding into how protease activities and their bioactive peptide products influence proper responses for survival, which is necessary to prevent reduction of agricultural yields and substantial economic losses due to plant stress.

Many MS-based proteomics and peptidomics approaches have been developed to discover protease/peptidase substrates and cleavage specificities. A branch of comparative proteomics techniques known as N-terminomics selectively isolates protein N-terminal peptides and distinguishes the N-termini of mature proteins from protease-generated N-termini to determine protease substrates. One example is terminal amine isotopic labeling of substrates (TAILS). Here, protein amine groups are isotopically labeled in samples with and without proteolytic activity, samples are combined, and blocked N-termini are enriched through negative selection for downstream LC-MS/MS analysis of N-terminal peptides[44]. Substrate trapping is another approach for identifying substrates and involves site-directed mutagenesis to generate catalytically inactive proteases. This prevents degradation and stabilizes the interaction between substrates and the protease, which allows for affinity enrichment of the protease and its interactors for LC-MS/MS analysis[45]. Negative controls are included to differentiate between true substrates and non-specific binding proteins. The technique called multiplex substrate profiling by mass spectrometry (MSP-MS) has been developed to determine peptidase cleavage specificity. This assay uses a diverse library of peptides incorporating all combinations of

neighbor and near-neighbor amino acids pairs, and the comparison of peptidase-treated and untreated samples at different time intervals reveals substrate specificity[46]. Differential, quantitative analyses can be performed to compare proteomes/peptidomes in the presence and absence of proteolytic activity. Protease/peptidase activities nullified via genetic or chemical approaches increases the accumulation of substrates, so the identification of proteins/peptides with an increased abundance following inhibited proteolysis provides a list of potential substrates[47]. This requires follow up validation to differentiate between true substrates and pleiotropic effects, but this method can be preferable over the previously described techniques. It can be more reproducible than TAILS since it requires less steps, identifies substrates that transiently interact with the protease/peptidase that may be lost during substrate trapping, and allows for the direct identification of substrates unlike MSP-MS. Nevertheless, all of these approaches enable the characterization of proteases/peptidases, which increases understanding of their functions and contributions to proteostasis and stress signaling.

1.4 Protein Phosphorylation

Protein phosphorylation is a ubiquitous PTM responsible for regulating a variety of cellular processes including cell growth and programmed cell death[48,49]. The addition or removal of a phosphate moiety by a kinase or phosphatase is responsible for conformational alterations of substrates which affect function and cause changes in protein interaction partners. The significant role phosphorylation has in processes such as protein synthesis, metabolism, cell division, and signal transduction has increased interest in understanding the biological significance of this modification[50–52]. In eukaryotes, phosphorylation is most prevalent on protein serine, threonine, and tyrosine residues, but noncanonical phosphorylation on histidine, lysine, arginine, and cysteine residues occur as well. In bacteria, histidine and arginine phosphorylation

in TCSs are involved with nitrogen regulation, chemotaxis, and sporulation signal transduction. Additional phosphorylation that is mediated by eukaryotic-like serine/threonine kinases/phosphatases influences cell division, cell wall biosynthesis, and stress response[53,54].

Historically, protein phosphorylation was analyzed using gel-based techniques and visualized by ^{32}P labeling or western blotting with phosphosite-specific antibodies, but it is challenging to identify novel phosphoproteins and localize the phosphosite using these methods[55]. Instead, MS-based phosphoproteomics has become the primary approach for investigating protein phosphorylation, due to its high sensitivity, extensive throughput, ability to detect previously unknown phosphoproteins, and capacity to localize the phosphosite within the protein[56]. However, this PTM is transient and sub-stoichiometric, which results in low abundance of phosphopeptides within a digested proteome, challenging MS-based phosphoproteomics analysis. Phosphopeptide enrichment prior to LC-MS/MS analysis is commonly used to decrease sample complexity and improve detection, leading to greater depth of coverage of the phosphoproteome[57,58].

Immobilized metal affinity chromatography (IMAC) and metal oxide affinity chromatography (MOAC) are the most common enrichments for peptides containing phosphorylated serine and threonine residues[59–61]. In IMAC, phosphopeptides are enriched due to the affinity between positively charged immobilized metal ions and the negatively charged phosphate group on phosphopeptides. For MOAC, the interaction of phosphate-containing peptides and the metal oxide occur by reversible Lewis acid-base chemistry. MOAC predominately enriches singly phosphorylated peptides, while IMAC tend to enrich more multiply phosphorylated peptides[62,63]. Both IMAC and MOAC suffer from non-specific binding of acidic peptides, but this can be alleviated with acidic loading and washing buffer to protonate

peptidyl carboxylic acids while keeping phosphorylated residues negatively charged[64]. Additionally, organic acid additives such as phthalic acid or lactic acid can be included in MOAC buffers to outcompete non-phosphorylated peptides for adsorption to the metal oxide, increasing enrichment specificity[65]. Phosphopeptide enrichment prior to LC-MS/MS analysis provides the ability to perform in-depth proteomic profiling of protein phosphorylation and examine its role in stress response signaling.

1.5 Protein-Protein Interactions

Protein-protein interactions (PPIs) form protein networks that are involved in important cellular pathways such as cell cycle control, translation, metabolism, and signal transduction[66–68]. PPIs can have a multitude of effects that control downstream signaling, including: 1) alteration of kinetic properties of proteins, 2) formation of new binding sites, 3) inactivation of proteins, and 4) modification of substrate specificity[69]. These physical interactions can induce PTMs that influence and regulate PPIs through conformational changes by altering affinity, cooperativity, and kinetic factors of different interactions[70]. Identifying and characterizing these interactions is not only critical to better understand protein function and regulation, but also to gain insights into cellular signaling.

Interactomics (*i.e.*, the study of PPIs) has historical relied on genetic (*e.g.*, yeast two-hybrid) and structural (*e.g.*, NMR spectroscopy and X-ray crystallography) approaches[71]. The yeast two-hybrid system is rapid and can be applied in large-scale interactome mapping, but it is associated with a high false positive rate and the yeast system lacks certain PTMs that could prove necessary for interactions and subsequent identification of PPIs[72]. Structural techniques can provide high-resolution, three-dimensional structural information for proteins and PPIs, but demand high protein concentrations and, in the case of X-ray crystallography, require that a

protein can be crystallized[73]. To complement these established methods, MS-based techniques such as immunoprecipitation-mass spectrometry (IP-MS) and crosslinking-mass spectrometry (XL-MS) have been developed to investigate the protein interactome. The advantages of these methods are that proteins can be captured or crosslinked at physiological pH, allowing for the identification of biologically relevant PPIs, and they can leverage the high sensitivity of MS to analyze low abundant proteins[71,74]. IP-MS is implemented when trying to investigate the protein-protein interactions of a specific protein, while XL-MS can be used to elucidate the global protein interactome.

Immunoprecipitation (IP) is a form of affinity purification that isolates a protein of interest as well as its interacting partners. An antibody immobilized on a solid support binds either an endogenous target protein or a protein tag that has been genetically fused to the protein of interest and purifies it from the protein lysate with its associated proteins. IPs using an antibody against a protein of interest ensures that its expression is not altered by the addition of a protein tag[75]. Incorporating epitope-tagged proteins is more favorable due to the limited availability of antibodies for certain proteins and the antibody specific to an endogenous protein may disrupt PPIs if the epitope coincides with an interface needed for its interactions[76]. After enrichment, proteins are identified using LC-MS/MS analysis and their abundances are compared to a negative control to distinguish between true interactors and false positives resulting from non-specific binding to the antibody or solid support. In XL-MS, proteins in close proximity are covalently bound using a crosslinker to preserve native conformations and PPIs. Following enzymatic digestion, crosslinked peptides can be identified using LC-MS/MS analysis, which infers various PPIs and interaction interfaces. The application of MS-based techniques to

investigate the global protein interactome can identify novel PPIs in a high throughput manner, providing a better understand of protein function and cellular signal transduction.

1.6 Scope of Dissertation

This dissertation describes method development and studies aimed at the investigation of PTMs and PPIs and their modulation in stress signaling pathways to produce responses for survival in various organisms. All chapters revolve around LC-MS-based top-down peptidomics or bottom-up proteomics, using label-free quantification when performing differential analyses for peptide or protein abundances across conditions or genotypes. **Chapter 2** details the development and optimization of a quantitative peptidomics method to profile the *Arabidopsis thaliana* peptidome. This method is applied in a differential analysis between wild type and peptidase knockout mutant plant lines to reveal potential peptidase substrates. Following this analysis, *in vitro* enzyme assays are employed for validation of direct substrates identified in the differential analysis. This quantitative peptidomics method is then implemented (**Chapter 3**) to characterize the peptidomes in locally infected tissue of these plant lines during the early stages of biotic stress response. This study provides a framework for peptidase signaling networks, through which the interplay between proteolytic pathways and defense signaling can be further characterized. **Chapter 4** focuses on the development of a quantitative phosphoproteomics workflow to study protein phosphorylation in *Enterococcus faecalis*. This platform is then applied in a differential analysis between wild type and kinase mutant strains after various cell wall-active antimicrobial treatments to investigate stress-modulated protein phosphorylation and discover additional substrates for a kinase known to be involved with antibiotic resistance. **Chapter 5** describes standard operating procedures (SOPs) for IP-MS and XL-MS workflows that were optimized for the model alga species *Chlamydomonas reinhardtii* to analyze the

protein interactome. These XL-MS procedures were then applied (**Chapter 6**) to examine the global protein interactome in *C. reinhardtii* and develop computational models for protein structures and PPIs. Finally, **Chapter 7** summarizes the findings presented in this dissertation and considers potential future directions related to these research areas.

REFERENCES

1. Simon, S.; Skůpa, P.; Viaene, T.; Zwiewka, M.; Tejos, R.; Klíma, P.; Čarná, M.; Rolčík, J.; Rycke, R. De; Moreno, I.; Dobrev, P. I.; Orellana, A.; Zažímalová, E.; Friml, J. PIN6 auxin transporter at endoplasmic reticulum and plasma membrane mediates auxin homeostasis and organogenesis in *Arabidopsis*. *New Phytol.* **2016**, *211*(1), 65–74.

2. Ljung, K. Auxin metabolism and homeostasis during plant development. *Development* **2013**, *140*(5), 943–950.

3. Zhang, J.; Peer, W. A. Auxin homeostasis: the DAO of catabolism. *J. Exp. Bot.* **2017**, *68*(12), 3145–3154.

4. Fulda, S.; Gorman, A. M.; Hori, O.; Samali, A. Cellular stress responses: Cell survival and cell death. *Int. J. Cell Biol.* **2010**, *2010*, 214074.

5. Galluzzi, L.; Yamazaki, T.; Kroemer, G. Linking cellular stress responses to systemic homeostasis. *Nat. Rev. Mol. Cell Biol.* **2018**, *19*(11), 731–745.

6. Galluzzi, L.; Vitale, I.; Aaronson, S. A.; Abrams, J. M.; Adam, D.; Agostinis, P.; Alnemri, E. S.; Altucci, L.; Amelio, I.; Andrews, D. W.; Annicchiarico-Petruzzelli, M.; Antonov, A. V.; Arama, E.; Baehrecke, E. H.; Barlev, N. A.; Bazan, N. G.; Bernassola, F.; Bertrand, M. J. M.; Bianchi, K.; Blagosklonny, M. V.; Blomgren, K.; Borner, C.; Boya, P.; Brenner, C.; Campanella, M.; Candi, E.; Carmona-Gutierrez, D.; Cecconi, F.; Chan, F. K.-M.; Chandel, N. S.; Cheng, E. H.; Chipuk, J. E.; Cidlowski, J. A.; Ciechanover, A.; Cohen, G. M.; Conrad, M.; Cubillos-Ruiz, J. R.; Czabotar, P. E.; D'Angiolella, V.; Dawson, T. M.; Dawson, V. L.; De Laurenzi, V.; De Maria, R.; Debatin, K.-M.; DeBerardinis, R. J.; Deshmukh, M.; Di Daniele, N.; Di Virgilio, F.; Dixit, V. M.; Dixon, S. J.; Duckett, C. S.; Dynlacht, B. D.; El-Deiry, W. S.; Elrod, J. W.; Fimia, G. M.; Fulda, S.; García-Sáez, A. J.; Garg, A. D.; Garrido, C.; Gavathiotis, E.; Golstein, P.; Gottlieb, E.; Green, D. R.; Greene, L. A.; Gronemeyer, H.; Gross, A.; Hajnoczky, G.; Hardwick, J. M.; Harris, I. S.; Hengartner, M. O.; Hetz, C.; Ichijo, H.; Jäättelä, M.; Joseph, B.; Jost, P. J.; Juin, P. P.; Kaiser, W. J.; Karin, M.; Kaufmann, T.; Kepp, O.; Kimchi, A.; Kitsis, R. N.; Klionsky, D. J.; Knight, R. A.; Kumar, S.; Lee, S. W.; Lemasters, J. J.; Levine, B.; Linkermann, A.; Lipton, S. A.; Lockshin, R. A.; López-Otín, C.; Lowe, S. W.; Luedde, T.; Lugli, E.; MacFarlane, M.; Madeo, F.; Malewicz, M.; Malorni, W.; Manic, G.; Marine, J.-C.; Martin, S. J.; Martinou, J.-C.; Medema, J. P.; Mehlen, P.; Meier, P.; Melino, S.; Miao, E. A.; Molkentin, J. D.; Moll, U. M.; Muñoz-Pinedo, C.; Nagata, S.; Nuñez, G.; Oberst, A.; Oren, M.; Overholtzer, M.; Pagano, M.; Panaretakis, T.; Pasparakis, M.; Penninger, J. M.; Pereira, D. M.; Pervaiz, S.; Peter, M. E.; Piacentini, M.; Pinton, P.; Prehn, J. H. M.; Puthalakath, H.; Rabinovich, G. A.; Rehm, M.; Rizzuto, R.; Rodrigues, C. M. P.; Rubinsztein, D. C.; Rudel, T.; Ryan, K. M.; Sayan, E.; Scorrano, L.; Shao, F.; Shi, Y.; Silke, J.; Simon, H.-U.; Sistigu, A.; Stockwell, B. R.; Strasser, A.; Szabadkai, G.; Tait, S. W. G.; Tang, D.; Tavernarakis, N.; Thorburn, A.; Tsujimoto, Y.; Turk, B.; Vanden Berghe, T.; Vandenabeele, P.; Vander Heiden, M. G.; Villunger, A.; Virgin, H. W.; Vousden, K. H.; Vucic, D.; Wagner, E. F.; Walczak, H.; Wallach, D.; Wang, Y.; Wells, J. A.; Wood, W.; Yuan, J.; Zakeri, Z.; Zhivotovsky, B.; Zitvogel, L.; Melino, G.; Kroemer,

G. Molecular mechanisms of cell death: recommendations of the Nomenclature Committee on Cell Death 2018. *Cell Death Differ.* **2018**, *25*(3), 486–541.

7. Spriggs, K. A.; Bushell, M.; Willis, A. E. Translational Regulation of Gene Expression during Conditions of Cell Stress. *Mol. Cell* **2010**, *40*(2), 228–237.

8. Lindquist, S. The heat-shock response. *Annu. Rev. Biochem.* **1986**, *55*, 1151–1191.

9. Samali, A.; Orrenius, S. Heat shock proteins: regulators of stress response and apoptosis. *Cell Stress Chaperones* **1998**, *3*(4), 228.

10. Morimoto, R. I.; Kroeger, P. E.; Cotto, J. J. The Transcriptional regulation of heat shock genes: A plethora of heat shock factors and regulatory conditions. *EXS* **1996**, *77*, 139–163.

11. Samali, A.; Cotter, T. G. Heat Shock Proteins Increase Resistance to Apoptosis. *Exp. Cell Res.* **1996**, *223*(1), 163–170.

12. Lavoie, J. N.; Lambert, H.; Hickey, E.; Weber, L. A.; Landry, J. Modulation of cellular thermoresistance and actin filament stability accompanies phosphorylation-induced changes in the oligomeric structure of heat shock protein 27. *Mol. Cell. Biol.* **1995**, *15*(1), 505–516.

13. Zschiedrich, C. P.; Keidel, V.; Szurmant, H. Molecular Mechanisms of Two-Component Signal Transduction. *J. Mol. Biol.* **2016**, *428*(19), 3752–3775.

14. Capra, E. J.; Laub, M. T. Evolution of Two-Component Signal Transduction Systems. *Annu. Rev. Microbiol.* **2012**, *66*, 325–347.

15. Raffa, R. G.; Raivio, T. L. A third envelope stress signal transduction pathway in *Escherichia coli*. *Mol. Microbiol.* **2002**, *45*(6), 1599–1611.

16. Wanner, B. L. Gene regulation by phosphate in enteric bacteria. *J. Cell. Biochem.* **1993**, *51*(1), 47–54.

17. Kato, A.; Groisman, E. A. The PhoQ/PhoP Regulatory Network of *Salmonella enterica*. *Adv. Exp. Med. Biol.* **2008**, *631*, 7–21.

18. Lee, S.; Hinz, A.; Bauerle, E.; Angermeyer, A.; Juhaszova, K.; Kaneko, Y.; Singh, P. K.; Manoil, C. Targeting a bacterial stress response to enhance antibiotic action. *Proc. Natl. Acad. Sci.* **2009**, *106*(34), 14570–14575.

19. Muñoz, J.; Heck, A. J. R. From the Human Genome to the Human Proteome. *Angew. Chemie Int. Ed.* **2014**, *53*(41), 10864–10866.

20. Qi, H. H.; Ongusaha, P. P.; Myllyharju, J.; Cheng, D.; Pakkanen, O.; Shi, Y.; Lee, S. W.; Peng, J.; Shi, Y. Prolyl 4-hydroxylation regulates Argonaute 2 stability. *Nature* **2008**, *455*(7211), 421–424.

21. Sahar, S.; Zocchi, L.; Kinoshita, C.; Borrelli, E.; Sassone-Corsi, P. Regulation of BMAL1 Protein Stability and Circadian Function by GSK3β-Mediated Phosphorylation. *PLoS One* **2010**, *5*(1), e8561.

22. Deribe, Y. L.; Pawson, T.; Dikic, I. Post-translational modifications in signal integration. *Nat. Struct. Mol. Biol.* **2010**, *17*(6), 666–672.

23. Liu, J.; Qian, C.; Cao, X. Post-Translational Modification Control of Innate Immunity. *Immunity* **2016**, *45*(1), 15–30.

24. Ahearn, I. M.; Haigis, K.; Bar-Sagi, D.; Philips, M. R. Regulating the regulator: Post-translational modification of RAS. *Nat. Rev. Mol. Cell Biol.* **2012**, *13*(1), 39–51.

25. Xie, Y.; Kang, R.; Sun, X.; Zhong, M.; Huang, J.; Klionsky, D. J.; Tang, D. Posttranslational modification of autophagy-related proteins in macroautophagy. *Autophagy* **2015**, *11*(1), 28–45.

26. Khoury, G. A.; Baliban, R. C.; Floudas, C. A. Proteome-wide post-translational modification statistics: Frequency analysis and curation of the swiss-prot database. *Sci. Rep.* **2011**, *1*(1), 1–5.

27. Audagnotto, M.; Dal Peraro, M. Protein post-translational modifications: In silico prediction tools and molecular modeling. *Comput. Struct. Biotechnol. J.* **2017**, *15*, 307–319.

28. Cohen, P. The regulation of protein function by multisite phosphorylation – a 25 year update. *Trends Biochem. Sci.* **2000**, *25*(12), 596–601.

29. Savickas, S.; Kastl, P.; auf dem Keller, U. Combinatorial degradomics: Precision tools to unveil proteolytic processes in biological systems. *Biochim. Biophys. Acta - Proteins Proteomics* **2020**, *1868*(6), 140392.

30. Couturier, J.; Chibani, K.; Jacquot, J.-P.; Rouhier, N. Cysteine–based redox regulation and signaling in plants. *Front. Plant Sci.* **2013**, *4*, 105.

31. Choudhary, C.; Mann, M. Decoding signalling networks by mass spectrometry-based proteomics. *Nat. Rev. Mol. Cell Biol.* **2010**, *11*(6), 427–439.

32. Olsen, J. V.; Mann, M. Status of large-scale analysis of posttranslational modifications by mass spectrometry. *Mol. Cell. Proteomics* **2013**, *12*(12), 3444–3452.

33. Zhang, Y.; Zhang, C.; Jiang, H.; Yang, P.; Lu, H. Fishing the PTM proteome with chemical approaches using functional solid phases. *Chem. Soc. Rev.* **2015**, *44*(22), 8260–8287.

34. Koga, H.; Kaushik, S.; Cuervo, A. M. Protein homeostasis and aging: The importance of exquisite quality control. *Ageing Res. Rev.* **2011**, *10*(2), 205–215.

35. Jarosz, D. F.; Taipale, M.; Lindquist, S. Protein homeostasis and the phenotypic manifestation of genetic diversity: principles and mechanisms. *Annu. Rev. Genet.* **2010**, *44*, 189–216.

36. Bard, J. A. M.; Goodall, E. A.; Greene, E. R.; Jonsson, E.; Dong, K. C.; Martin, A. Structure and Function of the 26S Proteasome. *Annu. Rev. Biochem.* **2018**, *87*, 697–724.

37. Brogden, K. A.; Guthmiller, J. M.; Salzet, M.; Zasloff, M. The nervous system and innate immunity: the neuropeptide connection. *Nat. Immunol.* **2005**, *6*(6), 558–564.

38. Hou, S.; Liu, Z.; Shen, H.; Wu, D. Damage-associated molecular pattern-triggered immunity in plants. *Front. Plant Sci.* **2019**, *10*, 646.

39. Dersh, D.; Hollý, J.; Yewdell, J. W. A few good peptides: MHC class I-based cancer immunosurveillance and immunoevasion. *Nat. Rev. Immunol.* **2020**, *21*(2), 116–128.

40. Neefjes, J.; Ovaa, H. A peptide's perspective on antigen presentation to the immune system. *Nat. Chem. Biol.* **2013**, *9*(12), 769–775.

41. Huffaker, A.; Pearce, G.; Ryan, C. A. An endogenous peptide signal in *Arabidopsis* activates components of the innate immune response. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*(26), 10098–10103.

42. Ma, Y.; Walker, R. K.; Zhao, Y.; Berkowitz, G. A. Linking ligand perception by PEPR pattern recognition receptors to cytosolic Ca2+ elevation and downstream immune signaling in plants. *Proc. Natl. Acad. Sci.* **2012**, *109*(48), 19852–19857.

43. Yamaguchi, Y.; Huffaker, A.; Bryan, A. C.; Tax, F. E.; Ryan, C. A. PEPR2 Is a Second Receptor for the Pep1 and Pep2 Peptides and Contributes to Defense Responses in *Arabidopsis*. *Plant Cell* **2012**, *22*(2), 508–522.

44. Kleifeld, O.; Doucet, A.; auf dem Keller, U.; Prudova, A.; Schilling, O.; Kainthan, R. K.; Starr, A. E.; Foster, L. J.; Kizhakkedathu, J. N.; Overall, C. M. Isotopic labeling of terminal amines in complex samples identifies protein N-termini and protease cleavage products. *Nat. Biotechnol.* **2010**, *28*(3), 281–288.

45. Rei Liao, J. Y.; van Wijk, K. J. Discovery of AAA+ Protease Substrates through Trapping Approaches. *Trends Biochem. Sci.* **2019**, *44*(6), 528–545.

46. O'Donoghue, A. J.; Eroy-Reveles, A. A.; Knudsen, G. M.; Ingram, J.; Zhou, M.; Statnekov, J. B.; Greninger, A. L.; Hostetter, D. R.; Qu, G.; Maltby, D. A.; Anderson, M. O.; DeRisi, J. L.; McKerrow, J. H.; Burlingame, A. L.; Craik, C. S. Global identification of peptidase specificity by multiplex substrate profiling. *Nat. Methods* **2012**, *9*(11), 1095–1100.

47. Lone, A. M.; Kim, Y. G.; Saghatelian, A. Peptidomics methods for the identification of peptidase-substrate interactions. *Curr. Opin. Chem. Biol.* **2013**, *17*(1), 83–89.

48. Ward, P. S.; Thompson, C. B. Signaling in control of cell growth and metabolism. *Cold Spring Harb. Perspect. Biol.* **2012**, *4*(7), 1–15.

49. Dhanasekaran, D. N.; Premkumar Reddy, E. JNK-signaling: A multiplexing hub in programmed cell death. *Genes and Cancer* **2017**, *8*(9–10), 682–694.

50. Humphrey, S. J.; James, D. E.; Mann, M. Protein Phosphorylation: A Major Switch Mechanism for Metabolic Regulation. *Trends Endocrinol. Metab.* **2015**, *26*(12), 676–687.

51. Dennis, M. D.; Jefferson, L. S.; Kimball, S. R. Role of p70S6K1-mediated phosphorylation of eIF4B and PDCD4 proteins in the regulation of protein synthesis. *J. Biol. Chem.* **2012**, *287*(51), 42890–42899.

52. Beilharz, K.; Nováková, L.; Fadda, D.; Branny, P.; Massidda, O.; Veening, J. W. Control of cell division in *Streptococcus pneumoniae* by the conserved Ser/Thr protein kinase StkP. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*(15), E905–E913.

53. Stock, J. B.; Ninfa, A. J.; Stock, A. M. Protein phosphorylation and regulation of adaptive responses in bacteria. *Microbiol. Rev.* **1989**, *53*(4), 450.

54. Janczarek, M.; Vinardell, J.-M.; Lipa, P.; Karaś, M. Hanks-Type Serine/Threonine Protein Kinases and Phosphatases in Bacteria: Roles in Signaling and Adaptation to Various Environments. *Int. J. Mol. Sci.* **2018**, *19*(10), 2872.

55. Macek, B.; Mann, M.; Olsen, J. V. Global and site-specific quantitative phosphoproteomics: Principles and applications. *Annu. Rev. Pharmacol. Toxicol.* **2009**, *49*, 199–221.

56. Riley, N. M.; Coon, J. J. Phosphoproteomics in the Age of Rapid and Deep Proteome Profiling. *Anal. Chem.* **2016**, *88*(1), 74–94.

57. Zhao, Y.; Jensen, O. N. Modification-specific proteomics: Strategies for characterization of post-translational modifications using enrichment techniques. *Proteomics* **2009**, *9*(20), 4632–4641.

58. Huang, J.; Wang, F.; Ye, M.; Zou, H. Enrichment and separation techniques for large-scale proteomics analysis of the protein post-translational modifications. *J. Chromatogr. A* **2014**, *1372*, 1–17.

59. Humphrey, S. J.; Azimifar, S. B.; Mann, M. High-throughput phosphoproteomics reveals *in vivo* insulin signaling dynamics. *Nat. Biotechnol.* **2015**, *33*(9), 990–995.

60. Liu, J. J.; Sharma, K.; Zangrandi, L.; Chen, C.; Humphrey, S. J.; Chiu, Y. T.; Spetea, M.; Liu-Chen, L. Y.; Schwarzer, C.; Mann, M. *In vivo* brain GPCR signaling elucidated by phosphoproteomics. *Science* **2018**, *360*(6395), eaao4927.

61. Tan, H.; Yang, K.; Li, Y.; Shaw, T. I.; Wang, Y.; Blanco, D. B.; Wang, X.; Cho, J. H.; Wang, H.; Rankin, S.; Guy, C.; Peng, J.; Chi, H. Integrative Proteomics and

Phosphoproteomics Profiling Reveals Dynamic Signaling Networks and Bioenergetics Pathways Underlying T Cell Activation. *Immunity* **2017**, *46*(3), 488–503.

62. Mann, M.; Ong, S. E.; Grønborg, M.; Steen, H.; Jensen, O. N.; Pandey, A. Analysis of protein phosphorylation using mass spectrometry: Deciphering the phosphoproteome. *Trends Biotechnol.* **2002**, *20*(6), 261–268.

63. Bodenmiller, B.; Mueller, L. N.; Mueller, M.; Domon, B.; Aebersold, R. Reproducible isolation of distinct, overlapping segments of the phosphoproteome. *Nat. Methods* **2007**, *4*(3), 231–237.

64. Beltran, L.; Cutillas, P. R. Advances in phosphopeptide enrichment techniques for phosphoproteomics. *Amino Acids* **2012**, *43*(3), 1009–1024.

65. Sugiyama, N.; Masuda, T.; Shinoda, K.; Nakamura, A.; Tomita, M.; Ishihama, Y. Phosphopeptide enrichment by aliphatic hydroxy acid-modified metal oxide chromatography for nano-LC-MS/MS in proteomics applications. *Mol. Cell. Proteomics* **2007**, *6*(6), 1103–1109.

66. Khmelinskii, A.; Roostalu, J.; Roque, H.; Antony, C.; Schiebel, E. Phosphorylation-Dependent Protein Interactions at the Spindle Midzone Mediate Cell Cycle Regulation of Spindle Elongation. *Dev. Cell* **2009**, *17*(2), 244–256.

67. Kandel, S. E.; Lampe, J. N. Role of Protein–Protein Interactions in Cytochrome P450-Mediated Drug Metabolism and Toxicity. *Chem. Res. Toxicol.* **2014**, *27*(9), 1474–1486.

68. Pawson, T.; Nash, P. Protein–protein interactions define specificity in signal transduction. *Genes Dev.* **2000**, *14*(9), 1027–1047.

69. Phizicky, E. M.; Fields, S. Protein-protein interactions: methods for detection and analysis. *Microbiol. Rev.* **1995**, *59*(1), 94–123.

70. Berggård, T.; Linse, S.; James, P. Methods for the detection and analysis of protein–protein interactions. *Proteomics* **2007**, *7*(16), 2833–2842.

71. Wetie, A. G. N.; Sokolowska, I.; Woods, A. G.; Roy, U.; Loo, J. A.; Darie, C. C. Investigation of stable and transient protein–protein interactions: Past, present, and future. *Proteomics* **2013**, *13*(3–4), 538–557.

72. Ngounou Wetie, A. G.; Sokolowska, I.; Woods, A. G.; Roy, U.; Deinhardt, K.; Darie, C. C. Protein–protein interactions: switch from classical methods to proteomics and bioinformatics-based approaches. *Cell. Mol. Life Sci.* **2013**, *71*(2), 205–228.

73. Sinz, A. Cross-Linking/Mass Spectrometry for Studying Protein Structures and Protein–Protein Interactions: Where Are We Now and Where Should We Go from Here? *Angew. Chemie Int. Ed.* **2018**, *57*(22), 6390–6396.

74. Chavez, J. D.; Lee, C. F.; Caudal, A.; Keller, A.; Tian, R.; Bruce, J. E. Chemical

Crosslinking Mass Spectrometry Analysis of Protein Conformations and Supercomplexes in Heart Tissue. *Cell Syst.* **2018**, *6*(1), 136-141.e5.

75. Moresco, J. J.; Carvalho, P. C.; Yates, J. R. Identifying components of protein complexes in *C. elegans* using co-immunoprecipitation and mass spectrometry. *J. Proteomics* **2010**, *73*(11), 2198–2204.

76. Morris, J. H.; Knudsen, G. M.; Verschueren, E.; Johnson, J. R.; Cimermancic, P.; Greninger, A. L.; Pico, A. R. Affinity purification–mass spectrometry and network analysis to understand protein-protein interactions. *Nat. Protoc.* **2014**, *9*(11), 2539–2554.

CHAPTER 2: Profiling Thimet Oligopeptidase-Mediated Proteolysis in *Arabidopsis thaliana*[1]

2.1 Introduction

Plants have evolved various strategies to maintain protein homeostasis (proteostasis), balancing protein synthesis, assembly, and degradation[1]. To limit protein misfolding and aggregation under stress conditions, cells make use of their proteostasis network comprised of molecular chaperones, proteasome system, and autophagy[2]. Proteolysis is recognized as critical for the plant's defense against pathogens and adaptive responses to environmental stress[3,4]. Cytosolic proteolytic cascades complete the processing of proteasome-released peptides, while organelle-localized proteolytic components (*e.g.* chloroplastic and mitochondrial proteolytic cascades) have a wide range of functions including the removal of damaged organellar proteins[5,6]. Apart from the process of complete proteolysis, controlled proteolysis, the process by which proteases/peptidases generate proteins/peptides with biological functions, is well understood in metazoans[7–9]. Animal oligopeptidases with limited substrate specificities are critical factors in generating biologically active peptides with roles in growth, development, and defense mechanisms through modulation of signal transduction pathways[10,11]. Controlled proteolysis in plants is less explored and requires more investigation to reveal the identity of the protease activities involved and its bioactive peptide products.

[1] Reprinted with permission from Iannetta, A. A.; Rogers, H. T.; Al-Mohanna, T.; O'Brien, J. N.; Wommack, A. J.; Popescu, S. C.; Hicks, L. M. Profiling thimet oligopeptidase-mediated proteolysis in *Arabidopsis thaliana*. *Plant J.* **2021**, *106*(2), 336-350.

Plant stress caused by pathogens or though abiotic means (*e.g.* drought or temperature) reduces agricultural yields, causing substantial economic losses while reducing food security at the global level[12,13]. It is critical to recognize how plants perceive stress signals to elicit responses for survival. Activation of the plant innate immune response through the recognition of pathogen-associated molecular patterns (PAMPs) by pattern recognition receptors (PRRs) and pathogenic effectors by nucleotide-binding leucine-rich repeat (LRR) proteins is well documented[14–18]. Endogenous plant peptidases and their peptide products also play an important role in the signaling of plant immune processes[19–21]. For example, the plant elicitor peptide 1 has increased release from metacaspase-mediated proteolysis of PROPEP upon plant wounding and activates defensin gene transcription[19,20,22,23]. However, uncovering the role of peptidases in plant stress response is limited by the ability to identify substrates, characterize features of substrate recognition, as well as understand how cleavage products elicit downstream signaling effects due to the perceived stress[24,25].

Thimet oligopeptidases (TOPs) are zinc-dependent peptide hydrolases with a conserved HEXXH active site motif[26,27]. These metallopeptidases are critical components in plant response to oxidative stress triggered by pathogens or abiotic factors and are required for a fully functioning immune response to certain pathogens[28,29]. Salicylic acid (SA), a plant defense hormone, binds TOPs to inhibit their activities; plants with impaired expression of TOP genes have increased susceptibility to pathogens that activate effector-triggered immunity (ETI)[28,29]. The Arabidopsis genome contains two genes encoding TOPs: *TOP1* and *TOP2*[30]. TOP1 (AT5G65620, also known as organellar oligopeptidase, OOP) contains an N-terminal signal peptide that mediates its localization to the chloroplast and mitochondria[29,31]. It was shown to cleave presequences containing 8-23 amino acids *in vitro* and is hypothesized to act downstream

of organellar proteases for intra-organelle peptide degradation and organelle import processing[31]. It is proposed that TOP1 may be involved in an array of proteolytic processes due to its broad enzymatic specificity[31]. TOP2 (AT5G10540, also known as cytosolic oligopeptidase, CyOP) is located in the cytosol, where it is predicted to act downstream of the 20S proteasome, degrading proteasome-generated peptides during oxidative stress[29,31,32]. Prior evidence suggests that TOP1 and TOP2 have functional overlap in ETI and programmed cell death (PCD)[28,29,31,32]. Both oligopeptidases are required for plant defense to avirulent strains of *Pseudomonas syringae* through the activation of the resistance proteins RPS2 or RPS4, and both are necessary to regulate the PCD[29]. In a current model, TOPs sustain interconnected organelle and cytosol proteolytic pathways that regulate the ETI oxidative burst and plant resistance to pathogens through SA, reactive oxygen species (ROS), and antioxidants[28]. Therefore further differential characterization of the TOP1/TOP2 null mutant with the wild type will allow for the delineation of TOP peptide substrates and their specificity in substrate selection, which will generate insights into the relevance of TOP-mediated proteolytic pathways in proteostasis.

Herein, a quantitative mass spectrometry-based peptidomics approach was implemented to characterize the *Arabidopsis thaliana* plant peptidome in the context of TOPs (Fig. 2.1). A comparison between wild type (Col-0) and *top1top2* null mutant revealed putative direct and indirect TOPs substrates *in vivo*; direct substrates were validated via *in vitro* enzyme assays. Ten novel TOPs substrates were identified, revealing a putative cleavage motif.

2.2 Materials and Methods

2.2.1 Plant Growth

Seeds for *Arabidopsis thaliana* (Arabidopsis) ecotype Columbia were obtained from the Arabidopsis Biological Resource Center (Columbus, OH). Seeds for *top1top2* used in this study

were obtained by growing *top1top2* plants[29] with a 12/12 h day/night photoperiod; plants grown from new seed batches were verified using quantitative PCR (Fig. S13, Supporting Information, 10.1111/tpj.15165) with SYBR Green Master Mix (Applied Biosystems, Waltham, MA, USA) and gene-specific primers. Seeds were sown individually in 32-hole plug trays containing soil. After stratification for 2 weeks with a 14/10 h day/night photoperiod at 20 °C, plants were cultivated in a growth chamber for 6 weeks with a 8/16 h day/night photoperiod at 22 °C and 75% relative humidity with 140 µmol/m^2s photon flux density. Rosette leaves from 6-week-old plants were excised and immediately frozen under liquid nitrogen before storage at -80 °C.

2.2.2 Peptide Extraction

Three biological replicates were used for each genotype (*i.e.* Col-0 and *top1top2* mutants). Rosette leaf tissue (4 g) was ground under liquid N_2 to a fine powder before homogenization in 4 mL of 10% trichloroacetic acid (TCA) in acetone. The homogenate was incubated at -20 °C for 30 min before centrifugation at 15,000 *g* for 5 min at 4 °C. The supernatant was collected, and an additional 4 mL of 10% TCA in acetone was added to the ground rosette leaves before homogenization via vortexing, centrifugation at 15,000 *g* for 5 min at 4 °C, and combination of the supernatants. Extracted peptides were dried under N_2 and resuspended in 1 mL of 5 mM ammonium formate, pH 2.7 with 20% acetonitrile.

2.2.3 Strong Cation Exchange Solid-Phase Extraction

The isolation of peptides from small molecules was performed using 100 mg/1.0 mL Hypersep SCX cartridges (Thermo Fisher Scientific, Waltham, MA) held in an SPE 24-position vacuum manifold (Phenomenex, Torrance, CA) at a flow rate of 1 drop/s. Resin was first pre-eluted using 1 mL of 500 mM ammonium formate, pH 3 with 20% acetonitrile before equilibration with 1 mL of 5 mM ammonium formate, pH 2.7 with 20% acetonitrile. Samples

were loaded onto the cartridges in two passes and then washed using 2 mL of 5 mM ammonium formate, pH 2.7 with 20% acetonitrile. Peptides were eluted using 1 mL of 500 mM ammonium formate, pH 3 with 20% acetonitrile and dried by vacuum centrifugation. Peptides were resuspended in 1 mL of 0.1% trifluoroacetic acid (TFA).

2.2.4 Reversed Phase Solid-Phase Extraction

Desalting of samples was performed using 50 mg/1.0 mL Sep-Pak C18 cartridges (Waters, Milford, MA) held in an SPE 24-position vacuum manifold (Phenomenex) at a flow rate of 1 drop/s. Resin was first pre-eluted using 1 mL of 50% acetonitrile/0.1% TFA before equilibration with 1 mL of 0.1% TFA. Samples were acidified to pH 3 using 10% TFA, loaded onto the cartridges in two passes, and then washed using 1 mL of 0.1% TFA. Peptides were eluted using 1 mL of 50% acetonitrile/0.1% TFA and dried under vacuum centrifugation.

2.2.5 Peptide Quantitation

Samples were resuspended in 200 µL of 5% acetonitrile/0.1% TFA, and a 20 µL aliquot for peptide quantitation was dried under vacuum centrifugation. These dried aliquots were resuspended in 20 µL of 100 mM HEPES, pH 8.5 before peptide concentrations were estimated using the Pierce Quantitative Colorimetric Peptide Assay (Thermo Fisher Scientific) according to the manufacturer's protocols. Based on these results, peptide concentrations in each experiment were normalized across replicates with additional 5% acetonitrile/0.1% TFA.

2.2.6 LC-MS/MS Analysis

Samples were analyzed using an Acquity UPLC M-Class System (Waters) coupled to a Q Exactive HF-X mass spectrometer (Thermo Fisher Scientific). Mobile phase A consisted of water with 0.1% formic acid (Thermo Fisher Scientific) and mobile phase B was acetonitrile with 0.1% formic acid. Injections (2 µL) were made to a Symmetry C18 trap column (100 Å,

5µm, 180µm x 20 mm; Waters) with a flow rate of 5 µL/min for 3 min using 99% A and 1% B. Peptides were then separated on a HSS T3 C18 column (100 Å, 1.8µm, 75µm x 250 mm; Waters) using a linear gradient of increasing mobile phase B at a flow rate of 300 nL/min. Mobile phase B increased from 5% to 40% in 90 min before ramping to 85% in 5 min, where it was held for 10 min before returning to 5% in 2 min and re-equilibrating for 13 min. The mass spectrometer was operated in positive polarity and the Nanospray Flex source had spray voltage floating at 2.1 kV, capillary temperature at 320 °C, and funnel RF level at 40. MS survey scans were collected with a scan range of 350 – 2000 *m/z* at a resolving power of 120,000 and an AGC target of 3×10^6 with a maximum injection time of 50 ms. A top 20 data-dependent acquisition was used where HCD fragmentation of precursor ions having +2 to +7 charge state was performed using a normalized collision energy setting of 28. MS/MS scans were performed at a resolving power of 30,000 and an AGC target of 1×10^5 with a maximum injection time of 100 ms. Dynamic exclusion for precursor *m/z* was set to a 10 s window.

2.2.7 Database Searching and Label-Free Quantification

Acquired spectral files (*.raw) were imported into Progenesis QI for proteomics (Waters, version 2.0). Peak picking sensitivity was set to the maximum of five and a reference spectrum was automatically assigned. Total ion chromatograms were then aligned to minimize run-to-run differences in peak retention time. Each sample received a unique factor to normalize all peak abundance values resulting from systematic experimental variation. Alignment was validated (>80% score) and a combined peak list (*.mgf) was exported out of Progenesis for peptide sequence determination by Mascot (Matrix Science, version 2.5.1; Boston, MA). Database searching was performed against the *Arabidopsis thaliana* UniProt database (https://www.uniprot.org/proteomes/UP000006548, 39,359 canonical entries) with sequences for

common laboratory contaminants (https://www.thegpm.org/cRAP/, 116 entries) appended. Target-decoy searches of MS/MS data used "None" as the enzyme specificity, peptide/fragment mass tolerances of 15 ppm/0.02 Da, and variable modifications of N-terminus acetylation, C-terminus amidation, and methionine oxidation. Significant peptide identifications above the identity or homology threshold were adjusted to less than 1% peptide false discovery rate using the embedded Percolator algorithm[33]. Mascot results (*.xml) were imported to Progenesis for peak matching. Identifications with a Mascot score less than 13 were removed from consideration in Progenesis before exporting both "Peptide Measurements" and "Protein Measurements" from the "Review Proteins" stage.

2.2.8 Data Analysis and Statistics

Data were parsed using custom scripts written in R for pre-processing and statistical analysis (https://github.com/hickslab/QuantifyR). The "Peptide Measurements" data contain peak features with distinct precursor mass and retention time coordinates matched with a peptide sequence identification from the database search results. Some features were duplicated and matched with peptides having identical sequence, modifications, and score, but alternate protein accessions. These groups were reduced to satisfy the principle of parsimony and represented by the protein accession with the highest number of unique peptides found in the "Protein Measurements" data for this experiment, else the protein with the largest confidence score assigned by Progenesis. Some features were also duplicated with differing peptide identifications and were reduced to just the peptide with the highest Mascot ion score. An identifier was created by joining the protein accession of each peptide to the identified peptide sequence. Each dataset was reduced to unique identifiers by summing the abundance of all contributing peak features (i.e., different peptide charge states and combinations of variable modifications). Identifiers were

represented by the peptide with the highest Mascot score in each group. The molecular weight distributions of the Arabidopsis proteome and proteins from which the Arabidopsis peptidome derived from were compared using a two-sided, equal variance *t*-test. For the differential peptidome analysis, identifiers were removed if there was not at least one condition with > 50% nonzero values across the Progenesis-normalized abundance columns. Values were $\log_2$-transformed and a conditional imputation strategy was applied using the imp4p package[34], where conditions with at least one nonzero value had missing values imputed using the *impute.rand* function with default parameters. For cases where a condition had only missing values, the *impute.pa* function was used to impute small numbers centered on the lower 2.5% of values in each replicate. Statistical significance was determined using a two-tailed, equal variance *t*-test. Fold change was calculated by the difference of the mean abundance values between conditions being compared. Only observations with p-value < 0.05 and $|\log_2$-transformed fold change$| \geq 1$ after a two-sided, equal variance *t*-test were considered significantly different. Gene ontology (GO) annotations were assigned from UniProt.

2.2.9 Motif Analysis

In preparation for sequence logo visualization, data were parsed using custom scripts written in R. Peptides were filtered for those significantly increasing in Col-0 (p-value < 0.05, $\log_2$-transformed fold change ≤ -1). Based on the originating protein sequence, the peptide termini were extended four amino acids; for peptides too close to either protein termini, peptides were extended until the protein terminus was reached. Each elongated peptide was truncated into two smaller peptides containing the extended amino acids and either the first or last four amino acids of the original peptide sequence. These truncated peptides were filtered for those with eight residues to satisfy the input condition requiring peptides of the same length. For motif analysis,

sequence logo visualizations were performed using pLOGO, Positions with significant residue presence are depicted as amino acid letters sized above the red line[35].

2.2.10 TOPs Expression and Purification

cDNA sequences encoding TOP1 (AT5G65620) and TOP2 (AT5G10540) were cloned into the NheI/SalI site of the expression vector pTA. The recombinant vectors pTA-TOP1 and pTA-TOP2 were transformed into competent *E. coli* NiCo21 (DE3) cells (New England Biolabs, Ipswich, MA). *E. coli* cultures were grown with shaking at 200 rpm and 37 °C in LB Broth media in the presence of 50 mg/L kanamycin. At OD_{600} 0.5-0.6, the cultures were induced with 100 mg/L isopropyl β-D-1-thiogalactopyranoside for 4 h and cells were collected. Pellets were thawed on ice and lysed in 10 mL of B-PER complete bacterial protein extraction reagent (Thermo Fisher Scientific). The lysed cells were centrifuged at 2000 x*g* for 30 minutes at 4 ˚C, protein extracts were concentrated in an Amicon Stirred Cell 50 mL with 30 kDa ultrafiltration discs (MilliporeSigma, Burlington, MA), and purified via batch HisPur Ni-NTA Resin (Thermo Fisher Scientific) following the manufacturer's protocol. His-tagged proteins were aliquoted and stored in 20% glycerol at -80 °C. TOP fractions were collected, concentrated, and loaded onto a Superdex 200 size exclusion column using 500 mM NaCl in 50 mM Tris, pH 8.0, at a flow rate of 0.5 mL/min. The purity of TOPs was judged by the presence of a single band in SDS-PAGE. The concentration of purified TOPs was determined by UV absorption peak at 280 nm and Bradford assay. Activities of all enzymes were confirmed in an *in vitro* fluorescence enzyme assay (λ_{ex} = 345 nm, λ_{em} = 405 nm) using a quenched fluorogenic known metallopeptidase synthetic substrate with the sequence of Mca-PLGPK(Dnp)-OH (Mca – 7-methoxycoumarin-4-acetyl, Dnp – 2,4-dinitrophenol).

2.2.11 Solid-Phase Peptide Synthesis

Peptides were synthesized either by SynPeptide (Beicai Pudong New Area, Shanghai, China), or via a semi-automated flow chemistry instrument built in-house[36]. In brief, synthesis used 200 mg of 2-chlorotrityl chloride resin (200-400 mesh) that was pre-loaded with the C-terminal amino acid (0.80 mmol/g). Fmoc-protected amino acids (1 mmol) were solubilized in 2.5 mL of freshly prepared 0.38 M hexafluorophosphate benzotriazole tetramethyl uranium in dimethylformamide (DMF). Rapid Fmoc solid-phase peptide chemistry is begun after the instrument is prepared with appropriate levels of DMF for washing (40 mL per residue) and 20% piperidine in DMF for Fmoc deprotection (7 mL per residue), heating the water bath to 60 ºC for immersion of the sealed reactor, programing the pump flow rate to 20 mL/min, and assembling the reactor containing pre-weighed resin. To activate the amino acid solutions for coupling, 450 μL of *N*,*N*-diisopropylethylamine (or 250 μL with His, Cys, or Trp) is added and gently mixed. After syringe pump administration of the activated coupling solution at 6 mL/min, the resin is immediately washed at 20 mL/min with DMF for 20 sec. The Fmoc protecting group was removed with 20% piperidine in DMF for 30 sec and the resin was washed with DMF for 1 min. Following the coupling and deprotection of the final amino acid, the reactor was disassembled and the resin-bound peptide was eluted with a 10 min incubation at 60 ºC using 10 mL of 94% TFA/4% triisopropylsilane/2% H2O. The crude peptide solution was filtered and the volume was reduced to 5 mL under N_2 gas. The crude peptide solution was cooled to 4 ºC before precipitating with cold diethyl ether. Following centrifugation 1000 x*g* for 10 min at 4 ºC to collect the crude peptide precipitate, the white solid was resuspended in cold diethyl ether and the centrifugation was repeated. The pellet was dried *in vacuo* to deliver crude peptide. The peptide, dissolved in 20% acetonitrile/0.1% TFA, was further purified by preparative HPLC.

2.2.12 *In Vitro* Enzyme Assay

Synthesized peptides were solubilized in 500 µL of 100 mM NaCl in 50 mM Tris, pH 7.5. To initiate the enzyme assay, either TOP1, $^{\Delta SP}$TOP1, or TOP2 was added at a peptide:TOP ratio of 10:1. The reaction mixture was incubated at 23 °C for 30 min. Reaction mixtures were desalted using 50 mg/1.0 mL Sep-Pak C18 cartridges (Waters) as described above.

2.2.13 ESI-MS Analysis

Samples were resuspended in 100 µL of 50% methanol/0.1% formic acid. Peptides were directly infused via ESI on a Q Exactive HF-X Hybrid mass spectrometer (Thermo Fisher Scientific) for intact mass analysis. Samples were injected at a flow rate of 10 µL/min and full MS scans were analyzed in the Orbitrap. The mass spectrometer was operated at a resolving power of 120,000, positive polarity, spray voltage of 3 kV, with 150–2000 *m/z* range, and collecting 100 scans/sample for averaging.

2.2.14 Data Availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository[37] and can be accessed with the dataset identifier PXD019928 and 10.6019/PXD019928.

2.3 Results and Discussion

It is imperative to examine peptidomes in an *in vivo* context, as *in vitro* peptidase cleavage assays may reveal activity against an array of putative substrates that are otherwise irrelevant. Plant peptidomics is a nascent research area, mainly focusing on plant development, thus requiring method development and optimization for robust quantitative analysis[24,25,38]. Peptidome extractions are often accompanied by unwanted metabolites and other non-peptidic features[39]. These interferences can induce peptide ion suppression; thus, their efficient removal is

essential to maximize peptidome identification[39]. To profile the Arabidopsis peptidome, peptide extracts were directly analyzed intact (similar to top-down proteomics). This enabled direct detection of multiple peptidoforms: peptides containing the same amino acid sequence with differing termini post-translational modifications (PTMs) due to differential processing or cleavages. Because these peptides are derived from *in vivo* proteolysis and analyzed without further digestion, MS abundances are solely indicative of these peptides and are not reflective of intact protein abundances (as is the case in traditional bottom-up proteomics experiments).

2.3.1 Defining the *Arabidopsis* peptidome

Extraction of Arabidopsis rosette leaves was optimized using 10% trichloroacetic acid (TCA) in acetone and yielded 16.2 ± 3.2 ng peptide/mg plant material, a four-fold increase relative to the same extraction in water. TCA in acetone extracts were assessed via LC-MS/MS across six technical replicates of wild type (Col-0) plants. Label-free quantification resulted in 1032 unique peptides identified, originating from 361 proteins (Table S1, Supporting Information, 10.1111/tpj.15165). Peptide abundances had a median coefficient of variation (CV) of 26% (Fig. S1A, Supporting Information, 10.1111/tpj.15165), and an average linear Pearson correlation coefficient of 0.90 (Fig. S1B, Supporting Information, 10.1111/tpj.15165).

The median peptide length detected within the peptidome is 15 residues with 99% of peptides having less than 39 residues (Fig. 2.2A). These peptides originate from proteins with a median molecular weight of 33 kDa (Fig. 2.2B). Plant peptides are typically characterized by having less than 50 amino acids[40] and the Arabidopsis proteome has a 40 kDa median MW distribution, suggesting a slight bias in coverage in the peptidomics dataset.

Peptidome amino acid composition was compared to the full Arabidopsis proteome (Fig. 2.2C) under these particular extraction conditions to discern possible compositional differences. Within the peptidome, Ala had the largest increase in relative frequency (+5.6%) while Leu had the largest decrease in relative frequency (-4.5%) compared to the proteome. Cys and Trp have low relative frequency in the Arabidopsis proteome (Cys: 1.9%; Trp: 1.2%), however their representation decreased further in the peptidome (0.10% and 0.045%, respectively). Given the importance of these residues in development and reproduction[41,42], we examined if coverage within the peptidome was limited due to possible residue modifications (either through *in vivo* modification or artificial oxidation during sample preparation). The peptidome dataset was examined for the presence of common Cys PTMs including *S*-nitrosylation, *S*-thiolation, and *S*-acylation (Table S2, Supporting Information, 10.1111/tpj.15165). The most abundant Cys PTM identified was *S*-sulfonylation (24 instances, Fig. S2A and S2B, Supporting Information, 10.1111/tpj.15165), an irreversible oxidation state representative of loss of protein function and degradation[43]. Cys-rich peptides stabilized by disulfide bonds may elude identification via the bioinformatics approach implemented herein. For Trp, a possible degradation route includes oxidation to oxindolylalanine, *N*-formylkynurenine (doubly oxidized), and kynurenine (loss of CO after being doubly oxidized)[44,45]. Altogether nine instances of these modifications were detected in the peptidome (Table S3, Supporting Information, 10.1111/tpj.15165), displaying that even with the consideration of PTMs, Trp is rarely detected in the identified Arabidopsis peptidome.

To examine localization within originating proteins, the most N- or C-terminal residue in each identified peptide was mapped onto the full-length protein sequence. This revealed that ~6% and ~10% of identified peptides were N- and C-terminal peptides, respectively (Fig. 2.2D).

Meanwhile the first/last residues of other detected peptides were found more frequently within the starting and end 15% of the intact protein sequence (Fig. 2.2E). The presence of these protein terminal-containing peptides was expected. Most mitochondrial and chloroplastic proteins contain N-terminal presequences and transit peptides that allow for organellar transportation and import[46,47], which are subsequently cleaved and further degraded via peptide processing pathways[48]. In plants, acetylated N-termini can be recognized as degrons, where slow or incorrect folding leads to their exposure and degradation[49]. Both N- and C-termini are often solvent exposed, making them likely targets for degradation in proproteins, resulting in peptides derived from protein termini[50].

2.3.2 Differential Peptidomics Reveals Putative TOPs Substrates

Peptidase activity nullified via genetic or chemical approaches decreases peptidase products and increases the accumulation of substrates[51]. Thus, the absence of TOP1 and TOP2 *in vivo* would result in an increased abundance of peptide substrates in the *top1top2* knock-out mutant vs. Col-0 WT. Likewise, a representative increased abundance of products derived from these substrates would be expected in Col-0 WT vs. the mutant (Fig. 2.3A). The double mutant was used in comparison with WT instead of *top* single mutants due to their documented shared roles in ETI and PCD[28,29,31,32]. While TOP1 and TOP2 have different subcellular localizations, their functional overlap and high sequence similarity (92%) suggest a potential for redundant substrates[26,29]. The use of the double mutant also ensures the detection of TOP substrate accumulation compared to the WT.

To determine potential TOPs substrates, three biological replicates across each genotype were extracted and 1111 unique peptides from 369 proteins were quantified (Table S4, Supporting Information, 10.1111/tpj.15165). There were 1071 peptides shared across the two

genotypes, with 9 peptides unique to Col-0 and 31 peptides unique to *top1top2*. Of these 1111 peptides, 373 had a significant fold change (Fig. 2.3B), with 350 peptides more abundant in *top1top2* (representing potential TOPs substrates, Table S5, Supporting Information, 10.1111/tpj.15165) and 23 more abundant in the WT (representing potential cleaved products of TOP substrates, Table S6, Supporting Information, 10.1111/tpj.15165).

Peptides found more abundant in WT plants were analyzed to examine TOP cleavage specificity as at least one termini of these products may result from TOP cleavage. Motif analysis was performed on these prospective TOP recognition sites after extending each peptide's terminus to include the surrounding amino acids[35]. This resulted in a significant enrichment of Gly/Lys at P2, Met at P1, Gly/Ala at P1', and Ala/Val at P2' (Fig. 2.4A). In *Salmonella typhimurium*, oligopeptidase A, another zinc metallopeptidase in the M3 enzyme family, has been shown to hydrolyze oligopeptides with broad specificity, but Gly and Ala commonly occur as P1 or P1' residues[52], while human TOP prefers hydrophobic residues in positions P1, P2, and P3'[53]. Based on this, it is proposed that TOPs have broad substrate recognition specificities within and between species.

A previous study analyzed the WT peptidome and qualitatively compared it to plant lines lacking a combination of the chloroplast and mitochondria localized oligopeptidases TOP1, presequence protease 1 (PreP1, AT3G1970), and presequence protease 2 (PreP2, AT1G49630)[54]. PreP (cleaving peptides ranging between 10 and 65 residues) and TOP1 (cleaving peptides ranging between 8 and 23 residues) function together to degrade their substrates down to 3-7 residues, where they are further broken down by aminopeptidases[31,55]. Their analysis identified 180 unique peptides across the WT, *prep1prep2* double knockout, and *prep1prep2top1* triple knockout mutant lines, whereas the method presented herein identified 1111 unique peptides, an

improvement in Arabidopsis peptidome coverage. Across the Kmiec *et al.* dataset and the differential analysis described above (Table S4, Supporting Information, 10.1111/tpj.15165), there was an overlap of 32 peptides (26 of these were chloroplastic), but this small overlap could be attributed to the low peptidome coverage in Kmiec *et al.* Within the WT peptidome, the authors examined the localization of peptides by mapping them onto their source proteins. Resembling the peptide localization analysis herein, ~6% were also determined to be N-terminal peptides[54]. A stress response attributed to peptide accumulation was observed in *prep1prep2top1* and not the *prep1prep2* mutant. In the triple mutant, chloroplastic peptides (25% of identified peptides) accumulated at levels above the WT, supporting the role of PreP and TOP1 in organellar peptide degradation[54]. To compare these results to our differential peptidome analysis, Gene Ontology (GO) cellular component terms were applied to the peptides more abundant in *top1top2*. Among these peptides, 55% were chloroplastic (Fig. S3, Supporting Information, 10.1111/tpj.15165), mirroring this highlighted study and further supporting the role of TOP1 in the chloroplast. There was also an accumulation of cytosolic peptides in *top1top2*, which can be attributed to the knockout of the cytosolic TOP2, an oligopeptidase that is present/active in the *prep1prep2top1* where cytosolic peptides showed WT levels of accumulation.

2.3.3 Validation of TOPs Substrates and Peptidase Cleavage Specificity

Peptides significantly more abundant in the *top1top2* mutant *in vivo* represent putative direct substrates or pleotropic effects displaying increased accumulation due to the loss of TOP activity[24]. Therefore *in vitro* validation is required to differentiate between these outcomes. We elected to take an unbiased approach, prioritizing peptides for synthesis using the following criteria: 1) consideration of previously determined TOP substrate length (8-23 amino acids), 2) the most significant *in vivo* fold changes, and 3) detection of increased *top1top2* mutant peptides

from the same protein with overlapping recognition sequences. To determine direct substrates and cleavage specificity of TOPs, *in vitro* enzyme assays were conducted with seventeen synthetic peptides (Table 2.1).

Substrate candidates were incubated with recombinant TOP1, $^{\Delta SP}$TOP1 (lacking N-terminal signal peptide), and TOP2 and analyzed via ESI-MS. To validate putative TOPs substrates, mass spectra of reaction mixtures were examined to identify unique ions detected in the enzyme-treated samples, indicating cleaved peptide products. The exact masses resulting from high resolving power MS were used to determine peptidase cleavage locations. The detection of the cleaved peptide products greatly relies on their ionization efficiencies, which is dictated by their amino acid compositions. Since it is not expected that all potential peptide products will be detected in these analyses, the identification of one product of a cleavage represents the presence of that cleavage. As an additional negative control, these assays were performed on 6 peptides that did not meet the fold change and significance thresholds to be considered more abundant in *top1top2* over WT. The presence of only the full-length peptides in all reaction mixtures provides evidence of TOP cleavage specificity and increases confidence in assay results (Fig. S4, Supporting Information, 10.1111/tpj.15165).

Of the seventeen candidate peptides, ten were identified as cleaved by TOP1, $^{\Delta SP}$TOP1, and/or TOP2, representing direct TOP substrates (Table 2.2). There were no TOP cleavages detected for some peptides that had large fold changes in *top1top2* compared to WT, demonstrating the ability of our *in vitro* validation to distinguish true substrates from pleiotropic effects. The greatest number of identified substrates and cleavage sites was determined for $^{\Delta SP}$TOP1 (10 confirmed substrates, 29 total cleavages), while TOP1 (6 confirmed substrates, 9 total cleavages) had slightly more cleavages than TOP2 (5 confirmed substrates, 7 total

cleavages). This difference could be attributed to the signal peptide, as it has been shown in *Pisum sativum* where the signal peptide region of the mature stromal processing peptidase (SPP) strongly reduces the ability of the enzyme to interact with its substrates; removal of the N-terminal region of the signal peptide may initiate the mechanism of SPP folding and activation[56,57]. It is not surprising that TOP1 and TOP2 show differences in activity, as the crystal structures of $^{\Delta SP}$TOP1 and TOP2 reveal dramatic structural differences[31,58]. The major domains of $^{\Delta SP}$TOP1, which enclose its catalytic cavity, are located much closer to each other resulting in a tighter active site, which could contribute to its increased enzymatic activity[31,58]. Motif analysis was performed across all enzymes, revealing a preference for Gly/Leu and Ala/Gly residues in the P1 and P1' positions, respectively (Fig. 2.4B). Arabidopsis TOPs display a similar cleavage recognition motif to *S. typhimurium* oligopeptidase A, a well-studied peptidase in the same enzyme family as TOPs[52].

While TOPs are able to cleave these peptides *in vitro*, the subcellular localization of the originating proteins of these substrates should be examined to determine feasible interactions *in vivo* and define true TOP substrates. Of the 10 $^{\Delta SP}$TOP1-cleaved peptides, 9 (all peptides except SDDEHHFEASESGAS) are derived from proteins localized to the chloroplast or mitochondria. Of the 5 TOP2-cleaved peptides, 4 (all peptides except VMLRPASPGTGVIAGGA) are derived from cytosolic proteins.

These validated TOP substrates are derived from proteins involved in various cellular processes including photosynthesis (oxygen-evolving enhancer protein 3-2), glycolysis (phosphoglycerate kinase 1), protein folding (chaperonin 60 subunit β 1), biogenesis (30S ribosomal protein S5, eukaryotic translation initiation factor 5A-2, 60S ribosomal protein L28-1, cysteine synthase 1, bifunctional protein FolD 2), and antioxidant defense (catalase-2). These

processes are negatively affected by biotic and abiotic stressors, resulting in aberrant or damaged proteins[59,60]. Removal of these undesired proteins is crucial for maintaining homeostasis via energy conservation, and proteolysis is necessary for proper plant defense against pathogens and environmental stress[61]. For example, a well characterized proteolysis-driven repair cycle turns over protein complexes in the electron transport chain after photodamage; these same proteolytic mechanisms are used to promote energy conservation under adverse conditions[62,63]. The identification of these peptide substrates implicates TOPs in the preservation of these important cellular processes through proteolysis.

Besides their role in protein maintenance, stress-triggered controlled proteolysis results in the release of bioactive peptides[64,65]. Many of these peptides function as damage-associated molecular patterns (DAMPs), which promote defense gene expression and synthesis of defensive secondary metabolites[66–68]. For instance, the aspartic protease Constitutive Disease Resistance 1 generates systemin-like peptides and induces plant defense response and systemic acquired resistance (SAR)[69]. Within these responses, ROS are major signaling molecules that have repeatedly linked defense and proteolytic pathways[70,71]. Following biotic or abiotic stress, the environment of a cell becomes highly oxidative as a result of a burst in ROS synthesis in various cellular components, modifying proteins for degradation and prompting retrograde defense signaling[72–74]. Our previous work compared the redoxomes in Col-0 and *top1top2* after pathogen infection, and significant differences were detected regarding the identity and number of reversibly oxidized cysteines as well as the amplitude of time-dependent fluctuations in protein oxidation[75]. These results supported a determining role for TOPs in maintaining the proper level and dynamics of proteome oxidation during plant stress response that may link to the coordinated proteolysis observed in the current study[75].

Further characterization of these confirmed TOP substrates will need to be performed to determine if they are protein turnover degradation products or if they are bioactive defense signaling molecules. However, the proteins from which the validated peptides were derived are linked by their previously reported association with the Arabidopsis stress response. Arabidopsis plants lacking chaperonin 60 subunit β 1 (AT1G55490, originating protein of TOP peptide substrate AGVNKLADLVGVTLGPK, Fig. 2.5A) developed lesions on their leaves and express SAR in the absence of pathogens. These plants were also more susceptible to heat shock stress in comparison to wild type plants[76]. The suppression of eukaryotic translation initiation factor 5A-2 (AT1G26630, originating protein of TOP peptide substrate SDDEHHFEASESGAS, Fig. 2.5B) in transgenic Arabidopsis plants prevented the onset of PCD after inoculation with virulent *Pseudomonas syringae* pv. *tomato* DC3000, suggesting that this protein regulates PCD caused by bacterial infection[77]. Cysteine synthase 1 (AT4G14880, originating protein of TOP peptide substrate FDATRKEAEAM, Fig. S5, Supporting Information, 10.1111/tpj.15165) coordinates the regulation of Cys levels in the cytosol, which strongly affects plant response to abiotic and biotic stress[78]. Arabidopsis plants lacking this protein cannot regulate the generation and removal of ROS, which results in enhanced disease susceptibility against virulent and non-virulent *P. syringae* pv. *tomato* DC3000 strains[79]. 30S ribosomal protein S5 (AT2G33800, originating protein of TOP peptide substrate VMLRPASPGTGVIAGGA, Fig. S6, Supporting Information, 10.1111/tpj.15165) is an important constituent of the plastid 30S subunit and affects proteins involved in cold stress responses to mediate plant growth and development[80]. Catalase 2 (AT4G35090, originating protein of TOP peptide substrate MDPYKYRPA, Fig. S7, Supporting Information, 10.1111/tpj.15165) coordinates salicylic acid-mediated repression of auxin accumulation and jasmonic acid biosynthesis during plant defense. Arabidopsis plants lacking

catalase 2 display increased susceptibility to necrotrophic pathogens and its activity is required to promote autophagy-dependent PCD[81,82].

The other confirmed TOP peptide substrates were analyzed by examining their originating proteins in the context of their associations with the plant response to different stressors. The Arabidopsis genes for 60S ribosomal protein L28-1 (AT2G19730, originating protein of TOP peptide substrate AADKDQAVVLATT, Fig. S8, Supporting Information, 10.1111/tpj.15165) were downregulated 48 h after inoculation with *Agrobacterium tumefaciens*[83]. Oxygen-evolving enhancer protein 3-2 (AT4G05180, originating protein of TOP peptide substrate FQTIDNLDYAARSKSSPD, Fig. S9, Supporting Information, 10.1111/tpj.15165) and other oxygen-evolving enhancer proteins were downregulated in the Arabidopsis immune response, triggered by lipid transfection reagent treatment[84]. Following abscisic acid treatment and drought, the stress-induced protein KIN2 (AT5G15970, originating protein of TOP peptide substrate AFQAGQAAGKAE, Fig. S10, Supporting Information, 10.1111/tpj.15165) was determined to be upregulated[85]. Phosphoglycerate kinase 1 (AT3G12780, originating protein of TOP peptide substrate GKVLPGVIALDEAIPVT, Fig. S11, Supporting Information, 10.1111/tpj.15165) is a target for *S*-nitrosylation in Arabidopsis leaves inoculated with *P. syringae* pv. *tomato avrB* and is upregulated during induced cold stress. Bifunctional protein FolD2 (AT3G12290, originating protein of TOP peptide substrate MLLRNTVDGAKRVFGE, Fig. S12, Supporting Information, 10.1111/tpj.15165) was downregulated in distal Arabidopsis leaves following mechanical wounding or inoculation with *Oilseed rape mosaic virus*[86]. The connection between TOP peptide substrates and these responses could implicate TOP proteolytic activities in plant stress signaling, as previously shown[28,29]. Compared to Col-0 WT, Arabidopsis *top* mutants are more susceptible to *P. syringae* pv. *tomato* DC3000 *avrRpt2* and *avrRpm1* and

have an impaired response to chloroplastic oxidative stress induced by methyl viologen treatment[28,29].

Previous work using *in vitro* activity assays were performed and determined that TOP1 can degrade peptide fragments between 8-23 amino acids[31]. This aligns with the presented work, where the validated TOP substrates contained 9-18 amino acids. The analysis of TOP1 *in vitro* suggests that this peptidase may have a diverse range of substrates; around 10% of plant presequences are ≤23 residues, presenting potential TOP1 substrates[87]. Typical transit peptide lengths would be too large for this range, however, short peptide fragments (3-5 amino acids) were detected after *in vitro* transit peptide degradation with presequence protease (PreP) and TOP1, displaying a multi-step peptidolytic cascade[55,88]. In addition, TOP1 is proposed to be involved in the complete degradation of oxidized and damaged proteins to limit peptide accumulation after proteolysis. These peptides are derived from mitochondria and chloroplast proteases such as caseinolytic protease P (ClpP) and degradation of periplasmic proteins (Deg); most of these cleaved peptide fragments are between 5-20 amino acids[89,90]. Unlike TOP1, similar studies have not been reported with Arabidopsis TOP2. Its structure and localization are comparable to human TOP, which acts downstream of the proteasome pathway and favors peptide substrates with 6-17 amino acids[10,58,91,92].

2.4 Conclusion

In summary, a quantitative *in vivo* peptidomics method was developed and optimized to allow for the profiling of plant peptides. The displayed analytical reproducibility enables the discovery of proteolytic substrates by detecting changes in abundance following peptidase nullification. Using this approach, putative TOP substrates were identified from differential label-free quantitative peptidomics between Col-0 WT and *top1top2* knock-out mutant. From

these, ten substrates were validated using *in vitro* enzyme assays with heterologously expressed TOPs and synthetic candidate peptides. These enzymatic assays allowed for the identification of TOP cleavage sites, revealing a slight preference for hydrophobic amino acids surrounding the site of hydrolysis. Many of these substrates were derived from proteins associated with plant response to stress, strengthening the relationship between TOPs and plant stress signaling. Further studies of validated TOP substrates will determine if these serve simply as proteolysis products or have any relevant signaling function / bioactivity. Defining the roles of these substrates will generate valuable insights into how peptidases contribute to the plant immune response.

2.5 Tables

Table 2.1 Synthetic AtTOP peptide substrates tested in *in vitro* enzyme assays. All peptides were significantly increasing in the *top1top2* mutants and the fold change represents this increase, except the peptides from protein accessions AT5G15970and AT3G12290, where a longer or shorter peptide sequence was significantly increasing (SETNKNAFQAGQAAGKAE for AFQAGQAAGKAE and NTVDGAKRVFGE for MLLRNTVDGAKRVFGE).

UniProt Accession	Description	Sequence	Fold Change	TOP Substrate?
P21240	Chaperonin 60 subunit β 1	AGVNKLADLVGVTLGPK	57.3	Yes
P93014	30S ribosomal protein S5	VMLRPASPGTGVIAGGA	42.0	Yes
Q41932	Oxygen-evolving enhancer protein 3-2	FQTIDNLDYAARSKSSPD	29.0	Yes
Q93VP3	Eukaryotic translation initiation factor 5A-2	SDDEHHFEASESGAS	16.1	Yes
O82204	60S ribosomal protein L28-1	AADKDQAVVLATT	10.7	Yes
P25819	Catalase-2	MDPYKYRPA	7.1	Yes
P47998	Cysteine synthase 1	FDATRKEAEAM	4.0	Yes
P31169	Stress-induced protein KIN2	AFQAGQAAGKAE	3.3	Yes
Q9LD57	Phosphoglycerate kinase 1	GKVLPGVIALDEAIPVT	2.3	Yes
Q9LHH7	Bifunctional protein FolD 2	MLLRNTVDGAKRVFGE	2.1	Yes
Q9ZUH5	Short-chain dehydrogenase/reductase 2b	SVEEGASSPVR	1576.6	No
Q9SSS9	ATP synthase subunit Δ	KRQVIDDIVKSS	458.7	No
Q9LK36	Adenosylhomocysteinase 2	TKLTKDQSDYVSIPVEGPYKP	156.8	No
P17745	Elongation factor Tu	SIGSSVAKKYDEIDAAPEERA	51.9	No
Q9LD57	Phosphoglycerate kinase 1	SVGDLTSADLKGK	40.2	No
P19366	ATP synthase subunit β	VINEQNLAESK	24.9	No
F4JBY2	Transketolase	TVEPTTDSSIVDKSVNSIR	11.5	No

Table 2.2 Synthetic AtTOP peptide substrates found to be cleaved by TOPs after analysis with ESI-MS. The arrows represent identified sites of cleavage (ND: none detected). Cleaved peptide products that are bolded and underlined were uniquely detected in the enzyme-treated samples compared to the analysis of the bare synthetic peptide (Fig. 2.5, S4-11, Supporting Information, 10.1111/tpj.15165).

TOP1 Cleavage Sites	$^{\Delta SP}$TOP1 Cleavage Sites	TOP2 Cleavage Sites
AGVNKLADLVGVTL↓GPK	**AGVNKLADLVGVTL**↓GPK **AGVNKLA**↓**DLVGVTL**↓GPK **AGVNKL**↓**ADLVGVTL**↓GPK AGVNKLADLVG↓**VTL**↓GPK	ND
ND	**SDDEHHFE**↓ASESGAS	ND
FDATRKE↓AEAM	**FDATRKE**↓**AEAM** **FD**↓**ATRKE**↓**AEAM**	**FDATRKE**↓AEAM
VMLRPASPGTG↓VIAGGA **VMLRPASPG**↓**TGVIAGGA**	**VMLRPASPGTG**↓**VIAGGA** **VMLRPASPG**↓**TGVIAGGA** **VMLRPA**↓SPGTGVIAGGA **VML**↓RPASPGTGVIAGGA	**VMLRPASPG**↓TGVIAGGA
MDPYKY↓RPA	**MDPYKY**↓RPA	**MDPYKY**↓RPA
ND	**AADKDQAVVL**↓ATT AADKD↓**QAVVL**↓ATT	ND
ND	**FQTIDNLD**↓YAARSKSSPD	ND
AFQAGQAA↓GKAE **AFQAG**↓QAAGKAE	**AFQAGQAA**↓GKAE **AFQAG**↓QAAGKAE **AFQA**↓GQAAGKAE **AFQ**↓AGQAAGKAE **AF**↓QAGQAAGKAE	**AFQAGQAA**↓GKAE **AFQAG**↓QAAGKAE
ND	GKVLPGVIALD↓**EAIPVT** **GKVLPG**↓**VIAL**↓**DEAIPVT** **GKVLP**↓GVIALDEAIPVT	ND
MLLRNTVD↓**GAKRVFGE** **MLLR**↓NTVDGAKRVFGE	MLLRNTVDGAK↓**RVFGE** MLLRNTVDG↓**AKRVFGE** MLLRNTVD↓**GAKRVFGE** **MLLR**↓NTVDGAKRVFGE **MLL**↓RNTVDGAKRVFGE	MLLRNTVD↓**GAKRVFGE** **MLLR**↓NTVDGAKRVFGE

2.6 Figures

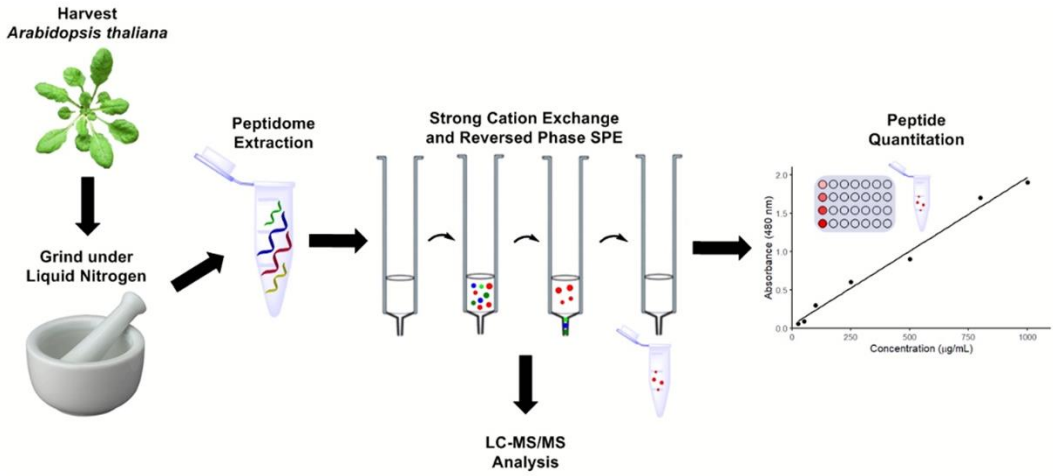


Figure 2.1 General workflow for label-free quantitative peptidomics. Six-week-old plant rosette leaves were ground under liquid nitrogen before extracting peptides with 10% TCA in acetone. Peptides were isolated from small molecules with SCX SPE and desalted with RP SPE before peptide quantitation. Peptide concentrations across replicates were normalized before liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis.

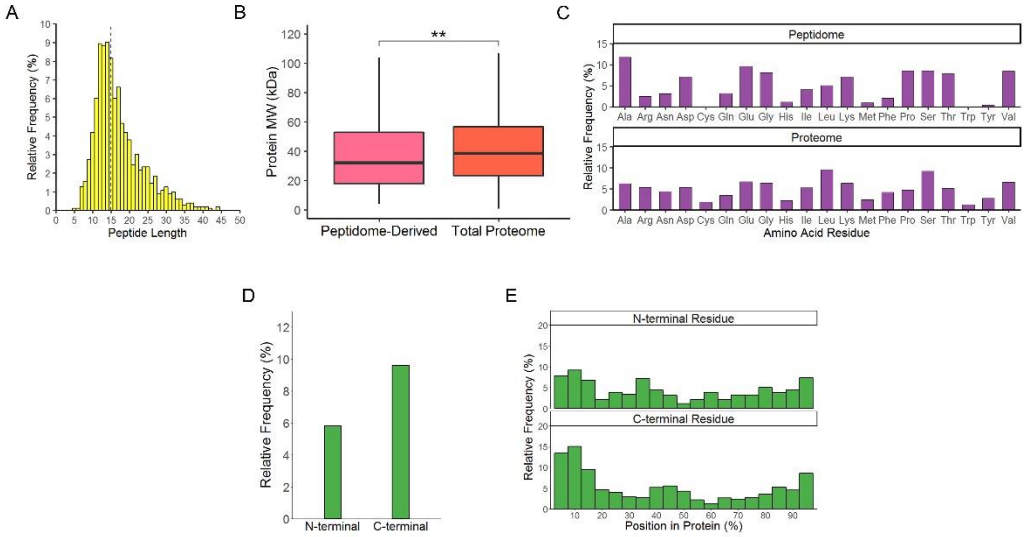


Figure 2.2 Characteristics of the Arabidopsis peptidome. (A) Histogram of the number of residues for identified peptides. The median peptide length is 15 residues (dashed line). (B) Box plots showing the distribution of protein molecular weights between proteins of identified peptides and the total Arabidopsis proteome (**: $p < 0.01$, following a two-sided, equal variance student's *t*-test). Outliers were removed for clarity of presentation. (C) Bar graph of the amino acid composition within the Arabidopsis peptidome and proteome. (D) Bar plots displaying the relative frequency of N- and C-terminal peptides identified in the Arabidopsis peptidome. (E) Histogram of the position of identified peptides within their proteins. This was determined twice based on whether the calculation made use of the position of the most N- or C-terminal residue.

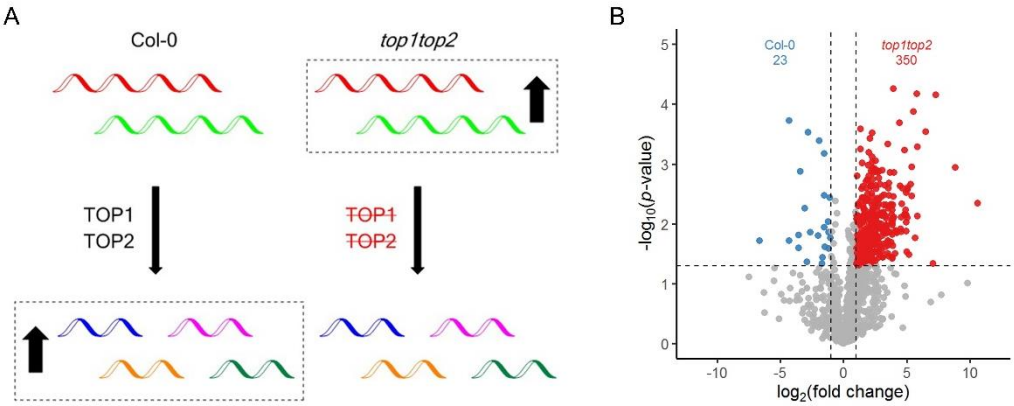


Figure 2.3 Differential analysis of the Arabidopsis global peptidome. (A) Explanation of expected results. The absence of AtTOPs should increase the abundance of their peptide substrates and decrease the abundance of peptide products derived from these substrates in *top1top2*. (B) Colored circles represent significantly changing peptides (p-value < 0.05, $|FC| \geq 2$) after a two-sided, equal variance t-test. Red indicates increasing peptides in the *top1top2* mutant line, which are potential TOP peptide substrates. Blue indicates significantly increasing peptides in the wild type, which are potential TOPs cleaved peptide products.

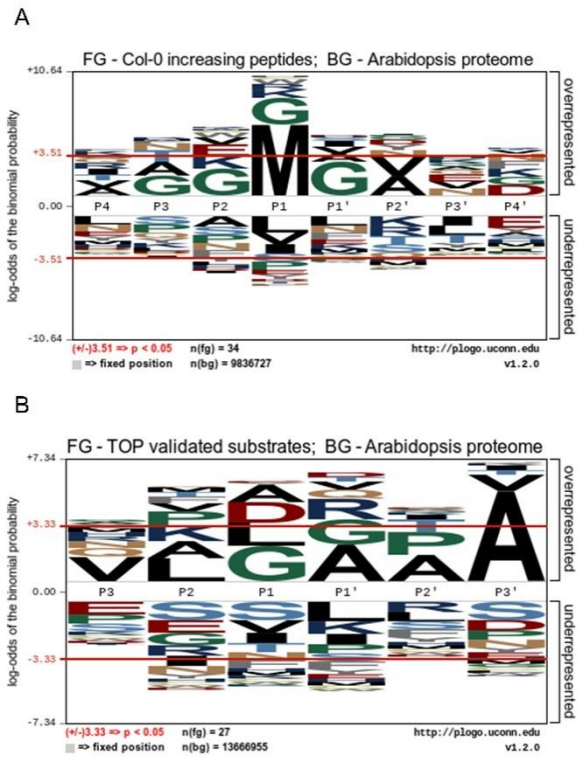


Figure 2.4 Sequence logo visualizations of TOP cleavage specificity using theoretical and determined TOP substrates. Positions with significant residue presence are depicted as amino acid letters sized above the red line [35]. (A) Motif analysis of the extended termini of peptides significantly increasing in the wild type plants. (B) Motif analysis of TOP cleavage sites on validated peptide substrates.

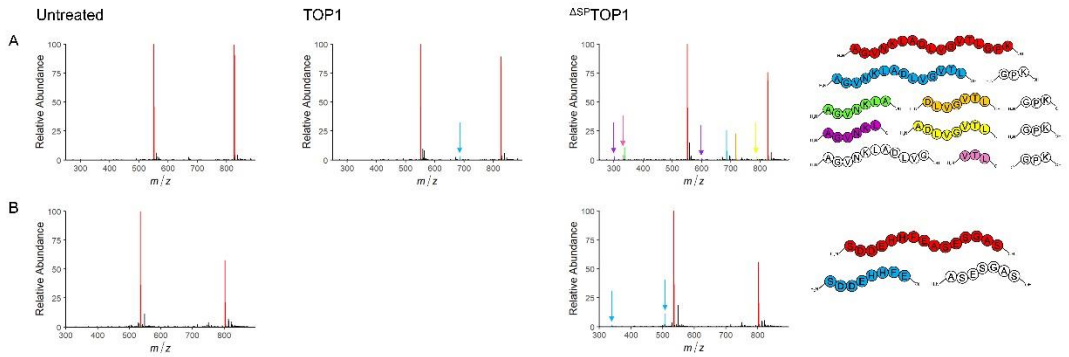


Figure 2.5 *In vitro* enzymatic assays. (A) The untreated AGVNKLADLVGVTLGPK peptide was detected in two charge states (red, *m/*z 826.49, +2 charge state and *m/*z 551.33, +3 charge state). One cleavage site was detected for TOP1. The N-terminal product (AGVNKLADLVGVTL, blue, *m/z* 685.41, +2 charge state) indicates this TOP cleavage site. Four cleavage sites were detected for $^{\Delta SP}$TOP1. The first is the one identified for TOP1. The N-terminal and C-terminal product pairs AGVNKLA/DLVGVTL (green, *m/z* 336.71, +2 charge state and orange, *m/z* 716.42, +1 charge state) and AGVNKL/ADLVGVTL (purple, *m/z* 301.19, +2 charge state and yellow, *m/z* 787.46, +1 charge state) indicate two cleavage sites, while the C-terminal product VTL (pink, *m/z* 332.22, +1 charge state) indicates the fourth cleavage site. (B) The untreated SDDEHHFEASESGAS peptide was detected in two charge states (red, *m/*z 802.81, +2 charge state and *m/*z 535.54, +3 charge state). One cleavage site was detected for $^{\Delta SP}$TOP1. The N-terminal product (SDDEHHFE, blue, *m/z* 508.19, +2 charge state) indicates this TOP cleavage site. All observed masses match with the theoretical peptide masses within 2 ppm mass error.

REFERENCES

1. Fernández-Fernández, M. R.; Valpuesta, J. M. Hsp70 chaperone: A master player in protein homeostasis. *F1000Research* **2018**, *7*, 1497.

2. Laskowska, E.; Kuczyńska-Wiśnik, D.; Lipińska, B. Proteomic analysis of protein homeostasis and aggregation. *J. Proteomics* **2019**, *198*, 98–112.

3. Shu, K.; Yang, W. E3 ubiquitin ligases: Ubiquitous actors in plant development and abiotic stress responses. *Plant Cell Physiol.* **2017**, *58*(9), 1461–1476.

4. Pogány, M.; Dankó, T.; Kámán-Tóth, E.; Schwarczinger, I.; Bozsó, Z. Regulatory Proteolysis in *Arabidopsis*-Pathogen Interactions. *Int. J. Mol. Sci.* **2015**, *16*(10), 23177–23194.

5. Kidrič, M.; Kos, J.; Sabotič, J. Proteases and their endogenous inhibitors in the plant response to abiotic stress. *Rev. Artic.* **2014**, *38*(1).

6. van Wijk, K. J. Protein Maturation and Proteolysis in Plant Plastids, Mitochondria, and Peroxisomes. *Annu. Rev. Plant Biol.* **2015**, *66*(1), 75–111.

7. Seidah, N. G.; Mayer, G.; Zaid, A.; Rousselet, E.; Nassoury, N.; Poirier, S.; Essalmani, R.; Prat, A. The activation and physiological functions of the proprotein convertases. *Int. J. Biochem. Cell Biol.* **2008**, *40*(6–7), 1111–1125.

8. Funkelstein, L.; Beinfeld, M.; Minokadeh, A.; Zadina, J.; Hook, V. Unique biological function of cathepsin L in secretory vesicles for biosynthesis of neuropeptides. *Neuropeptides* **2010**, *44*(6), 457–466.

9. Benigni, A.; Cassis, P.; Remuzzi, G. Angiotensin II revisited: New roles in inflammation, immunology and aging. *EMBO Mol. Med.* **2010**, *2*(7), 247–257.

10. Kessler, J. H.; Khan, S.; Seifert, U.; Le Gall, S.; Chow, K. M.; Paschen, A.; Bres-Vloemans, S. A.; de Ru, A.; van Montfoort, N.; Franken, K. L. M. C.; Benckhuijsen, W. E.; Brooks, J. M.; van Hall, T.; Ray, K.; Mulder, A.; Doxiadis, I. I. N.; van Swieten, P. F.; Overkleeft, H. S.; Prat, A.; Tomkinson, B.; Neefjes, J.; Kloetzel, P. M.; Rodgers, D. W.; Hersh, L. B.; Drijfhout, J. W.; van Veelen, P. A.; Ossendorp, F.; Melief, C. J. M. Antigen processing by nardilysin and thimet oligopeptidase generates cytotoxic T cell epitopes. *Nat. Immunol.* **2011**, *12*(1), 45–53.

11. Ferro, E. S.; Rioli, V.; Castro, L. M.; Fricker, L. D. Intracellular peptides: From discovery to function. *EuPA Open Proteomics* **2014**, *3*, 143–151.

12. McDonald, B. A.; Stukenbrock, E. H. Rapid emergence of pathogens in agro-ecosystems: global threats to agricultural sustainability and food security. *Philos. Trans. R. Soc. B Biol. Sci.* **2016**, *371*(1709), 20160026.

13. Savary, S.; Willocquet, L.; Pethybridge, S. J.; Esker, P.; McRoberts, N.; Nelson, A. The

global burden of pathogens and pests on major food crops. *Nat. Ecol. Evol.* **2019**, *3*(3), 430–439.

14. Jones, J. D. G.; Dangl, J. L. The plant immune system. *Nature* **2006**, *444*(7117), 323–329.

15. Chisholm, S. T.; Coaker, G.; Day, B.; Staskawicz, B. J. Host-microbe interactions: Shaping the evolution of the plant immune response. *Cell* **2006**, *124*(4), 803–814.

16. Bent, A. F.; Mackey, D. Elicitors, Effectors, and R Genes: The New Paradigm and a Lifetime Supply of Questions. *Annu. Rev. Phytopathol.* **2007**, *45*(1), 399–436.

17. Boller, T.; He, S. Y. Innate immunity in plants: An arms race between pattern recognition receptors in plants and effectors in microbial pathogens. *Science* **2009**, *324*(5928), 742–744.

18. Qi, Y.; Tsuda, K.; Glazebrook, J.; Katagiri, F. Physical association of pattern-triggered immunity (PTI) and effector-triggered immunity (ETI) immune receptors in *Arabidopsis*. *Mol. Plant Pathol.* **2011**, *12*(7), 702–708.

19. Huffaker, A.; Pearce, G.; Ryan, C. A. An endogenous peptide signal in *Arabidopsis* activates components of the innate immune response. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*(26), 10098–10103.

20. Qia, Z.; Verma, R.; Gehring, C.; Yamaguchi, Y.; Zhao, Y.; Ryan, C. A.; Berkowitz, G. A. Ca 2+ signaling by plant *Arabidopsis thaliana* Pep peptides depends on AtPepR1, a receptor with guanylyl cyclase activity, and cGMP-activated Ca 2+ channels. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*(49), 21193–21198.

21. Misas-Villamil, J. C.; van der Hoorn, R. A. L.; Doehlemann, G. Papain-like cysteine proteases as hubs in plant immunity. *New Phytol.* **2016**, *212*(4), 902–907.

22. Hander, T.; Fernández-Fernández, Á. D.; Kumpf, R. P.; Willems, P.; Schatowitz, H.; Rombaut, D.; Staes, A.; Nolf, J.; Pottie, R.; Yao, P.; Gonçalves, A.; Pavie, B.; Boller, T.; Gevaert, K.; Van Breusegem, F.; Bartels, S.; Stael, S. Damage on plants activates Ca 2+ -dependent metacaspases for release of immunomodulatory peptides. *Science* **2019**, *363*(6433).

23. Shen, W.; Liu, J.; Li, J. F. Type-II Metacaspases Mediate the Processing of Plant Elicitor Peptides in *Arabidopsis*. *Mol. Plant* **2019**, *12*(11), 1524–1533.

24. Rei Liao, J. Y.; van Wijk, K. J. Discovery of AAA+ Protease Substrates through Trapping Approaches. *Trends Biochem. Sci.* **2019**, *44*(6), 528–545.

25. Majsec, K.; Bhuiyan, N. H.; Sun, Q.; Kumari, S.; Kumar, V.; Ware, D.; Van Wijk, K. J. The plastid and mitochondrial peptidase network in *Arabidopsis thaliana*: A foundation for testing genetic interactions and functions in organellar proteostasis. *Plant Cell* **2017**, *29*(11), 2687–2710.

26. Kmiec, B.; Teixeira, P. F.; Murcha, M. W.; Glaser, E. Divergent evolution of the M3A family of metallopeptidases in plants. *Physiol. Plant.* **2016**, *157*(3), 380–388.

27. Gomiz-Rüth, F. X. Catalytic domain architecture of metzincin metalloproteases. *J. Biol. Chem.* **2009**, *284*(23), 15353–15357.

28. Westlake, T. J.; Ricci, W. A.; Popescu, G. V.; Popescu, S. C. Dimerization and thiol sensitivity of the salicylic acid binding thimet oligopeptidases TOP1 and TOP2 define their functions in redox-sensitive cellular pathways. *Front. Plant Sci.* **2015**, *6*(327), 1–16.

29. Moreau, M.; Westlake, T.; Zampogna, G.; Popescu, G.; Tian, M.; Noutsos, C.; Popescu, S. The *Arabidopsis* oligopeptidases TOP1 and TOP2 are salicylic acid targets that modulate SA-mediated signaling and the immune response. *Plant J.* **2013**, *76*(4), 603–614.

30. García-Lorenzo, M.; Sjödin, A.; Jansson, S.; Funk, C. Protease gene families in Populus and Arabidopsis. *BMC Plant Biol.* **2006**, *6*(30).

31. Kmiec, B.; Teixeira, P. F.; Berntsson, R. P. A.; Murcha, M. W.; Branca, R. M. M.; Radomiljac, J. D.; Regberg, J.; Sèensson, L. M.; Bakali, A.; Langel, Ü.; Lehtiö, J.; Whelan, J.; Stenmark, P.; Glaser, E. Organellar oligopeptidase (OOP) provides a complementary pathway for targeting peptide degradation in mitochondria and chloroplasts. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*(40), E3761–E3769.

32. Polge, C.; Jaquinod, M.; Holzer, F.; Bourguignon, J.; Walling, L.; Brouquisse, R. Evidence for the existence in *Arabidopsis thaliana* of the proteasome proteolytic pathway: Activation in response to cadmium. *J. Biol. Chem.* **2009**, *284*(51), 35412–35424.

33. Käll, L.; Canterbury, J. D.; Weston, J.; Noble, W. S.; MacCoss, M. J. Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat. Methods* **2007**, *4*(11), 923–925.

34. Gianetto, Q. G. imp4p: Imputation for Proteomics https://cran.r-project.org/web/packages/imp4p/index.html (accessed Aug 6, 2019).

35. O'Shea, J. P.; Chou, M. F.; Quader, S. A.; Ryan, J. K.; Church, G. M.; Schwartz, D. PLogo: A probabilistic approach to visualizing sequence motifs. *Nat. Methods* **2013**, *10*(12), 1211–1212.

36. Simon, M. D.; Heider, P. L.; Adamo, A.; Vinogradov, A. A.; Mong, S. K.; Li, X.; Berger, T.; Policarpo, R. L.; Zhang, C.; Zou, Y.; Liao, X.; Spokoyny, A. M.; Jensen, K. F.; Pentelute, B. L. Rapid Flow-Based Peptide Synthesis. *ChemBioChem* **2014**, *15*(5), 713–720.

37. Perez-Riverol, Y.; Csordas, A.; Bai, J.; Bernal-Llinares, M.; Hewapathirana, S.; Kundu, D. J.; Inuganti, A.; Griss, J.; Mayer, G.; Eisenacher, M.; Erez, E. P.; Uszkoreit, J.; Pfeuffer, J.; Sachsenberg, T.; Yılmaz, S.; Tiwary, S.; Cox, J.; Audain, E.; Walzer, M.; Jarnuczak, A. F.; Ternent, T.; Brazma, A.; Vizcaíno, J. A. The PRIDE database and

related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* **2019**, *47*(D1), D442–D450.

38. Gemperline, E.; Keller, C.; Li, L. Mass Spectrometry in Plant-omics. *Anal. Chem.* **2016**, *88*(7), 3422–3434.

39. Maes, E.; Oeyen, E.; Boonen, K.; Schildermans, K.; Mertens, I.; Pauwels, P.; Valkenborg, D.; Baggerman, G. The challenges of peptidomics in complementing proteomics in a clinical context. *Mass Spectrom. Rev.* **2019**, *38*(3), 253–264.

40. Das, D.; Jaiswal, M.; Khan, F. N.; Ahamad, S.; Kumar, S. PlantPepDB: A manually curated plant peptide database. *Sci. Rep.* **2020**, *10*(1), 1–8.

41. Chae, K.; Kieslich, C. A.; Morikis, D.; Kim, S. C.; Lord, E. M. A gain-of-function mutation of *Arabidopsis* lipid transfer protein 5 disturbs pollen tube tip growth and fertilization. *Plant Cell* **2009**, *21*(12), 3902–3914.

42. Srivastava, R.; Liu, J.-X.; Guo, H.; Yin, Y.; Howell, S. H. Regulation and processing of a plant peptide hormone, AtRALF23, in *Arabidopsis*. *Plant J.* **2009**, *59*(6), 930–939.

43. Couturier, J.; Chibani, K.; Jacquot, J. P.; Rouhier, N. Cysteine-based redox regulation and signaling in plants. *Front. Plant Sci.* **2013**, *4*(105).

44. Simat, T. J.; Steinhart, H. Oxidation of Free Tryptophan and Tryptophan Residues in Peptides and Proteins. *J. Agric. Food Chem.* **1998**, *46*(2), 490–498.

45. Plan, M. R. R.; Göransson, U.; Clark, R. J.; Daly, N. L.; Colgrave, M. L.; Craik, D. J. The cyclotide fingerprint in *Oldenlandia affinis*: Elucidation of chemically modified, linear and novel macrocyclic peptides. *ChemBioChem* **2007**, *8*(9), 1001–1011.

46. Ghifari, A. S.; Huang, S.; Murcha, M. W. The peptidases involved in plant mitochondrial protein import. *J. Exp. Bot.* **2019**, *70*(21), 6005–6018.

47. Lee, D. W.; Lee, S.; Lee, J.; Woo, S.; Razzak, M. A.; Vitale, A.; Hwang, I. Molecular Mechanism of the Specificity of Protein Import into Chloroplasts and Mitochondria in Plant Cells. *Mol. Plant* **2019**, *12*(7), 951–966.

48. Kmiec, B.; Teixeira, P. F.; Glaser, E. Shredding the signal: Targeting peptide degradation in mitochondria and chloroplasts. *Trends Plant Sci.* **2014**, *19*(12), 771–778.

49. Varshavsky, A. The N-end rule pathway and regulation by proteolysis. *Protein Sci.* **2011**, *20*(8), 1298–1345.

50. Lobanov, M. Y.; Furletova, E. I.; Bogatyreva, N. S.; Roytberg, M. A.; Galzitskaya, O. V. Library of disordered patterns in 3D protein structures. *PLoS Comput. Biol.* **2010**, *6*(10), e1000958.

51. Lone, A. M.; Kim, Y. G.; Saghatelian, A. Peptidomics methods for the identification of

peptidase-substrate interactions. *Curr. Opin. Chem. Biol.* **2013**, *17*(1), 83–89.

52. Conlin, C. A.; Miller, C. G. Dipeptidyl carboxypeptidase and oligopeptidase A from *Escherichia coli* and *Salmonella typhimurium*. *Methods Enzymol.* **1995**, *248*, 567–579.

53. Dando, P. M.; Brown, M. A.; Barrett, A. J. Human thimet oligopeptidase. *Biochem. J.* **1993**, *294*(2), 451–457.

54. Kmiec, B.; Branca, R. M. M.; Berkowitz, O.; Li, L.; Wang, Y.; Murcha, M. W.; Whelan, J.; Lehtiö, J.; Glaser, E.; Teixeira, P. F. Accumulation of endogenous peptides triggers a pathogen stress response in *Arabidopsis thaliana*. *Plant J.* **2018**, *96*(4), 705–715.

55. Teixeira, P. F.; Kmiec, B.; Branca, R. M. M.; Murcha, M. W.; Byzia, A.; Ivanova, A.; Whelan, J.; Drag, M.; Lehtiö, J.; Glaser, E. A multi-step peptidolytic cascade for amino acid recovery in chloroplasts. *Nat. Chem. Biol.* **2017**, *13*(1), 15–17.

56. Richter, S.; Lamppa, G. K. Structural properties of the chloroplast stromal processing peptidase required for its function in transit peptide removal. *J. Biol. Chem.* **2003**, *278*(41), 39497–39502.

57. Park, M. H.; Zhong, R.; Lamppa, G. Chloroplast stromal processing peptidase activity is modulated by transit peptide determinants that include inhibitory roles for its N-terminal domain and initial Met. *Biochem. Biophys. Res. Commun.* **2018**, *503*(4), 3149–3154.

58. Wang, R.; Rajagopalan, K.; Sadre-Bazzaz, K.; Moreau, M.; Klessig, D. F.; Tong, L. Structure of the *Arabidopsis thaliana* TOP2 oligopeptidase. *Acta Crystallogr. Sect. F, Struct. Biol. Commun.* **2014**, *70*(Pt 5), 555–559.

59. Das, K.; Roychoudhury, A. Reactive oxygen species (ROS) and response of antioxidants as ROS-scavengers during environmental stress in plants. *Front. Environ. Sci.* **2014**, *2*(53).

60. Bassham, D. C. Plant autophagy-more than a starvation response. *Curr. Opin. Plant Biol.* **2007**, *10*(6), 587–593.

61. Díaz-Villanueva, J. F.; Díaz-Molina, R.; García-González, V. Protein folding and mechanisms of proteostasis. *Int. J. Mol. Sci.* **2015**, *16*(8), 17193–17230.

62. Kato, Y.; Sun, X.; Zhang, L.; Sakamoto, W. Cooperative D1 degradation in the photosystem II repair mediated by chloroplastic proteases in *Arabidopsis*. *Plant Physiol.* **2012**, *159*(4), 1428–1439.

63. Nishimura, K.; Kato, Y.; Sakamoto, W. Essentials of Proteolytic Machineries in Chloroplasts. *Mol. Plant* **2017**, *10*(1), 4–19.

64. Delauré, S. L.; Van Hemelrijck, W.; De Bolle, M. F. C.; Cammue, B. P. A.; De Coninck, B. M. A. Building up plant defenses by breaking down proteins. *Plant Sci.* **2008**, *174*(4), 375–385.

65. Tavormina, P.; De Coninck, B.; Nikonorova, N.; De Smet, I.; Cammuea, B. P. A. The plant peptidome: An expanding repertoire of structural features and biological functions. *Plant Cell* **2015**, *27*(8), 2095–2118.

66. Hou, S.; Liu, Z.; Shen, H.; Wu, D. Damage-associated molecular pattern-triggered immunity in plants. *Front. Plant Sci.* **2019**, *10*, 646.

67. Albert, M. Peptides as triggers of plant defence. *J. Exp. Bot.* **2013**, *64*(17), 5269–5279.

68. Krol, E.; Mentzel, T.; Chinchilla, D.; Boller, T.; Felix, G.; Kemmerling, B.; Postel, S.; Arents, M.; Jeworutzki, E.; Al-Rasheid, K. A. S.; Becker, D.; Hedrich, R. Perception of the *Arabidopsis* danger signal peptide 1 involves the pattern recognition receptor AtPEPR1 and its close homologue AtPEPR2. *J. Biol. Chem.* **2010**, *285*(18), 13471–13479.

69. Xia, Y.; Suzuki, H.; Borevitz, J.; Blount, J.; Guo, Z.; Patel, K.; Dixon, R. A.; Lamb, C. An extracellular aspartic protease functions in *Arabidopsis* disease resistance signaling. *EMBO J.* **2004**, *23*(4), 980–988.

70. Wrzaczek, M.; Brosché, M.; Kangasjärvi, J. ROS signaling loops - production, perception, regulation. *Curr. Opin. Plant Biol.* **2013**, *16*(5), 575–582.

71. Wang, L.; Kim, C.; Xu, X.; Piskurewicz, U.; Dogra, V.; Singh, S.; Mahler, H.; Apel, K. Singlet oxygen- and EXECUTER1-mediated signaling is initiated in grana margins and depends on the protease FtsH2. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*(26), E3792–E3800.

72. Maruta, T.; Noshi, M.; Tanouchi, A.; Tamoi, M.; Yabuta, Y.; Yoshimura, K.; Ishikawa, T.; Shigeoka, S. H_2O_2-triggered retrograde signaling from chloroplasts to nucleus plays specific role in response to stress. *J. Biol. Chem.* **2012**, *287*(15), 11717–11729.

73. Møller, I. M.; Sweetlove, L. J. ROS signalling - specificity is required. *Trends Plant Sci.* **2010**, *15*(7), 370–374.

74. Choudhury, S.; Panda, P.; Sahoo, L.; Panda, S. K. Reactive oxygen species signaling in plants under abiotic stress. *Plant Signal. Behav.* **2013**, *8*(4), e23681.

75. McConnell, E. W.; Berg, P.; Westlake, T. J.; Wilson, K. M.; Popescu, G. V.; Hicks, L. M.; Popescu, S. C. Proteome-wide analysis of cysteine reactivity during effector-triggered immunity. *Plant Physiol.* **2019**, *179*(4), 1248–1264.

76. Ishikawa, A.; Tanaka, H.; Nakai, M.; Asahi, T. Deletion of a Chaperonin 60 Beta Gene Leads to Cell Death in the *Arabidopsis* Lesion Initiation 1 Mutant. *Plant Cell Physiol.* **2003**, *44*(3), 255–261.

77. Hopkins, M. T.; Lampi, Y.; Wang, T. W.; Liu, Z.; Thompson, J. E. Eukaryotic translation initiation factor 5A is involved in pathogen-induced cell death and development of disease symptoms in *Arabidopsis*. *Plant Physiol.* **2008**, *148*(1), 479–489.

78. Gotor, C.; Laureano-Marín, A. M.; Moreno, I.; Aroca, Á.; García, I.; Romero, L. C. Signaling in the plant cytosol: Cysteine or sulfide? *Amino Acids* **2015**, *47*(10), 2155–2164.

79. Tahir, J.; Watanabe, M.; Jing, H.-C.; Hunter, D. A.; Tohge, T.; Nunes-Nesi, A.; Brotman, Y.; Fernie, A. R.; Hoefgen, R.; Dijkwel, P. P. Activation of R -mediated innate immunity and disease susceptibility is affected by mutations in a cytosolic O -acetylserine (thiol) lyase in *Arabidopsis*. *Plant J.* **2013**, *73*(1), 118–130.

80. Zhang, J.; Yuan, H.; Yang, Y.; Fish, T.; Lyi, S. M.; Thannhauser, T. W.; Zhang, L.; Li, L. Plastid ribosomal protein S5 is involved in photosynthesis, plant development, and cold stress tolerance in *Arabidopsis*. *J. Exp. Bot.* **2016**, *67*(9), 2731–2744.

81. Yuan, H. M.; Liu, W. C.; Lu, Y. T. CATALASE2 Coordinates SA-Mediated Repression of Both Auxin Accumulation and JA Biosynthesis in Plant Defenses. *Cell Host Microbe* **2017**, *21*(2), 143–155.

82. Hackenberg, T.; Juul, T.; Auzina, A.; Gwizdz, S.; Małolepszy, A.; Van Der Kelen, K.; Dam, S.; Bressendorff, S.; Lorentzen, A.; Roepstorff, P.; Nielsen, K. L.; Jørgensen, J. E.; Hofius, D.; Van Breusegem, F.; Petersen, M.; Andersen, S. U. Catalase and NO CATALASE ACTIVITY1 promote autophagy-dependent cell death in *Arabidopsis*. *Plant Cell* **2013**, *25*(11), 4616–4626.

83. Ditt, R. F.; Kerr, K. F.; De Figueiredo, P.; Delrow, J.; Comai, L.; Nester, E. W. The Arabidopsis thaliana transcriptome in response to *Agrobacterium tumefaciens*. *Mol. Plant-Microbe Interact.* **2006**, *19*(6), 665–681.

84. Grandellis, C.; Garavaglia, B. S.; Gottig, N.; Lonez, C.; Ruysschaert, J. M.; Ottado, J. DOTAP, a lipidic transfection reagent, triggers *Arabidopsis* plant defense responses. *Planta* **2019**, *249*(2), 469–480.

85. Huang, D.; Wu, W.; Abrams, S. R.; Cutler, A. J. The relationship of drought-related gene expression in *Arabidopsis thaliana* to hormonal and environmental factors. *J. Exp. Bot.* **2008**, *59*(11), 2991–3007.

86. Niehl, A.; Zhang, Z. J.; Kuiper, M.; Peck, S. C.; Heinlein, M. Label-free quantitative proteomic analysis of systemic responses to local wounding and virus infection in *Arabidopsis thaliana*. *J. Proteome Res.* **2013**, *12*(6), 2491–2503.

87. Huang, S.; Taylor, N. L.; Whelan, J.; Millar, A. H. Refining the Definition of Plant Mitochondrial Presequences through Analysis of Sorting Signals, N-Terminal Modifications, and Cleavage Motifs. *Plant Physiol.* **2009**, *150*(3), 1272–1285.

88. Zybailov, B.; Rutschow, H.; Friso, G.; Rudella, A.; Emanuelsson, O.; Sun, Q.; van Wijk, K. J. Sorting signals, N-terminal modifications and abundance of the chloroplast proteome. *PLoS One* **2008**, *3*(4), e1994–e1994.

89. Choi, K.-H.; Licht, S. Control of Peptide Product Sizes by the Energy-Dependent Protease ClpAP. *Biochemistry* **2005**, *44*(42), 13921–13931.

90. Krojer, T.; Pangerl, K.; Kurt, J.; Sawa, J.; Stingl, C.; Mechtler, K.; Huber, R.; Ehrmann, M.; Clausen, T. Interplay of PDZ and protease domain of DegP ensures efficient elimination of misfolded proteins. *Proc. Natl. Acad. Sci.* **2008**, *105*(22), 7702–7707.

91. Saric, T.; Graef, C. I.; Goldberg, A. L. Pathway for Degradation of Peptides Generated by Proteasomes: A KEY ROLE FOR THIMET OLIGOPEPTIDASE AND OTHER METALLOPEPTIDASES. *J. Biol. Chem.* **2004**, *279*(45), 46723–46732.

92. Berti, D. A.; Morano, C.; Russo, L. C.; Castro, L. M.; Cunha, F. M.; Zhang, X.; Sironi, J.; Klitzke, C. F.; Ferro, E. S.; Fricker, L. D. Analysis of intracellular substrates and products of thimet oligopeptidase in human embryonic kidney 293 cells. *J. Biol. Chem.* **2009**, *284*(21), 14105–14116.

CHAPTER 3: The Contributions of Thimet Oligopeptidase-Directed Proteostasis Towards Effector-Triggered Immunity in *Arabidopsis thaliana*

*****Submitted for publication. Authors: Iannetta, A. A.; Berg, P.; Nejat, N.; Smythers, A. L.; Purvis, A. L.; Brown, Z. A.; Wommack, A. J.; Popescu, S. C.; Hicks, L. M.; and Popescu, G.V.

3.1 Introduction

Plants are sessile organisms that have developed an innate immune response with complex, chemical-based signaling pathways that sense and respond to unfavorable environments[1]. It is critical to understand how plants perceive stress signals to elicit responses for survival. One aspect of this evolved plant innate immune response is effector-triggered immunity (ETI), a robust resistance response to virulence effectors deployed by pathogens to suppress and interfere with pathogen-associated molecular pattern-triggered immunity[2]. ETI is activated when nucleotide-binding leucine-rich repeat immune receptors recognize effectors, such as avrRpt2 from *Pseudomonas syringae*[2–4]. The detection of these effectors elicits the hypersensitive response (HR), a form of programmed cell death (PCD), that restricts pathogen growth[5,6]. HR increases both salicylic acid (SA) and reactive oxygen species (ROS) and maintains a positive feedback loop between SA and ROS until a threshold is reached to initiate ETI and SA signaling[7,8]. ROS accumulation elicited by ETI is biphasic with a low amplitude, transient first phase, followed by a sustained phase of much higher magnitude[9].

ROS accumulation following the HR damages proteins via irreversible oxidation and removal of these oxidized proteins is crucial to maintain proteostasis[10,11]. Therefore, proteolysis is necessary for proper plant defense against pathogenic stress[12]. Cytosolic proteolytic cascades

complete the processing of proteasome-released peptides, while organelle-localized proteolytic components (*e.g.,* chloroplastic and mitochondrial proteolytic cascades) have a wide range of functions including the removal of damaged organellar proteins[13,14]. Although connectivity between organellar and cytosolic proteolytic networks is not well characterized in plants, it is an essential component of the metazoan response to oxidative stress[15,16]. Within these pathways, metazoan oligopeptidases with specificity limited to few substrates are critical factors in generating bioactive peptides for stress response signaling through controlled proteolysis[17,18]. This is less explored in plants, but plant peptidases have been shown to have other functions besides their role in protein maintenance, including the release of defense response peptides[19]. Controlled proteolysis in plants requires more investigation to reveal the identity of the protease activities involved and cognate bioactive peptide products.

Plant thimet oligopeptidases (TOPs) are zinc-dependent peptide hydrolases with a conserved HEXXH active site motif[20,21]. These metallopeptidases are critical components in plant response to oxidative stress through SA-mediated signaling pathways and are required for a fully functioning immune response to pathogens that activate ETI[22,23]. The Arabidopsis genome contains two genes encoding TOPs: *TOP1* and *TOP2*[24]. TOP1 (AT5G65620, also known as organellar oligopeptidase, OOP) contains an N-terminal signal peptide that mediates its localization to the chloroplast and mitochondria[23,25]. TOP1 cleaves presequences containing 8-23 amino acids *in vitro* and is hypothesized to act downstream of organellar proteases for intra-organelle peptide degradation and organelle import processing[25]. TOP2 (AT5G10540, also known as cytosolic oligopeptidase, CyOP) is located in the cytosol, where it is predicted to act downstream of the 20S proteasome, degrading proteasome-generated peptides during oxidative stress[23,25,26]. Prior evidence suggests that TOP1 and TOP2 have functional overlap in ETI and

PCD[22,23,25,26]. Both oligopeptidases are required for plant defense to avirulent strains of *P. syringae* through the activation of the resistance proteins RPS2 or RPS4, and both are necessary to regulate the PCD[23]. In a current model, TOPs sustain interconnected organelle and cytosol proteolytic pathways that regulate the ETI oxidative burst and plant resistance to pathogens through SA, ROS, and antioxidants[22].

Our prior work delineated TOP peptides substrates via quantitative *in vivo* peptidomics comparing Arabidopsis Col-0 and the *top1top2* null mutant[27]. Herein, a quantitative mass spectrometry-based peptidomics approach was implemented to characterize the peptidomes in locally infected tissue of *Arabidopsis thaliana* wild type (Col-0) and *top1top2* null mutant plants during the early stages of ETI.

3.2 Materials and Methods

3.2.1 Plant Growth and Infection Assays

Seeds of *Arabidopsis thaliana* ecotype Columbia (Col-0) and *top1top2* mutant line were sterilized by standardized methods as described in[28], and grown on MS media for 10 days, then transferred to individual jiffy pellets under controlled conditions with a 12h light (12:00pm to 12:00am; 100 µmol m^{-2} s^{-1} photon flux density) / 12 h dark period and a relative humidity of 60% to 65%. Day and night temperatures were set to 23°C and 21°C, respectively. Experiments were performed with 4 to 5-week-old, uniform appearance naïve plants. *Pseudomonas syringae* pv. *tomato* pathovar DC3000 (*Pst*) were cultivated at 28°C in King's B medium (Sigma Aldrich) containing the appropriate antibiotics, Rifampicin and Kanamycin[29]. Overnight log phase cultures were diluted to final optical densities of 600 nm (OD_{600}) for leaf inoculations of Col-0 and *top1top2* plants. To activate ETI, two to three mature leaves at similar developmental stages were infiltrated with PstAvrRpt2 suspensions in 10 mM $MgCl_2$ buffer at 5×10^5 CFU ml^{-1};

control plants were infiltrated with 10 mM $MgCl_2$. Infiltrated leaves were harvested at the required timepoints for peptidome analysis.

3.2.2 Peptide Extraction

Three biological replicates were used for each genotype (*i.e.,* Col-0 and *top1top2* mutant) and infection timepoint. The preparation of peptidome samples from local tissue followed the method described in Iannetta *et al.*[27]. Briefly, rosette leaf tissue was ground under liquid N_2 and peptides were extracted from plant material in two rounds using 10% trichloroacetic acid (TCA) in acetone. The isolation of peptides from small molecules in this extract was performed using strong cation exchange solid-phase extraction (SPE) and peptides were desalted using reversed phase SPE. Peptide concentrations were estimated using the Pierce Quantitative Colorimetric Peptide Assay (Thermo Fisher Scientific) according to the manufacturer's protocols. Based on these results, peptide concentrations in each experiment were normalized across replicates before LC-MS/MS analysis.

3.2.3 LC-MS/MS Analysis

Samples were analyzed using an Acquity UPLC M-Class System (Waters) coupled to a Q Exactive HF-X mass spectrometer (Thermo Fisher Scientific). Mobile phase A consisted of water with 0.1% formic acid (Thermo Fisher Scientific) and mobile phase B was acetonitrile with 0.1% formic acid. Injections (1 µL) were made to a Symmetry C18 trap column (100 Å, 5µm, 180µm x 20 mm; Waters) with a flow rate of 5 µL/min for 3 min using 99% A and 1% B. Peptides were then separated on a HSS T3 C18 column (100 Å, 1.8µm, 75µm x 250 mm; Waters) using a linear gradient of increasing mobile phase B at a flow rate of 300 nL/min. Mobile phase B increased from 5% to 40% in 90 min before ramping to 85% in 5 min, where it was held for 10 min before returning to 5% in 2 min and re-equilibrating for 13 min. The mass

spectrometer was operated in positive polarity and the Nanospray Flex source had spray voltage floating at 2.1 kV, capillary temperature at 320 °C, and funnel RF level at 40. MS survey scans were collected with a scan range of 350 – 2000 *m/z* at a resolving power of 120,000 and an AGC target of 3×10^6 with a maximum injection time of 50 ms. A top 20 data-dependent acquisition was used where HCD fragmentation of precursor ions having +2 to +7 charge state was performed using a normalized collision energy setting of 28. MS/MS scans were performed at a resolving power of 30,000 and an AGC target of 1×10^5 with a maximum injection time of 100 ms. Dynamic exclusion for precursor *m/z* was set to a 10 s window.

3.2.4 Database Searching and Label-Free Quantification

Acquired spectral files (*.raw) were imported into Progenesis QI for proteomics (Waters, version 2.0). Peak picking sensitivity was set to the maximum of five and a reference spectrum was automatically assigned. Total ion chromatograms (TICs) were then aligned to minimize run-to-run differences in peak retention time. Each sample received a unique factor to normalize all peak abundance values resulting from systematic experimental variation. Alignment was validated (>80% score) and a combined peak list (*.mgf) was exported out of Progenesis for peptide sequence determination by Mascot (Matrix Science, version 2.5.1; Boston, MA). Database searching was performed against the *Arabidopsis thaliana* UniProt database (https://www.uniprot.org/proteomes/UP000006548, 39,345 canonical entries, accessed 03/2021) with sequences for common laboratory contaminants (https://www.thegpm.org/cRAP/, 116 entries, accessed 03/2021) appended. Target-decoy searches of MS/MS data used “None” as the enzyme specificity, peptide/fragment mass tolerances of 15 ppm/0.02 Da, and variable modifications of N-terminus acetylation, C-terminus amidation, and methionine oxidation. Significant peptide identifications above the identity or homology threshold were adjusted to less

than 1% peptide FDR using the embedded Percolator algorithm[30]. Mascot results (*.xml) were imported to Progenesis for peak matching. Identifications with a Mascot score less than 13 were removed from consideration in Progenesis before exporting both “Peptide Measurements” and “Protein Measurements” from the “Review Proteins” stage.

3.2.5 Data Analysis and Statistics

Data were parsed using custom scripts written in R for pre-processing and statistical analysis (https://github.com/hickslab/QuantifyR). The “Peptide Measurements” data contain peak features with distinct precursor mass and retention time coordinates matched with a peptide sequence identification from the database search results. Some features were duplicated and matched with peptides having identical sequence, modifications, and score, but alternate protein accessions. These groups were reduced to satisfy the principle of parsimony and represented by the protein accession with the highest number of unique peptides found in the “Protein Measurements” data for this experiment, else the protein with the largest confidence score assigned by Progenesis. Some features were also duplicated with differing peptide identifications and were reduced to just the peptide with the highest Mascot ion score. An identifier was created by joining the protein accession of each peptide to the identified peptide sequence. Each dataset was reduced to unique identifiers by summing the abundance of all contributing peak features (*i.e.*, different peptide charge states and combinations of variable modifications). Identifiers were represented by the peptide with the highest Mascot score in each group. Identifiers were removed if there was not at least one condition with > 50% nonzero values across the Progenesis-normalized abundance columns. Values were $\log_2$-transformed and a conditional imputation strategy was applied using the imp4p package[31], where conditions with at least one nonzero value had missing values imputed using the *impute.rand* function with default parameters. For

cases where a condition had only missing values, the *impute.pa* function was used to impute small numbers centered on the lower 2.5% of values in each replicate.

For comparisons within genotypes, statistical significance was determined using a one-way analysis of variance (ANOVA). Only observations p-value < 0.05 were considered significantly different. Unsupervised hierarchical clustering was performed on significantly different identifiers to group together similarly changing abundance trends across conditions (*i.e.*, across treatment times). For visualization purposes, abundance values across treatments for each identifier were converted to standardized z-scores. For pairwise comparisons, statistical significance was determined using a two-tailed, equal variance *t*-test. Fold change was calculated by the difference of the mean abundance values between conditions being compared. Only observations with p-value < 0.05 and $|\log_2$-transformed fold change$| \geq 1$ after a two-sided, equal variance *t*-test were considered significantly different. Gene ontology (GO) annotations were assigned from UniProt. GO term enrichment was performed using the PANTHER Overrepresentation Test (Fisher's Exact with false discovery rate correction).

3.2.6 Motif Analysis

In preparation for sequence logo visualization, data were parsed using custom scripts written in R. Peptides were filtered for those significantly increasing in Col-0 (p-value < 0.05, $\log_2$-transformed fold change ≤ -1). Based on the originating protein sequence, the peptide termini were extended four amino acids; for peptides too close to either protein termini, peptides were extended until the protein terminus was reached. Each elongated peptide was truncated into two smaller peptides containing the extended amino acids and either the first or last four amino acids of the original peptide sequence. These truncated peptides were filtered for those with eight residues to satisfy the input condition requiring peptides of the same length. For motif analysis,

sequence logo visualizations were performed using pLOGO, Positions with significant residue presence are depicted as amino acid letters sized above the red line[32].

3.2.7 *In Vitro* Enzyme Assay

Heterologously expressed and purified TOP enzymes and synthetic peptides were produced as described in Iannetta *et al.*[27]. Synthesized peptides were solubilized in 500 µL of 100 mM NaCl in 50 mM Tris, pH 7.5. To initiate the enzyme assay, either TOP1, $^{\Delta SP}$TOP1, or TOP2 was added at a peptide:TOP ratio of 10:1. The reaction mixture was incubated at 23 °C for 30 min. Reaction mixtures were desalted using reversed phase SPE.

3.2.8 ESI-MS Analysis

Samples from the *in vitro* enzyme assays were resuspended in 100 µL of 50% methanol/0.1% formic acid. Peptides were directly infused via ESI on a Q Exactive HF-X Hybrid mass spectrometer (Thermo Fisher Scientific) for intact mass analysis. Samples were injected at a flow rate of 10 µL/min and full MS scans were analyzed in the Orbitrap. The mass spectrometer was operated at a resolving power of 120,000, positive polarity, spray voltage of 3 kV, with 150–2000 *m/z* range, and collecting 100 scans/sample for averaging.

3.2.9 Quantitative ATP Luciferase Assay

Four biological replicates of frozen rosette leaf tissue (~0.05 g) were pulverized via three, 5 min rounds using a TissueLyser II (Qiagen, Germantown, MD) cell disrupter at 30 Hz before homogenization in 250 µL of water. The homogenate was vortexed and incubated at 100 °C for 30 min before centrifugation at 10,000 *g* for 15 min at 4 °C. The supernatant was collected, protein concentrations were estimated using the CB-X assay (G-Biosciences, St. Louis, MO) according to the manufacturer's protocol, and these concentrations were used to normalize replicates following luminescence analysis. The ATP determination kit (Invitrogen, Waltham,

MA) was used to quantify cellular ATP concentrations according to the manufacturer's protocol. Briefly, 20 µL of sample and 180 µL of 1 mM DTT, 0.5 mM luciferin, and 1.25 µg/mL firefly luciferin in 1x reaction buffer were added to a 96-well plate and incubated in the dark for 5 min at RT. Luminescence was measured and compared to an ATP standard calibration curve to convert luminescence to ATP concentration. Each biological replicate was analyzed in technical triplicates.

3.2.10 Quantitative $NADP^+$/NADPH Enzyme Cycling Assay

The NADP/NADPH Quantitation Kit (Sigma-Aldrich, Burlington, MA) was used to quantify cellular $NADP^+$/NADPH concentrations according to the manufacturer's protocol. Briefly, five biological replicates of frozen rosette leaf tissue (~0.05 g) were pulverized via three, 5 min rounds using a TissueLyser II (Qiagen) cell disrupter at 30 Hz before homogenization in 500 µL of the provided extraction buffer. The homogenate was vortexed and incubated at -20 °C for 10 min before centrifugation at 10,000 *g* for 15 min at 4 °C. The supernatant was collected and filtered through a 10 kDa molecular weight cutoff filter by centrifugation at 3,200 *g* for 30 min at 4 °C. Prior to filtering, an aliquot was reserved for protein quantification via the CBX assay (G-Biosciences); this protein concentration was used to normalize $NADP^+$/NADPH on the basis of protein quantification following analysis. For $NADP^+$ quantification, 50 µL of filtrate and 100 µL of the provided reaction mixture were added to a 96-well plate and incubated for 5 min at RT. After incubation, 10 µL of developer solution was added and mixtures were incubated in the dark for 30 min at RT. Absorbances were measured (λ = 450 nm) and compared to an $NADP^+$ standard calibration curve to convert absorbance to $NADP^+$ concentration. For total $NADP^+$/NADPH quantification, 200 µL of filtrate was incubated for 30 min at 60 °C before centrifugation at 10,000 *g* for 5 min at 4 °C. The following steps mirrored the procedure for

$NADP^+$ quantification. The difference between the total $NADP^+$/NADPH and $NADP^+$ concentrations was calculated to determine the NADPH concentration for each replicate. Each biological replicate was analyzed in technical duplicates.

3.2.11 Data Availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository[33] and can be accessed with the dataset identifier PXD019812 and 10.6019/PXD019812.

3.3 Results and Discussion

3.3.1 ETI-Triggered Peptidomics

To measure peptidome changes during the initial stages of ETI and elucidate TOP-mediated proteolytic pathways during plant defense, Col-0 and *top1top2* rosette leaves were analyzed 0 minutes post inoculation (mpi), 30 mpi, and 180 mpi with *Pseudomonas syringae* pv. tomato DC3000 (*Pst*) carrying the avirulence gene avrRpt2. Quantitative peptidomics was processed in a single experiment and included local tissue from both genotypes following pathogen inoculation at all timepoints (Figure 3.1, Table S1, Lab_Members\Tony\ThesisSupp). Two approaches were taken to assess peptide abundance differences across these conditions. First, peptide abundances were compared across infection timepoints within each plant line to determine peptidome differences during plant defense response. Second, peptide abundances were compared across plant lines at each timepoint to reveal stress-mediated TOP proteolysis.

The analysis of the Col-0 samples produced 2810 quantifiable peptides from 698 proteins (Table S2, Lab_Members\Tony\ThesisSupp), while *top1top2* revealed 2793 quantifiable peptides from 693 proteins (Table S3, Lab_Members\Tony\ThesisSupp). A one-way ANOVA within each

genotype revealed that 1260 peptides in Col-0 and 727 peptides in *top1top2* were significantly changing (p-value < 0.05) across the infection timepoints. Hierarchical clustering on the significantly changing peptides in Col-0 (Figure 3.2A, Table S2, column F, Lab_Members\Tony\ThesisSupp) and *top1top2* (Figure 3.2B, Table S3, column F, Lab_Members\Tony\ThesisSupp) resulted in six clusters for each plant line, displaying temporal trends in the peptidomes following pathogen inoculation.

In Col-0, three of the trends (Clusters A-C) contain peptides that are less abundant at 30 mpi than 0 mpi. These could represent peptides that are prompted to be degraded by the ETI oxidative burst. Two of the trends (Clusters D/E) contain peptides that are more abundant at 30 mpi than 0 mpi and may indicate peptides that are derived from proteins whose transcription/translation may be triggered by the ETI oxidative burst or peptides that are stabilized by direct oxidation. The last trend (Cluster F) contains peptides that have a similar abundance at 0 and 30 mpi, but then are less abundant at 180 mpi suggesting a delay in degradation. ROS accumulation elicited by ETI is biphasic with a low amplitude, transient first phase, followed by a sustained phase of much higher magnitude[9]. The degradation of peptides in this last trend could require higher ROS levels produced in the second oxidative burst phase.

For *top1top2*, three of the trends (Clusters A-C) contain peptides that are more abundant at 30 mpi than 0 mpi. These could represent peptides that are direct or indirect targets of TOP1 and/or TOP2. The other three trends (Clusters D-F) contain peptides that decrease in abundance post-inoculation, either at a single or both timepoints. These may signify peptides that are degraded through TOP-independent pathways or a TOP-regulated pathway that is suppressed in Col-0.

3.3.2 Differential Peptidomics Reveals Potential TOP Substrates During ETI

Peptidase activity nullified via genetic or chemical approaches decreases peptidase products and increases the accumulation of substrates[34,35]. Thus, the absence of TOP1 and TOP2 *in vivo* results in an increased abundance of peptide substrates in the *top1top2* knock-out mutant vs. Col-0. Likewise, a representative increased abundance of products derived from these substrates would be expected in Col-0 vs. the mutant. The double mutant was used in comparison with Col-0 instead of *top* single mutants due to their documented shared roles in ETI and PCD[22,23,25,26]. While TOP1 and TOP2 have different subcellular localizations, their functional overlap and high sequence similarity suggest a potential for redundant substrates[20,23]. The use of the double mutant also ensures the detection of TOP substrate accumulation compared to Col-0.

Col-0 and *top1top2* peptidomes of local tissue were compared at each timepoint to distinguish TOP-mediated proteolysis during the initial stages of ETI. Peptides that are significantly accumulated in the null *top1top2* mutant may represent direct substrates or pleotropic effects due to the absence of TOP activity. Across these samples, 2830 peptides from 698 proteins were quantified, revealing peptides that accumulated in either plant line at the three infection timepoints (Figure 3.3A-C, Table S4-9, Lab_Members\Tony\ThesisSupp). There is little overlap within each genotype when comparing peptides that were significantly more abundant in Col-0 (potential TOP cleavage peptide products) and *top1top2* (potential TOP substrates) across the timepoints (Figure 3.3D,E). This suggests that TOP peptide substrates could change during ETI as many unique peptide substrates and peptide cleavage products accumulate post inoculation.

Peptides more abundant in Col-0 across the infection timepoints were analyzed to examine TOP cleavage specificity as at least one termini of these products may result from TOP cleavage.

Motif analysis was performed on these prospective TOP recognition sites after extending peptide termini to include the surrounding amino acids[32]. This resulted in a significant enrichment of Val at P4, Ala at P3, Val at P2, Lys at P1, Ala at P1', Val at P2', Ala at P3', and Asp at P4' (Figure 3.4). This aligns with our previous work that revealed TOP cleavage preference for nonpolar residues in the positions surrounding the cleavage site[27], except for preference for the basic Lys residue at the P1 position. This is congruent with other studies that have shown that Arabidopsis and *Homo sapiens* TOPs, as well as other metallopeptidases in the same enzyme family do not have strict cleavage specificities[25,36,37].

Since peptides accumulated in either genotype represent potential TOP substrates or TOP cleaved peptide products, the combination of these peptides at each timepoint can reveal TOP-mediated proteolysis pre- and post-inoculation. Gene Ontology analysis of these significantly changing peptides between infection timepoints was performed to uncover effects of TOP1 and TOP2 on different processes during ETI (Figure S1, Table S10, Lab_Members\Tony\ThesisSupp). Even though the overlap of significantly changing peptides was low, similar GO biological processes and molecular functions were enriched at all timepoints. Many photosynthesis related GO biological processes including carbon fixation, reductive pentose-phosphate cycle, and photosynthetic electron transport in photosystem II were enriched at each timepoint. Likewise, biological processes associated with ATP synthesis and gluconeogenesis were also enriched at all timepoints. Glycine catabolic and metabolic processes were only enriched at 30 mpi, while purine nucleoside/ribonucleoside diphosphate metabolic processes were only enriched at 180 mpi. For GO molecular functions, activities of oxidoreductases that catalyze the reactions between $NADP^+$/NADPH and NAD^+/NADH, as well as phosphoglycerate kinase were enriched at all timepoints. Dihydrolipoyl dehydrogenase and

spermidine synthase activities were only enriched at 30 mpi, while ribulose-bisphosphate carboxylase and fructose-bisphosphate aldolase activities were only enriched at 180 mpi. From this it shows that TOPs maintain proteostasis within certain processes such as photosynthesis and ATP synthesis throughout ETI, in addition to other metabolic processes as ETI progresses.

3.3.3 Validation of TOP Substrates

From pairwise comparisons of local tissue between the genotypes across infection timepoints, peptides significantly more abundant in the *top1top2* mutant *in vivo* likely represent either direct substrates or pleiotropic effects displaying increased accumulation to compensate for the loss of TOP activity[38]. Validating these peptides *in vitro* can delineate between these hypotheses as direct substrates can be confirmed *in vitro* following incubation of candidate peptides with heterologously expressed TOPs. Peptides were selected for screening using the following criteria: 1) consideration of previously determined TOP substrate length (8-23 amino acids), 2) detection of increased *top1top2* mutant peptides over multiple timepoints, 3) detection of increased *top1top2* mutant peptides from the same protein with overlapping recognition sequences and 4) the largest *in vivo* fold changes. These *in vitro* enzyme assays were conducted with 13 synthetic peptides (Table S11, Lab_Members\Tony\ThesisSupp) that met the above criteria. The selected peptides are derived from proteins involved with photosynthesis, ATP synthesis/binding, carbon fixation, and fatty acid synthesis/beta-oxidation. Substrate candidates were incubated with recombinant $^{\Delta SP}$TOP1 and TOP2 and analyzed via LC-MS. Of the 13 candidate peptides, 10 were identified as cleaved by $^{\Delta SP}$TOP1 and/or TOP2, representing direct TOP substrates (Table 3.1, Figure S2-11, Lab_Members\Tony\ThesisSupp). There were a greater number of identified substrates and cleavage sites for $^{\Delta SP}$TOP1 (10 confirmed substrates, 18 total cleavages) compared to TOP2 (8 confirmed substrates, 15 total cleavage sites).

3.3.4 TOPs are Required for Proper Metabolic State During Immune Response

Our results from *in vivo* peptidome differential analysis and *in vitro* enzyme assays display that ETI influences TOP-mediated proteolysis of peptides derived from proteins involved in ATP synthesis. Many peptides from chloroplastic and mitochondrial ATP synthase subunits accumulated in the *top1top2* mutant vs. Col-0 and were decreased in Col-0 post-inoculation. This suggests that TOPs are necessary for ATP synthase proteostasis during ETI, and lack of TOPs in the *top1top2* mutant could lead to dysfunction of ATP synthesis. To determine if this proteolysis plays a role in ATP synthesis after pathogen challenge, ATP concentrations were measured in Col-0 and *top1top2* following inoculation with *Pst* avrRpt2 (Figure 3.5A). ATP concentrations between Col-0 and *top1top2* were similar at 0 mpi. In Col-0, there was a burst in ATP at 30 mpi, before ATP concentrations returned to original levels at 180 mpi. This differs in *top1top2* where ATP concentrations remained low and steady during all timepoints. This reveals defective ATP synthesis in the *top1top2* mutant and that TOPs are necessary for ATP accumulation during plant defense. Cellular damage causes the release of ATP into the extracellular matrix where it is recognized as a damage-associated molecular pattern by P2 receptor kinase 1 on the plasma membrane[39,40]. This causes increased formation of the secondary messengers cytosolic Ca^{2+}, nitric oxide, and ROS, which activates jasmonate-mediated gene expression for plant defense responses[41]. Previously, *top1top2* was shown to allow statistically significantly higher growth of *Pst* avrRpt2 compared to Col-0[23]. The lack of ATP available for defense signaling in the mutant could help explain this increased susceptibility to this pathogen.

GO analysis of potential TOP substrates revealed an enrichment for oxidoreductases that catalyze the reaction between $NADP^+$/NADPH at all timepoints. This indicates that TOPs mediate proteostasis in this process, and the regulation of $NADP^+$ reduction could be affected by

the absence of TOPs. To confirm this, $NADP^+$ concentrations were measured in Col-0 and *top1top2* following inoculation with *Pst* avrRpt2 (Figure 3.5B). At 0 mpi, Col-0 had a higher concentration of $NADP^+$ compared to *top1top2*. In Col-0, there was a decrease in $NADP^+$ concentration at 30 mpi to levels similar to *top1top2* at 0 mpi and this was maintained 180 mpi. In *top1top2*, the opposite trend was observed as $NADP^+$ concentrations increased at 30 mpi and remained at this level at 180 mpi. NADPH is an important reductant during plant response to oxidative stress[42], so this dysregulation of $NADP^+$/NADPH in the absence of TOPs could contribute to the increased susceptibility of *top1top2* to pathogens that trigger ETI.

3.4 Tables

Table 3.1 Synthetic AtTOP peptide substrates found to be cleaved by TOPs after analysis with LC-MS. The arrows represent identified sites of cleavage (ND: none detected). Cleaved peptide products that are bolded and underlined were uniquely detected in the enzyme-treated samples compared to the analysis of the bare synthetic peptide (Figure S2-11).

Peptide Sequence	$^{\Delta SP}$TOP1 Cleavage Sites	TOP2 Cleavage Sites
IKTDKPFGIN	**IKTDKPF**↓GIN IKTD↓**KPFGIN**	**IKTDKPF**↓GIN IKTD↓**KPFGIN**
MDSDFGIPR	**MDSDFG**↓IPR	**MDSDFG**↓IPR
TGDQRLLDAS	**TGDQRLL**↓DAS	**TGDQRLL**↓DAS
VVISAPSKDAPM	**VVIS**↓APSKDAPM	**VVISAPSKD**↓APM **VVIS**↓APSKDAPM
SVVKLEAPQLAQ	**SVVKLEAPQ**↓LAQ **SVVKLE**↓**APQLAQ** **SVVKL**↓EAPQLAQ	**SVVKLEAPQ**↓LAQ **SVVKLE**↓**APQLAQ** **SVVKL**↓EAPQLAQ
GSSFLDPK	**GSSFL**↓DPK	**GSSFL**↓DPK
VLNTGAPITVPVGRATLG	VLNTGAPITVPVG↓**RATLG** VLNTGAPITVP↓**VGRATLG** **VLNTGAPIT**↓VPVGRATLG	**VLNTGAPITVPVGRA**↓TLG VLNTGAPITVPVG↓**RATLG** VLNTGAPITVP↓**VGRATLG** **VLNTGAPIT**↓VPVGRATLG
DPFGLGKPA	**DPFGLG**↓KPA	**DPFGLG**↓KPA
AKDELAGSIQKGV	**AKDELAGSIQ**↓KGV	ND
TGGGASLELLEGKPLPG	**TGGGASLELLE**↓**GKPLPG** TGGGASLELL↓**EGKPLPG** **TGGGASLE**↓LLEGKPLPG **TGGGASL**↓ELLEGKPLPG	ND

3.5 Figures

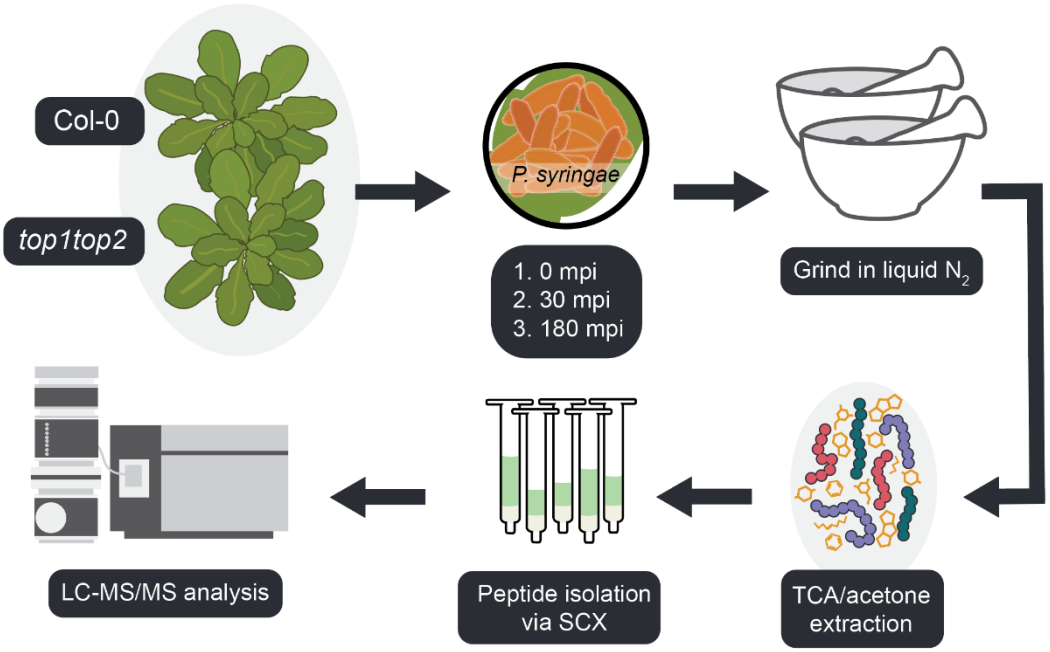


Figure 3.1 The peptidomes of Col-0 and *top1top2* rosette leaves were analyzed following pathogen inoculation with *Pst* avrRpt2 to measure peptidome changes during the initial stages of ETI and elucidate TOP-mediated proteolytic pathways during plant defense. Rosette leaves were ground under liquid nitrogen before extracting peptides with 10% TCA in acetone. Peptides were isolated from small molecules with SCX SPE before peptide quantitation. Peptide concentrations across replicates were quantified and normalized before liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis.

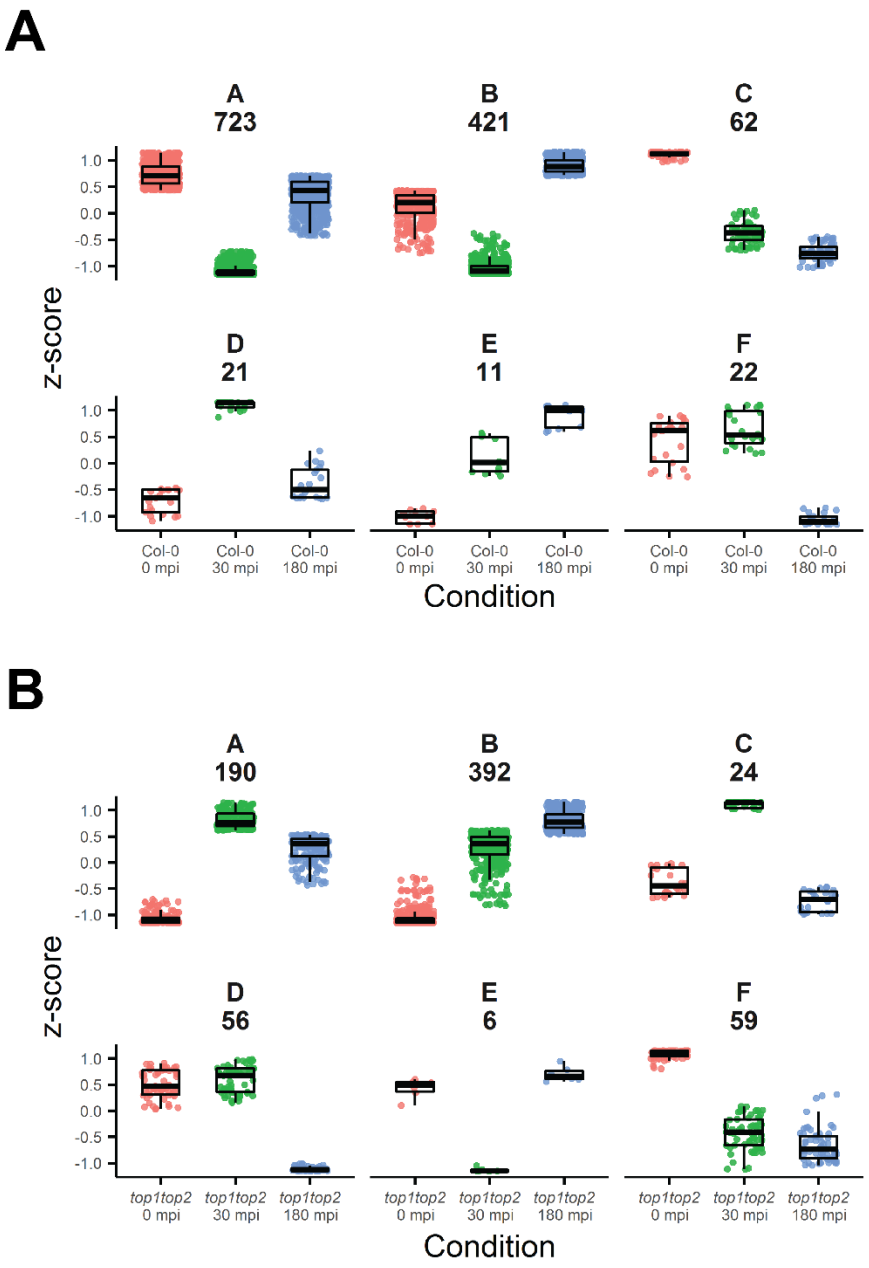


Figure 3.2 Unsupervised hierarchical clustering of the peptides significantly changing across the infection timepoints in each plant genotype after a one-way ANOVA (*p*-value < 0.05), displaying TOP-mediated peptidome changes in the initial stages of ETI. (A) The 1260 peptides in Col-0 that were significantly changing across the timepoints were grouped into six clusters. (B) The 727 peptides in *top1top2* that were significantly changing across the timepoints were grouped into six clusters.

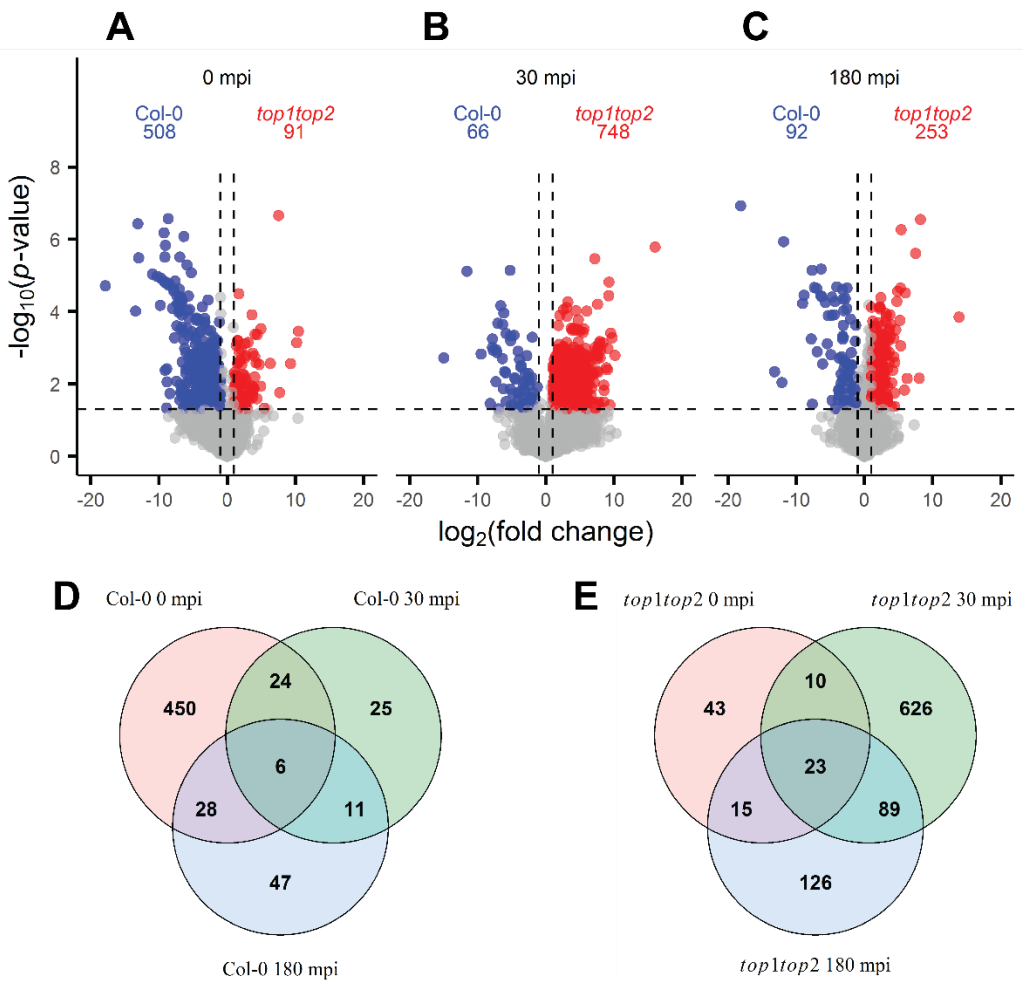


Figure 3.3 Differential analysis of the Arabidopsis peptidome during the initial stages of ETI. Colored circles represent significantly changing peptides (p-value < 0.05, $|FC| \geq 2$) after a two-sided, equal variance t-test. Red circles indicate increasing peptides in the *top1top2* mutant line, which are potential TOP peptide substrates. Blue circles indicate significantly increasing peptides in the wild type, which are potential TOPs cleaved peptide products. (A) Comparison of plant genotypes 0 mpi with *Pst* avrRpt2. At this timepoint, 599 peptides had a significant fold change, with 91 peptides more abundant in *top1top2* and 508 more abundant in Col-0. (B) Comparison of plant genotypes 30 mpi with *Pst* avrRpt2. At this timepoint, 814 peptides had a significant fold change, with 748 peptides more abundant in *top1top2* and 66 more abundant in Col-0. (C) Comparison of plant genotypes 180 mpi with *Pst* avrRpt2. At this timepoint, 345 peptides had a significant fold change, with 253 peptides more abundant in *top1top2* and 92 more abundant in Col-0. (D) Overlap of peptides that were significantly more abundant in Col-0 across the timepoints. (E) Overlap of peptides that were significantly more abundant in *top1top2* across the timepoints.

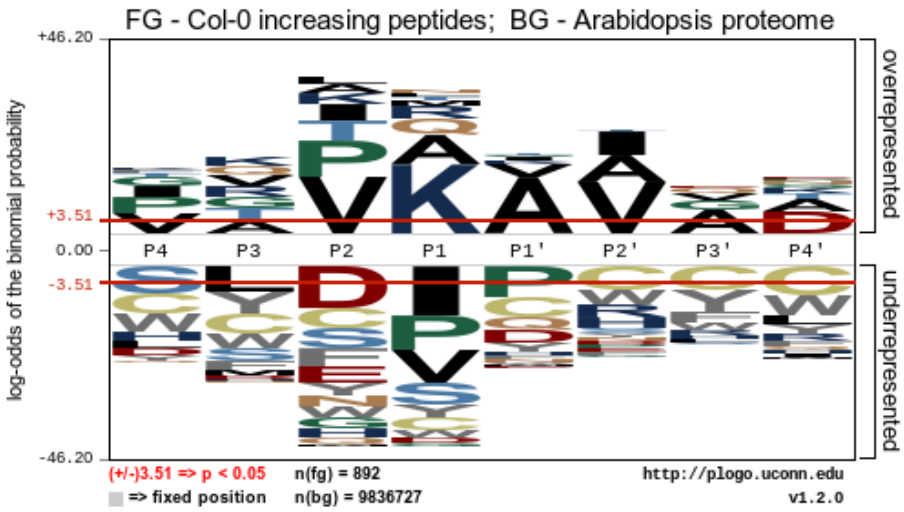


Figure 3.4 Sequence logo visualizations of TOP cleavage specificity using theoretical TOP substrates. Positions with significant residue presence are depicted as amino acid letters sized above the red line[32].

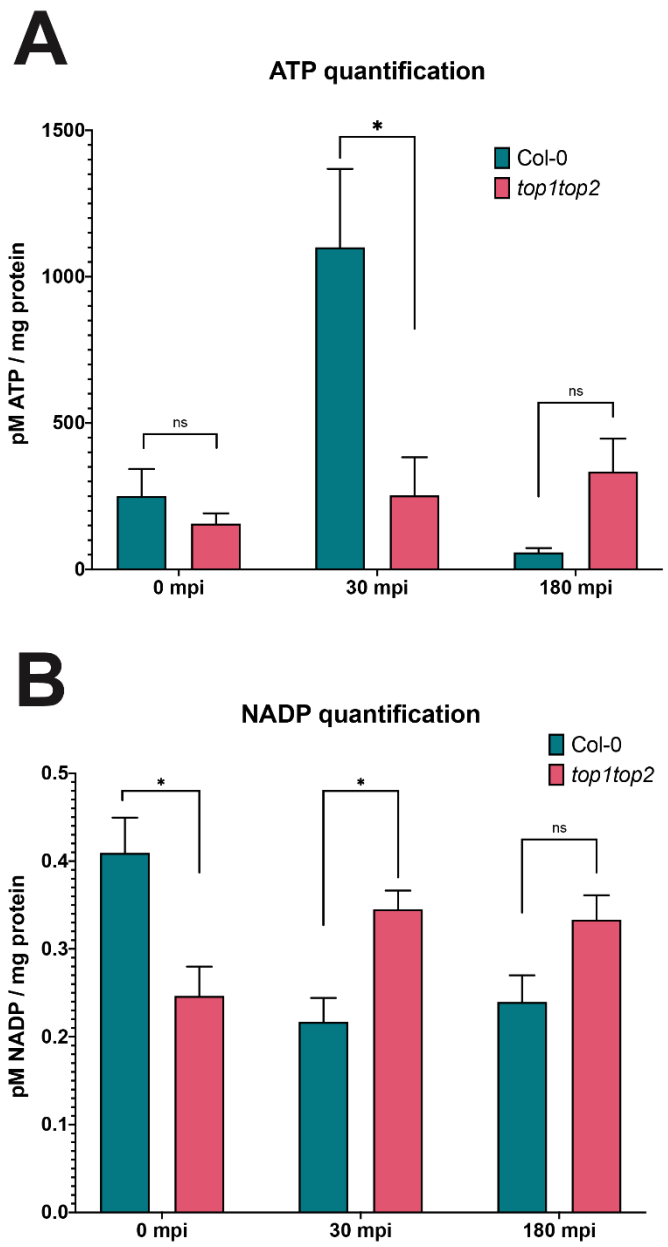


Figure 3.5 Metabolic state of Arabidopsis Col-0 and *top1top2* plant lines following inoculation with *Pst* avrRpt2. (A) ATP was quantified using a luciferase assay (*: $p < 0.05$, ns: not significant, following a paired student's *t*-test). (B) $NADP^+$ was quantified using an enzyme cycling assay (*: $p < 0.05$, ns: not significant, following a paired student's *t*-test).

REFERENCES

1. Chagas, F. O.; Pessotti, R. D. C.; Caraballo-Rodríguez, A. M.; Pupo, M. T. Chemical signaling involved in plant-microbe interactions. *Chem. Soc. Rev.* **2018**, *47*(5), 1652–1704.

2. Jones, J. D. G.; Dangl, J. L. The plant immune system. *Nature* **2006**, *444*(7117), 323–329.

3. Cui, H.; Tsuda, K.; Parker, J. E. Effector-triggered immunity: From pathogen perception to robust defense. *Annu. Rev. Plant Biol.* **2015**, *66*, 487–511.

4. Axtell, M. J.; Staskawicz, B. J. Initiation of RPS2-specified disease resistance in *Arabidopsis* is coupled to the AvrRpt2-directed elimination of RIN4. *Cell* **2003**, *112*(3), 369–377.

5. Lam, E.; Kato, N.; Lawton, M. Programmed cell death, mitochondria and the plant hypersensitive response. *Nature* **2001**, *411*(6839), 848–853.

6. Heath, M. C. Hypersensitive response-related death. *Program. Cell Death High. Plants* **2000**, *44*(3), 77–90.

7. Yoshimoto, K.; Jikumaru, Y.; Kamiya, Y.; Kusano, M.; Consonni, C.; Panstruga, R.; Ohsumi, Y.; Shirasu, K. Autophagy negatively regulates cell death by controlling NPR1-dependent salicylic acid signaling during senescence and the innate immune response in *Arabidopsis*. *Plant Cell* **2009**, *21*(9), 2914–2927.

8. Zurbriggen, M. D.; Carrillo, N.; Hajirezaei, M. R. ROS signaling in the hypersensitive response: When, where and what for? *Plant Signal. Behav.* **2010**, *5*(4), 393–396.

9. Torres, M. A.; Jones, J. D. G.; Dangl, J. L. Reactive oxygen species signaling in response to pathogens. *Plant Physiol.* **2006**, *141*(2), 373–378.

10. Das, K.; Roychoudhury, A. Reactive oxygen species (ROS) and response of antioxidants as ROS-scavengers during environmental stress in plants. *Front. Environ. Sci.* **2014**, *2*(53).

11. Bassham, D. C. Plant autophagy-more than a starvation response. *Curr. Opin. Plant Biol.* **2007**, *10*(6), 587–593.

12. Díaz-Villanueva, J. F.; Díaz-Molina, R.; García-González, V. Protein folding and mechanisms of proteostasis. *Int. J. Mol. Sci.* **2015**, *16*(8), 17193–17230.

13. Kidrič, M.; Kos, J.; Sabotič, J. Proteases and their endogenous inhibitors in the plant response to abiotic stress. *Rev. Artic.* **2014**, *38*(1), 139–158.

14. van Wijk, K. J. Protein Maturation and Proteolysis in Plant Plastids, Mitochondria, and Peroxisomes. *Annu. Rev. Plant Biol.* **2015**, *66*(1), 75–111.

15. Suhm, T.; Kaimal, J. M.; Dawitz, H.; Peselj, C.; Masser, A. E.; Hanzén, S.; Ambrožič, M.; Smialowska, A.; Björck, M. L.; Brzezinski, P.; Nyström, T.; Büttner, S.; Andréasson, C.; Ott, M. Mitochondrial Translation Efficiency Controls Cytoplasmic Protein Homeostasis. *Cell Metab.* **2018**, *27*(6), 1309–1322.

16. Samant, R. S.; Livingston, C. M.; Sontag, E. M.; Frydman, J. Distinct proteostasis circuits cooperate in nuclear and cytoplasmic protein quality control. *Nature* **2018**, *563*(7731), 407–411.

17. Kessler, J. H.; Khan, S.; Seifert, U.; Le Gall, S.; Chow, K. M.; Paschen, A.; Bres-Vloemans, S. A.; de Ru, A.; van Montfoort, N.; Franken, K. L. M. C.; Benckhuijsen, W. E.; Brooks, J. M.; van Hall, T.; Ray, K.; Mulder, A.; Doxiadis, I. I. N.; van Swieten, P. F.; Overkleeft, H. S.; Prat, A.; Tomkinson, B.; Neefjes, J.; Kloetzel, P. M.; Rodgers, D. W.; Hersh, L. B.; Drijfhout, J. W.; van Veelen, P. A.; Ossendorp, F.; Melief, C. J. M. Antigen processing by nardilysin and thimet oligopeptidase generates cytotoxic T cell epitopes. *Nat. Immunol.* **2011**, *12*(1), 45–53.

18. Ferro, E. S.; Rioli, V.; Castro, L. M.; Fricker, L. D. Intracellular peptides: From discovery to function. *EuPA Open Proteomics* **2014**, *3*, 143–151.

19. Tavormina, P.; De Coninck, B.; Nikonorova, N.; De Smet, I.; Cammuea, B. P. A. The plant peptidome: An expanding repertoire of structural features and biological functions. *Plant Cell* **2015**, *27*(8), 2095–2118.

20. Kmiec, B.; Teixeira, P. F.; Murcha, M. W.; Glaser, E. Divergent evolution of the M3A family of metallopeptidases in plants. *Physiol. Plant.* **2016**, *157*(3), 380–388.

21. Gomiz-Rüth, F. X. Catalytic domain architecture of metzincin metalloproteases. *J. Biol. Chem.* **2009**, *284*(23), 15353–15357.

22. Westlake, T. J.; Ricci, W. A.; Popescu, G. V.; Popescu, S. C. Dimerization and thiol sensitivity of the salicylic acid binding thimet oligopeptidases TOP1 and TOP2 define their functions in redox-sensitive cellular pathways. *Front. Plant Sci.* **2015**, *6*(327), 1–16.

23. Moreau, M.; Westlake, T.; Zampogna, G.; Popescu, G.; Tian, M.; Noutsos, C.; Popescu, S. The *Arabidopsis* oligopeptidases TOP1 and TOP2 are salicylic acid targets that modulate SA-mediated signaling and the immune response. *Plant J.* **2013**, *76*(4), 603–614.

24. García-Lorenzo, M.; Sjödin, A.; Jansson, S.; Funk, C. Protease gene families in Populus and Arabidopsis. *BMC Plant Biol.* **2006**, *6*(30).

25. Kmiec, B.; Teixeira, P. F.; Berntsson, R. P. A.; Murcha, M. W.; Branca, R. M. M.; Radomiljac, J. D.; Regberg, J.; Sèensson, L. M.; Bakali, A.; Langel, Ü.; Lehtiö, J.; Whelan, J.; Stenmark, P.; Glaser, E. Organellar oligopeptidase (OOP) provides a complementary pathway for targeting peptide degradation in mitochondria and chloroplasts. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*(40), E3761–E3769.

26. Polge, C.; Jaquinod, M.; Holzer, F.; Bourguignon, J.; Walling, L.; Brouquisse, R.

Evidence for the existence in *Arabidopsis thaliana* of the proteasome proteolytic pathway: Activation in response to cadmium. *J. Biol. Chem.* **2009**, *284*(51), 35412–35424.

27. Iannetta, A. A.; Rogers, H. T.; Al-Mohanna, T.; O'Brien, J. N.; Wommack, A. J.; Popescu, S. C.; Hicks, L. M. Profiling thimet oligopeptidase-mediated proteolysis in *Arabidopsis thaliana*. *Plant J.* **2021**, *106*(2), 336–350.

28. Lindsey III, B. E.; Rivero, L.; Calhoun, C. S.; Grotewold, E.; Brkljacic, J. Standardized method for high-throughput sterilization of *Arabidopsis* seeds. *J. Vis. Exp.* **2017**, *128*, e56587.

29. Whalen, M. C.; Innes, R. W.; Bent, A. F.; Staskawicz, B. J. Identification of *Pseudomonas syringae* pathogens of *Arabidopsis* and a bacterial locus determining avirulence on both *Arabidopsis* and soybean. *Plant Cell* **1991**, *3*(1), 49–59.

30. Käll, L.; Canterbury, J. D.; Weston, J.; Noble, W. S.; MacCoss, M. J. Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat. Methods* **2007**, *4*(11), 923–925.

31. Gianetto, Q. G. imp4p: Imputation for Proteomics https://cran.r-project.org/web/packages/imp4p/index.html (accessed Aug 6, 2019).

32. O'Shea, J. P.; Chou, M. F.; Quader, S. A.; Ryan, J. K.; Church, G. M.; Schwartz, D. PLogo: A probabilistic approach to visualizing sequence motifs. *Nat. Methods* **2013**, *10*(12), 1211–1212.

33. Perez-Riverol, Y.; Csordas, A.; Bai, J.; Bernal-Llinares, M.; Hewapathirana, S.; Kundu, D. J.; Inuganti, A.; Griss, J.; Mayer, G.; Eisenacher, M.; Erez, E. P.; Uszkoreit, J.; Pfeuffer, J.; Sachsenberg, T.; Yılmaz, S.; Tiwary, S.; Cox, J.; Audain, E.; Walzer, M.; Jarnuczak, A. F.; Ternent, T.; Brazma, A.; Vizcaíno, J. A. The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* **2019**, *47*(D1), D442–D450.

34. Lone, A. M.; Kim, Y. G.; Saghatelian, A. Peptidomics methods for the identification of peptidase-substrate interactions. *Curr. Opin. Chem. Biol.* **2013**, *17*(1), 83–89.

35. Cavalcanti, D. M. L. P.; Castro, L. M.; Neto, J. C. R.; Seelaender, M.; Neves, R. X.; Oliveira, V.; Forti, F. L.; Iwai, L. K.; Gozzo, F. C.; Todiras, M.; Schadock, I.; Barros, C. C.; Bader, M.; Ferro, E. S. Neurolysin knockout mice generation and initial phenotype characterization. *J. Biol. Chem.* **2014**, *289*(22), 15426–15440.

36. Conlin, C. A.; Miller, C. G. Dipeptidyl carboxypeptidase and oligopeptidase A from *Escherichia coli* and *Salmonella typhimurium*. *Methods Enzymol.* **1995**, *248*, 567–579.

37. Dando, P. M.; Brown, M. A.; Barrett, A. J. Human thimet oligopeptidase. *Biochem. J.* **1993**, *294*(2), 451–457.

38. Rei Liao, J. Y.; van Wijk, K. J. Discovery of AAA+ Protease Substrates through Trapping

Approaches. *Trends Biochem. Sci.* **2019**, *44*(6), 528–545.

39. Choi, J.; Tanaka, K.; Cao, Y.; Qi, Y.; Qiu, J.; Liang, Y.; Lee, S. Y.; Stacey, G. Identification of a plant receptor for extracellular ATP. *Science* **2014**, *343*(6168), 290–294.

40. Tanaka, K.; Choi, J.; Cao, Y.; Stacey, G. Extracellular ATP acts as a damage-associated molecular pattern (DAMP) signal in plants. *Front. Plant Sci.* **2014**, *5*(446).

41. Tripathi, D.; Zhang, T.; Koo, A. J.; Stacey, G.; Tanaka, K. Extracellular ATP Acts on Jasmonate Signaling to Reinforce Plant Defense. *Plant Physiol.* **2018**, *176*(1), 511–523.

42. Kozuleva, M.; Goss, T.; Twachtmann, M.; Rudi, K.; Trapka, J.; Selinski, J.; Ivanov, B.; Garapati, P.; Steinhoff, H. J.; Hase, T.; Scheibe, R.; Klare, J. P.; Hanke, G. T. Ferredoxin: NADP(H) oxidoreductase abundance and location influences redox poise and stress tolerance. *Plant Physiol.* **2016**, *172*(3), 1480–1493.

CHAPTER 4: IreK-Mediated, Cell Wall Protective Phosphorylation in *Enterococcus faecalis*

*****Submitted for publication. Authors: Iannetta, A. A.; Minton, N. E.; Uitenbroek, A. A.; Little, J. L.; Stanton, C. R.; Kristich, C. J.; and Hicks, L. M.

4.1 Introduction

Enterococcus faecalis is a commensal Gram-positive bacterium found in the human gastrointestinal (GI) tract. This bacterium is an opportunistic pathogen with intrinsic resistance to a wide variety of cell wall-active antibiotics and disinfectants, including the β-lactam containing cephalosporins and the antiseptic chlorhexidine[1–4]. Treatment with cell wall-active antimicrobials creates an opportunity for resistant *E. faecalis* to proliferate in the gut and translocate to other locations in the body[5–7], which can result in serious nosocomial infections such as bacteremia and endocarditis[5,6,8]. Understanding the *E. faecalis* response to cell wall-active antimicrobials could inform preventive treatments for these infections.

Protein phosphorylation has been recognized for its role in bacterial pathogenesis[9–12]. In *E. faecalis*, the eukaryotic-type Ser/Thr kinase IreK (intrinsic resistance of enterococci kinase) belongs to a family of transmembrane kinases defined by the presence of multiple extracellular PASTA (penicillin-binding protein and serine/threonine kinase associated) domains[13]. These PASTA kinases are ubiquitous in *Actinobacteria* and *Firmicutes* and regulate a variety of important functions, including cell division, virulence, antibiotic resistance, and toxin production[14–17]. Collectively, PASTA kinases have conserved functions across Gram-positive bacteria involving monitoring and modulating processes associated with the cell wall. The

proposed model for PASTA kinase signaling in bacteria includes ligand binding (*e.g.*, released peptidoglycan fragments) to PASTA domains outside the cell, causing kinase dimerization[18–21]. This promotes kinase autophosphorylation inducing a conformational change that increases kinase activity and subsequent phosphorylation of downstream substrates. Antimicrobial-triggered PASTA kinase phosphorylation events have been previously identified in methicillin-resistant *Staphylococcus aureus*[22].

E. faecalis IreK maintains cell wall integrity, possibly by regulating peptidoglycan biosynthesis and metabolism[7]. It is hypothesized that antibiotic-induced cell wall stress leads to activation of IreK-mediated phosphorylation signaling pathways to mitigate and repair the damage[21]. Absence of IreK leads to cell envelope defects and increased susceptibility to a variety of cephalosporins, including ceftriaxone[21,23,24]. Further, knock out mutants display a loss in the long-term ability to colonize the GI tract, an essential factor for pathogenesis[23]. Increased IreK phosphorylation in response to ceftriaxone treatment has been shown to correlate with antimicrobial resistance[21], and three conserved sites on the activation loop (T163, T166, and T168) are essential for IreK autophosphorylation and activation, and increase antibiotic resistance when phosphorylated[24].

IreK's role in *E. faecalis* cell wall-active antimicrobial resistance is demonstrably regulated through phosphorylation[21,24,25]. However, the only confirmed direct IreK substrate is IreB, a small protein that is well conserved in Gram-positive bacteria[26]. While the function of IreB is not known in *E. faecalis*, the phosphorylation of its homolog ReoM in *Listeria monocytogenes* controls proteolytic degradation of MurA, which catalyses the first step in peptidoglycan biosynthesis[27]. IreK-mediated phosphorylation of IreB on a conserved set of threonine residues (Thr4/Thr7) was identified through *in vitro* kinase assays and phosphoamino acid analysis[28], and

an IreB homolog in *Streptococcus* provided evidence of phosphorylation on conserved sites[29]. Ceftriaxone treatment of wild type *E. faecalis* led to an increase in IreB phosphorylation, showing that these events are stress-modulated[21]. Additionally, IreK was shown to positively affect two component system (TCS) signaling, a common signal transduction mechanism used by bacteria to sense and respond to changing environments[30–32]. The histidine kinase CroS of the *E. faecalis* TCS was determined to be phosphorylated in response to cell wall-active antimicrobial agents in an IreK-dependent manner, but CroS could not be confirmed as a direct IreK substrate as *in vitro* kinase assays did not lead to increased CroS phosphorylation[30]. Other direct substrates of IreK and downstream effectors in this signaling pathway are unclear. Elucidation of additional substrates is critical to provide increased understanding of how IreK influences antimicrobial resistance and identify new strategies and targets for therapeutic intervention to prevent or overcome infections by multidrug-resistant *E. faecalis*.

Herein, stress-modulated phosphorylation events between untreated and antimicrobial-treated wild type *E. faecalis* are revealed by mass spectrometry-based label free quantitative (LFQ) phosphoproteomics (Figure 4.1). *E. faecalis* was cultured with antimicrobials possessing different mechanisms of action (ceftriaxone - inhibition of penicillin-binding proteins that catalyze essential bacterial peptidoglycan crosslinking, and chlorhexidine - disruption of cell membranes), allowing for the determination of treatment-specific modifications. Similar examination of an Δ*ireK* mutant knockout strain identified potential *in vivo* IreK phosphorylation substrates and antimicrobial resistance effectors that begin to define the IreK signaling network. Both of the previously described IreK-dependent phosphorylation events were identified in the differential phosphoproteomics data, and two newly discovered IreK substrates were confirmed by orthogonal assays to undergo IreK-dependent phosphorylation *in vivo* and/or *in vitro*.

4.2 Materials and Methods

4.2.1 Cell Culturing and Antimicrobial Treatment

For phosphoproteomics analyses, *E. faecalis* strains (Table S1, Lab_Members\Tony\ThesisSupp) were cultured in semi-defined MM9YE medium[33] supplemented with 0.5% glucose. Stationary-phase cultures were diluted to OD_{600} = 0.01 and incubated at 37 °C until mid-exponential phase (OD_{600} = 0.5). Cultures were split, antimicrobials (ceftriaxone, 256 μg/ml; or chlorhexidine, 0.00024%) were added to their corresponding biological replicates, and incubation continued at 37 °C for 20 min. The concentration of ceftriaxone was chosen based on prior work demonstrating that exposure of *E. faecalis* cells to similar ceftriaxone concentrations led to robust enhancement of signaling by IreK[21]. The concentration of chlorhexidine chosen was determined to be 0.5x of the inhibitory concentration in preliminary experiments. Cells were then chilled in an ice-water bath and collected by centrifugation before flash-freezing the cell pellets in a dry-ice/ethanol bath.

4.2.2 Protein Extraction

Five biological replicates were used for each genotype (*i.e.*, wild type and Δ*ireK* null mutant) and treatment (*i.e.*, untreated, ceftriaxone, chlorhexidine). Unless noted, all steps were performed on ice. Cell pellets (2 g) were resuspended in 10 mL of PBS, pH 7.4 containing 1x cOmplete protease inhibitor and phosSTOP phosphatase inhibitor cocktails (Roche, Indianapolis, IN). Cells were lysed via three, 1 min rounds of probe sonication (Heat Systems Inc., Farmingdale, NY) with 1 s pulses containing a 50% duty cycle (0.5 s on, 0.5 s off). Following sonication, 1% SDS was added to the homogenate and samples were vortexed. Cellular debris was cleared by centrifugation at 2,000 x*g* for 10 min at 4°C and proteins were precipitated from the supernatant using 5 volumes of cold 100 mM ammonium acetate in methanol. Following incubation for 16 h

at -80 °C, proteins were pelleted by centrifugation at 2000 x*g* for 10 min at 4 °C. To remove excess reagents, the pellets were washed once with 10 mL of 100 mM ammonium acetate in methanol and once with 10 mL of 70% ethanol. Proteins were resuspended in 2 mL of 50 mM Tris-HCl, pH 8.0 with 0.5% SDS and 4 M urea, and the remaining cellular debris was removed by centrifugation. Protein concentrations were estimated using the CB-X assay (G-Biosciences, St. Louis, MO) according to the manufacturer's protocol, and were normalized across replicates with additional resuspension buffer prior to digestion.

4.2.3 Protein Digestion

Samples were reduced using 10 mM dithiothreitol for 30 min at room temperature (RT) and alkylated with 40 mM iodoacetamide for 45 min in darkness at RT. Proteins were precipitated with 1 mL of cold acetone and centrifuged at 15,000 x*g* and 4 °C for 5 min. Proteins were resuspended in 500 µL of 100 mM Tris-HCl, pH 8 with 4 M urea and diluted two-fold to 2 M urea with additional 100 mM Tris-HCl, pH 8. Digestion was performed with Trypsin Gold (Promega, Madison, WI) at RT for 16 h using a protease:protein ratio of 1:100 (w/w).

4.2.4 Reversed-Phase Solid-Phase Extraction

Samples were desalted with 50 mg/1.0 mL Sep-Pak C18 cartridges (Waters, Milford, MA) using a 24-position vacuum manifold (Phenomenex, Torrance, CA) at a flow rate of 1 drop/s. Resin was first pre-eluted using 1 mL of 80% acetonitrile/0.1% trifluoroacetic acid (TFA) before equilibration with 1 mL of 0.1% TFA. Samples were acidified to pH < 3 using 10% TFA, loaded onto the cartridges in two passes, and then washed using 1 mL of 0.1% TFA. Peptides were eluted using 1 mL of 80% acetonitrile/0.1% TFA. Aliquots (100 µg) were taken for global proteomic analysis before concentration by vacuum centrifugation.

4.2.5 Phosphopeptide Enrichment

The rest of the sample not taken for global proteomics analysis was resuspended in 150 µL of 80% acetonitrile/1% TFA containing 25 mg/mL phthalic acid. Phosphopeptide enrichment was performed using 3 mg/200 µL Titansphere Phos-TiO tips (GL Sciences Inc., Torrance, CA), as described previously[34]. Each step in the enrichment was followed with centrifugation at 1000 x*g* for 2 min. The resin was pre-eluted using 100 µL of 20% acetonitrile/5% ammonium hydroxide before conditioning with 100 µL of 80% acetonitrile/1% TFA and 80% acetonitrile/1% TFA containing 25 mg/mL phthalic acid. Samples were loaded on the tips in three passes and then washed in two passes using 100 µL of 80% acetonitrile/1% TFA containing 25 mg/mL phthalic acid and three passes using 100 µL of 80% acetonitrile/1% TFA to remove non-specifically bound peptides. Phosphopeptides were eluted using 200 µL of 20% acetonitrile/5% ammonium hydroxide and concentrated by vacuum centrifugation. Samples were desalted by reversed-phase solid-phase extraction as described above prior to LC-MS/MS.

4.2.6 LC-MS/MS Analysis

Samples were analyzed using an Acquity UPLC M-Class System (Waters) coupled to a Q Exactive HF-X mass spectrometer (Thermo Fisher Scientific, Waltham, MA). Mobile phase A consisted of water with 0.1% formic acid (Thermo Fisher Scientific) and mobile phase B was acetonitrile with 0.1% formic acid. Injections were made to a Symmetry C18 trap column (100 Å, 5µm, 180µm x 20 mm; Waters) with a flow rate of 5 µL/min for 3 min using 99% A and 1% B. Peptides were then separated on a HSS T3 C18 column (100 Å, 1.8µm, 75µm x 250 mm; Waters) using a linear gradient of increasing mobile phase B at a flow rate of 300 nL/min. Mobile phase B increased from 5% to 35% in 90 min before ramping to 85% in 5 min, where it was held for 10 min before returning to 5% in 2 min and re-equilibrating for 13 min. The mass

spectrometer was operated in positive polarity and the Nanospray Flex source had spray voltage floating at 2.1 kV, capillary temperature at 320 °C, and funnel RF level at 40. MS survey scans were collected with a scan range of 350 – 2000 *m/z* at a resolving power of 120,000 and an AGC target of 3×10^6 with a maximum injection time of 50 ms. A top 20 data-dependent acquisition was used where HCD fragmentation of precursor ions having +2 to +7 charge state was performed using a normalized collision energy setting of 28. MS/MS scans were performed at a resolving power of 30,000 and an AGC target of 1×10^5 with a maximum injection time of 100 ms. Dynamic exclusion for precursor *m/z* was set to a 10 s window.

4.2.7 Proteomics Database Searching and Label-Free Quantification

Acquired spectral files (*.raw) were imported into Progenesis QI for proteomics (Waters, version 2.0). Peak picking sensitivity was set to the maximum of five and a reference spectrum was automatically assigned. Total ion chromatograms (TICs) were then aligned to minimize run-to-run differences in peak retention time. Each sample received a unique factor to normalize all peak abundance values resulting from systematic experimental variation. Alignment was validated (> 80% score) and a combined peak list (*.mgf) was exported for peptide sequence determination by Mascot (Matrix Science, version 2.5.1; Boston, MA). Database searching was performed against the *Enterococcus faecalis* National Center for Biotechnology Information database (https://www.ncbi.nlm.nih.gov/genome/808?genome_assembly_id=168518, 2,559 canonical entries, accessed 08/2020) with sequences for common laboratory contaminants (https://www.thegpm.org/cRAP/, 116 entries, accessed 08/2020) appended. For global proteome samples, target-decoy searches of MS/MS data used a trypsin protease specificity with the possibility of two missed cleavages, peptide/fragment mass tolerances of 15 ppm/0.02 Da, fixed modification of cysteine carbamidomethylation, and variable modifications of N-terminus

acetylation and methionine oxidation. For phosphoproteome enriched samples, the same parameters were used with the addition of variable modifications of serine, threonine, and tyrosine phosphorylation. Significant peptide identifications above the identity or homology threshold were adjusted to less than 1% peptide FDR using the embedded Percolator algorithm[35] and imported to Progenesis for peak matching. Identifications with a Mascot score less than 13 were removed from consideration in Progenesis before exporting both "Peptide Measurements" and "Protein Measurements" from the "Review Proteins" stage.

4.2.8 Proteomics Data Processing

Data were parsed using custom scripts written in R for pre-processing and statistical analysis (https://github.com/hickslab/QuantifyR).

For global proteomic analysis, leading protein accessions were considered from the "Protein Measurements" data and kept if there were ≥ 2 shared peptides and ≥ 1 unique peptide assigned. Proteins were removed if there was not at least one condition with > 50% nonzero values across the Progenesis-normalized abundance columns. Values were $\log_2$-transformed and a conditional imputation strategy was applied using the imp4p package[36], where conditions with at least one nonzero value had missing values imputed using the *impute.rand* function with default parameters. For cases where a condition had only missing values, the *impute.pa* function was used to impute small numbers centered on the lower 2.5% of values in each replicate.

For the phosphoproteome analysis, we initially summarized the "Peptide Measurements" data such that each row was a unique MS1 peak feature (*i.e.*, with distinct precursor mass and retention time coordinates) matched with a peptide sequence identification from the database searching results. Peak features with an identification from peptide ion deconvolution were

removed from the data. Some features were initially duplicated and matched with peptides having identical sequences, modifications, and score but alternate protein accessions. These groups were reduced to satisfy the principle of parsimony and represented by the protein accession with the highest number of unique peptides found in the "Protein Measurements" data or else the protein with the largest confidence score assigned by Progenesis. Some features were also duplicated with differing peptide identifications and were reduced to include only the peptide with the highest Mascot ion score. Results were then filtered for peptides containing at least one phosphorylation-modified Ser, Thr, or Tyr residue. An identifier was created by joining the protein accession of each peptide to the particular site(s) of modification in the protein sequence. Each dataset was reduced to unique identifiers by summing the abundance of all contributing peak features (*i.e.*, different peptide charge states, missed cleavages, and combinations of additional variable modifications). Summarized identifiers were represented by the information on the peptide with the highest Mascot score in each group. Identifiers were removed if there was not at least one condition with > 50% nonzero values across the Progenesis-normalized abundance columns. Phosphopeptides were normalized by dividing the phosphopeptide abundances with their protein abundances from the global proteome data in each replicate[37–40]. The phosphopeptides derived from proteins that were not identified in the global proteome data could not be normalized and were not included in the phosphoproteome statistical analysis. Values were $\log_2$-transformed and a conditional imputation strategy was applied using the imp4p package[36], as described above. Phosphosite localization was appended using an implementation of the Mascot Delta score[41] with confident site localization considered a Mascot Delta score > 10, which is a localization probability > 90%.

4.2.9 Proteomics Statistical Analysis

For the analysis concerning the optimization of phosphoproteome methods, identifiers were removed if there was a zero value across any of the loading amount raw abundance columns. Linear regression analysis was performed for each identifier using abundance values measured across increasing peptide loading amounts, and the median coefficient of determination (R^2) was calculated. For visualization purposes, abundance values across peptide loading amounts for each identifier were converted to standardized z-scores.

For the differential global proteome analysis, statistical significance was determined using a two-tailed, equal variance *t*-test and the method of Benjamini and Hochberg (BH) was used to correct *p*-values for multiple comparisons[42]. Fold change was calculated by the difference of the mean abundance values between conditions being compared. Only observations with FDR-adjusted *p*-value < 0.05 and $\log_2$-transformed fold change ± 1 were considered significantly different.

For the differential phosphoproteome analysis, statistical significance was determined using a two-tailed, equal variance *t*-test or a one-way analysis of variance (ANOVA), both with BH-correction. For the *t*-test, fold change was calculated by the difference of the mean abundance values between conditions being compared. Only observations with FDR-adjusted *p*-value < 0.05 and $\log_2$-transformed fold change ± 1 after a two-sided, equal variance *t*-test were considered significantly different. For the ANOVA, only observations with FDR-adjusted *p*-value < 0.05 were considered significantly different. Unsupervised hierarchical clustering was performed on significantly different identifiers to group together similarly changing abundance trends across conditions (*i.e.*, with treatment). For visualization purposes, abundance values across treatments for each identifier were converted to standardized z-scores.

4.2.10 Enrichment of GpsB-His$_6$ from *E. faecalis* Cells

Stationary-phase cultures of *E. faecalis* cells harboring the empty vector (as control) or plasmid overexpressing GpsB-His$_6$ were diluted to OD_{600} = 0.01 and grown to exponential phase in Mueller-Hinton broth supplemented with 10 µg/mL chloramphenicol (for plasmid maintenance) at 37 °C. Cells were then crosslinked with 1% formaldehyde for 10 min at RT and subsequently quenched with 0.5 M glycine for at least 5 min at RT. Cells were lysed by treatment with 15 mg/mL lysozyme + 1x phosphatase inhibitor for 30 min at 37 °C followed by bead beating 6 times on/off for 30 s each at RT. The lysates were boiled for at least 30 min to reverse crosslinks and incubated with binding-buffer-equilibrated Ni-charged resin for 1 h at RT. The resin was washed with binding buffer (50 mM Tris, 300 mM NaCl, 5 mM imidazole [pH = 8.0]) and GpsB-His$_6$ was recovered from the resin with elution buffer (50 mM Tris, 300 mM NaCl, 500 mM imidazole [pH = 8.0]). The input and elution fractions were subjected to SDS-PAGE and immunoblotting was performed using antisera to detect total GpsB or anti-pThreonine (Cell Signaling Technology, Danvers, MA) to detect phosphorylated GpsB.

4.2.11 Protein Purification

Recombinant His$_6$-SUMO-GpsB, wild type His$_6$-IreK-n (intracellular domain only), and a catalytically impaired His$_6$-IreK-n K41R mutant were each purified from *Escherichia coli* BL21 (DE3) or NiCo21 (DE3) cells. Overnight cultures were diluted 1:50 into LB supplemented with 50 µg/mL kanamycin and grown for 3 h at 37 °C, then induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for 1 h. Cells were pelleted by centrifugation at 10,000 rpm for 10 min at 4°C, resuspended in binding buffer (50 mM Tris, 300 mM NaCl, 5 mM imidazole [pH = 8.0]), and then lysed using a French Press. The lysates were centrifuged at 4,500 rpm for at least 25 min at 4 °C and passed through a 0.2 µM filter. Filtered lysates were applied to Ni-charged

resin previously equilibrated with binding buffer. The protein-bound resin was then washed with wash buffer (50 mM Tris, 300 mM NaCl, 20 mM imidazole [pH = 8.0]) and bound His_6-tagged proteins were subsequently recovered using elution buffer (50 mM Tris, 300 mM NaCl, 500 mM imidazole [pH = 8.0]). The elution fractions containing the His_6-IreK-n proteins were each dialyzed against 1 L of 50 mM Tris, 25 mM NaCl (pH = 7.5) for 1 h, then overnight at 4 °C. The elution fraction containing His_6-SUMO-GpsB was dialyzed for 1 h at 4 °C against 1 L of 50 mM Tris, 150 mM NaCl (pH = 8.0), then His-GB1-Ulp1 protease and 0.1% β-mercaptoethanol were added to the dialysis bag and dialyzed overnight at 4 °C in 1 L of fresh buffer. After dialysis, cleaved (untagged) GpsB was applied to binding-buffer-equilibrated Ni-charged resin and GpsB was collected in the flowthrough and wash fractions as verified by SDS-PAGE and GelCode Blue staining. Untagged GpsB was subjected to size exclusion chromatography using a HiLoad Superdex 200 pg 16/600 column (Cytiva, Marlborough, MA) with a running buffer of 50 mM Tris, 25 mM NaCl (pH = 7.5) at a flow rate of 1 mL/min. Fractions containing GpsB were then concentrated using an Amicon® Ultra 15 mL 3 kDa cutoff concentrator.

4.2.12 *In Vitro* Kinase Assays

Purified untagged GpsB (14.2 µM) was incubated at 37 °C in the presence or absence of recombinant wild type His_6-IreK-n (0.33 µM) or catalytically impaired His_6-IreK-n K41R mutant (0.33 µM) in 50 mM Tris (pH = 7.5), 25 mM NaCl, and 5 mM $MgCl_2$. ATP (2 mM) was added to initiate the reactions and aliquots were removed at intervals (0, 75, and 150 min), mixed with. 5X SDS-PAGE Laemmli buffer, boiled for 5 min, and subjected to SDS-PAGE. The gels were stained with Pro-Q™ Diamond Phosphoprotein Gel Stain (Invitrogen, Waltham, MA) followed by SYPRO™ Ruby Protein Gel Stain (Invitrogen) according to the manufacturer's instructions to detect phosphorylated GpsB and total GpsB, respectively.

4.2.13 Immunoblot Analysis on Whole-Cell Lysates

Stationary-phase cultures of *E. faecalis* cells were diluted to OD_{600} = 0.01 and grown to exponential phase in Mueller-Hinton broth at 37 °C. Cells were harvested by mixing with an equal volume of cold ethanol/acetone (1:1), collected by centrifugation, and washed once with water. Cell pellets were suspended in lysozyme buffer (10 mM Tris, 20% sucrose, 50 mM NaCl [pH 8.0]), treated with lysozyme (4 mg/ml) for 20 min at 37 °C, and mixed with 5x Laemmli SDS-PAGE loading buffer to create whole-cell lysates. For electrophoresis, acrylamide gels were prepared the day of the experiment (10% acrylamide for standard gels; 6% acrylamide with 50 uM Phos-tag and 100 uM $Zn(NO_3)_2$ for Phos-tag gels). Phos-tag gels were soaked in 5 mM EDTA for 20 min twice before transfer to PVDF. Membranes were probed with custom antisera for MltG or RpoA (loading control) and HRP-conjugated secondary antibody before developing with SuperSignal (Thermo Fisher Scientific) using a ChemiDoc touch imaging system (BioRad, Hercules, CA).

4.2.14 Data Availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository[43] and can be accessed with the dataset identifier PXD023265 and 10.6019/PXD023265.

4.3 Results and Discussion

4.3.1 Optimization of Phosphopeptide Enrichment to Maximize Coverage

The low prevalence of bacterial protein phosphorylation compared to eukaryotic systems presents a substantial technical challenge for the analysis of bacterial phosphoproteomics[44]. To ensure robust coverage of the *E. faecalis* phosphoproteome, TiO_2 phosphopeptide enrichment was optimized. Six biological replicates of wild type *E. faecalis* were extracted, combined, and

aliquoted to produce six different protein loading amounts ranging from 0.5-5 mg. The upper limit was determined by the ability to consistently extract 5 mg protein from each biological replicate (~2 g cell pellet weight), given the limit in cell culture volumes to perform the full experiment at the same time. LC-MS/MS revealed 91 phosphopeptides from 61 proteins quantified across this enriched sample set (Table S2, Lab_Members\Tony\ThesisSupp). This optimization experiment yielded lower phosphoproteome coverage compared to the full experiment described below due to the combination of six replicates at varying protein loading amounts vs. thirty replicates at the 5 mg protein loading amount. To examine quantification accuracy and ensure resin saturation was not occurring due to non-specific binding of unphosphorylated peptides, an analysis of the phosphopeptide abundances was performed (Figure S1A, Lab_Members\Tony\ThesisSupp). Quantitative ratios (R) were calculated by dividing the phosphopeptide abundances of the 5 mg enriched sample by each of the other samples (Figure S1B, Lab_Members\Tony\ThesisSupp). Measured median R values across the higher loading amounts were accurate and closely matched the expected ratios (<10% error), while the median R values for the 0.5 and 1 mg samples were exaggerated (>50% error) suggesting that these enrichments provided phosphopeptides near the lower limit of quantification for LFQ analysis. Linear regression analysis was performed for each phosphopeptide. The median coefficient of determination (R^2) for the phosphopeptides was 0.96 and 86% of the phosphopeptides had an $R^2 > 0.8$ across samples, demonstrating that the relative abundance of each phosphopeptide was highly correlated with the changing loading amounts. While these methods are optimized for robust bacterial phosphoproteome coverage, it does not compare to the depth detected for eukaryotic systems (20,000 phosphopeptides identified in *Saccharomyces cerevisiae*[45] and 50,000 phosphopeptides identified in a HeLa S3 human cell

line[46]), displaying the lower prevalence of bacterial protein phosphorylation. Among other Gram-positive bacteria, typically less than 200 phosphopeptides are detected, again demonstrating lower detection of protein phosphorylation in bacteria[47–49]. Recently, efforts to optimize bacterial phosphoproteome coverage using Fe^{3+}-IMAC produced the identification of ~4000 phosphopeptides in *Staphylococcus aureus*, but this optimized method still only resulted in the identification of ~300 phosphopeptides in *Bacillus subtilis*[50]. Thus, additional reflection into current phosphoproteomics methodologies is necessary to determine if the low coverage of Gram-positive bacterial phosphoproteomes obtained previously is due to technical limitations or actual low prevalence of protein phosphorylation in these species. As it stands, these results provide confidence that TiO_2 resin saturation is not occurring and enabled phosphoproteome analysis on 5 mg enriched loading amounts.

4.3.2 Cell Wall-Active Antimicrobial-Modulated Phosphorylation

Phosphorylation plays an important role in the *E. faecalis* cell wall stress response, which drives its virulence and resistance to antimicrobials[21,24,25]. To gain insights into the signaling processes involved, phosphorylation dynamics of antimicrobial-treated *E. faecalis* were examined. While ceftriaxone inhibits penicillin-binding proteins, chlorhexidine disrupts bacterial cytoplasmic membranes. These differing mechanisms of action facilitated determination of treatment-specific modifications. Quantitative phosphoproteomics was processed in a single experiment and included two *E. faecalis* strains, wild type and Δ*ireK*, under untreated, chlorhexidine-, and ceftriaxone-treated conditions (Table S3, Lab_Members\Tony\ThesisSupp). To describe the general landscape of protein phosphorylation during stress response signal transduction, the phosphoproteomes of wild type untreated, chlorhexidine-, and ceftriaxone-treated were investigated (Table S3, columns K-Y, Lab_Members\Tony\ThesisSupp). Following

this, a comparison between the wild type and Δ*ireK* phosphoproteomes across these different treatments was conducted to distinguish the stress-mediated phosphorylation dependent on IreK (Table S3, columns K-AN, Lab_Members\Tony\ThesisSupp).

To assess cell wall-active antimicrobial-modulated phosphorylation events, differential LFQ phosphoproteomics was performed across biological replicates of untreated, ceftriaxone-, and chlorhexidine-treated wild type *E. faecalis*. Overall, 167 phosphosites (45% phosphosites localized) from 161 unique phosphopeptides and 94 unique proteins were quantified (Table S3, columns K-Y, Lab_Members\Tony\ThesisSupp). To ensure that observed phosphorylation changes were due to differential phosphorylation, protein abundance changes were assessed for the unenriched global wild type proteome. From this, 1351 proteins were quantified (Table S4, Lab_Members\Tony\ThesisSupp), including 95% of the proteins identified in the phosphoproteome. This revealed 95 proteins with a significant fold change (FC ≥ 2, FDR-adjusted *p*-value < 0.05) between untreated vs. chlorhexidine-treated samples (Figure S2A, Lab_Members\Tony\ThesisSupp) and 183 proteins with a significant fold change between untreated vs. ceftriaxone-treated wild type samples (Figure S2B, Lab_Members\Tony\ThesisSupp). When chlorhexidine- and ceftriaxone-treated wild type samples were compared, 368 proteins had a significant fold change between the different treatments (Figure S2C, Lab_Members\Tony\ThesisSupp). Given the apparent changes in protein expression or turnover in response to these treatments, phosphopeptide abundances were normalized by dividing each replicate with their protein abundances from the global proteome data (Table S5, columns J-X, Lab_Members\Tony\ThesisSupp).

Post-normalization, one-way ANOVA with BH correction revealed that 87 phosphopeptides from 53 proteins in the wild type strain were significantly changing (FDR-adjusted *p*-value <

0.05) across the treatments (Table S5, Lab_Members\Tony\ThesisSupp). Hierarchical clustering was performed on the significantly changing phosphopeptides, resulting in four clusters (Figure 4.2, Table S5, column I, Lab_Members\Tony\ThesisSupp). The largest cluster, containing 45 phosphopeptides, revealed peptides that increased in phosphorylation within both the chlorhexidine- and ceftriaxone-treated samples compared to the untreated samples (cluster A). Delineation of stress-specific phosphorylation changes following chlorhexidine and ceftriaxone treatments revealed 11 phosphopeptides were significantly increased specifically after chlorhexidine treatment (cluster B), while 13 phosphopeptides were significantly increased specifically after ceftriaxone treatment (cluster C). The last cluster, containing 18 phosphopeptides, included peptides that decreased in phosphorylation within both the chlorhexidine- and ceftriaxone-treated samples compared to the untreated samples (cluster D).

The known IreK-mediated IreB (AEA93661.1) phosphosites Thr4 and Thr7 exhibited increased phosphorylation after both chlorhexidine and ceftriaxone treatments (Table 4.1, Figure S3A, Lab_Members\Tony\ThesisSupp). While the function of IreB in *E. faecalis* has not been determined, this phosphorylation is known to disrupt its ability to dimerize, which is essential for its biological activity[26]. Prior work assayed antimicrobial susceptibility with highly antimicrobial-resistant Δ*ireB* mutants complemented with various IreB variants[26]. These variants contained a series of mutations along the dimer interface, among others. When Δ*ireB* is complemented with these dimerization interface mutants, the antimicrobial-resistant phenotype persisted. This displays a strong connection between IreB phosphorylation, dimerization, and its function as a negative regulator of *E. faecalis* resistance. Our results confirm increases in IreB Thr4/Thr7 phosphorylation after ceftriaxone treatment[21,28], while discovering these increases

after chlorhexidine treatment for the first time. This indicates that IreB is involved with defense responses to diverse cell wall perturbations.

Many phosphopeptides with phosphorylation increases after both chlorhexidine and ceftriaxone treatments originated from multiple cell division proteins, supporting the connection between phosphorylation-modulated peptidoglycan synthesis under cell wall stress. These include components of the divisome (GpsB, FtsL, FtsY, and FtsZ), a protein complex located in the cell membrane that modulates cell division and helps coordinate division with septal peptidoglycan synthesis[51]. GpsB (AEA93615.1) had the largest number of phosphopeptides increasing in abundance (seven out of its twelve in total) after both chlorhexidine and ceftriaxone treatments (Table 4.1, Figure 4.3). This phosphorylation increase could be representative of increased IreK activity in preserving the cell wall based on the relationship between IreK homologs and GpsB in other Gram-positive bacteria. During the bacterial cell cycle, GpsB acts as an adaptor, docking peptidoglycan synthases to other cell wall enzymes[52]. In *Bacillus subtilis* GpsB interacts with the IreK homolog PrkC, stimulating its autophosphorylation and kinase activity[53]. GpsB is phosphorylated by PrkC, which inhibits kinase autophosphorylation and activity, regulating PrkC activity through a negative feedback loop[54]. The detected increase in phosphorylation after the antimicrobial treatments suggests that GpsB plays a similar role in *E. faecalis*.

The recombinase RecA (AEA95126.1) phosphosite Ser161 was among those increased in abundance specifically after chlorhexidine treatment (Table 4.1, Figure S3B, Lab_Members\Tony\ThesisSupp). While this RecA phosphopeptide clustered within this group, the chlorhexidine treatment only produced slightly higher abundances of this phosphoeptide when compared to the ceftriaxone treatment. RecA plays an important role in mediating the

bacterial SOS regulatory network where it is activated upon DNA damage or inhibition to cell division, facilitating the inactivation of repressor LexA and the expression of SOS genes[55–57]. These data suggest that there is a chlorhexidine-specific phosphorylation of RecA resulting in its activation, inducing the expression of genes aimed at DNA repair. While chlorhexidine is known to bind and disrupt the bacterial membrane, its degradation product 4-chloroaniline induces DNA damage in bacteria, which could prompt this response[58].

The SsrA-binding protein SmpB (AEA94680.1) phosphosite Thr98 was among those increased in abundance specifically after ceftriaxone treatment (Table 4.1, Figure S3C, Lab_Members\Tony\ThesisSupp). SmpB, along with SsrA, participate in bacterial trans-translation, where incomplete protein products produced under stress conditions are tagged by SsrA and then bound by SmpB to signal these products for degradation[59]. SmpB is required for effective persister formation against antibiotics that inhibit protein and DNA synthesis[59,60]. Previous *E. faecalis* transcriptome studies display a downregulation of ribosomal proteins when cell wall synthesis is inhibited following antimicrobial treatment[61,62]. In the global proteome data analyzed here, proteins related with translation are decreased in abundance specifically after ceftriaxone treatment (which inhibits cell wall synthesis by preventing crosslinking of peptidoglycan outside the cytoplasmic membrane) and this could result in the production of incomplete proteins in the cytoplasm for processing by SsrA/SmpB. These results suggest that this quality control and degradation pathway is controlled through phosphorylation.

4.3.3 Differential Phosphoproteomics Uncovers Potential IreK Substrates

To determine the dependence of stress-mediated phosphorylation on IreK, the phosphoproteomic dataset was expanded to include untreated and treated *E. faecalis* Δ*ireK* null mutant results (Table S3, columns Z-AN, Lab_Members\Tony\ThesisSupp). The 167

phosphosites from 161 unique phosphopeptides and 94 unique proteins described previously were compared across the wild type and Δ*ireK* strains (Table S3, columns K-AN, Lab_Members\Tony\ThesisSupp). Given the protein level abundance differences between these strains (Figure S4, Table S6, Lab_Members\Tony\ThesisSupp), phosphopeptide abundances were normalized (Table S7, columns L-AO, Lab_Members\Tony\ThesisSupp).

A comparison between the untreated strains (wild type vs. Δ*ireK*) revealed 44 phosphopeptides significantly more abundant (FC ≥ 2, FDR-adjusted p-value < 0.05) in the wild type, indicating potential IreK substrates under basal conditions (Figure 4.4A). IreK exhibits some activity during normal growth (in the absence of exogenous antimicrobial stress), and hence these phosphopeptides likely represent IreK substrates that are important for basic growth and cell division. IreK-dependent phosphopeptides increased to 79 and 83 after treatment with chlorhexidine and ceftriaxone, respectively, showing an amplification of possible IreK-mediated phosphorylation events after cell wall stress (Figure 4.4B and 4.4C).

Of the IreK-dependent phosphopeptides significantly increased in the ceftriaxone- and chlorhexidine-treated wild type, 70 phosphopeptides were shared across the treatments, with 9 distinct to chlorhexidine and 13 distinct to ceftriaxone (Figure S5, Table S7, column K, Lab_Members\Tony\ThesisSupp). This high overlap suggests that there is a similar IreK-mediated protein phosphorylation response to cell wall perturbation across distinct antimicrobial mechanisms of action. Previous work suggested that IreK is capable of recognizing activating signals through at least two distinct mechanisms[21], and it remains unclear if ceftriaxone and chlorhexidine trigger IreK activity through distinct or common mechanisms. Regardless of the mechanism by which IreK recognizes stress caused by ceftriaxone or chlorhexidine, the downstream phosphorylation signaling pathways appear to be very similar.

Among the phosphosites determined to be modulated by chlorhexidine and/or ceftriaxone treatments in the wild type, several were increased in abundance in wild type vs. Δ*ireK*, suggesting that these changes are regulated directly or indirectly by IreK. Unsurprisingly, the known IreK substrates IreB Thr4 and Thr7 were increased in the wild type across the untreated, chlorhexidine-, and ceftriaxone-treated samples (Table 4.1, Figure S6, Lab_Members\Tony\ThesisSupp). This trend was the same for SmpB Thr98 (Table 4.1, Figure S6, Lab_Members\Tony\ThesisSupp). GpsB contained the most phosphopeptides increased in abundance in the wild type over Δ*ireK* (10 increased in untreated, 12 increased in chlorhexidine- and ceftriaxone-treated), displaying that it is heavily modified by IreK-mediated phosphorylation (Table 4.1, Figure 4.5A).

To gain a better understanding of the influence of IreK in antimicrobial resistance, IreK-mediated phosphorylation changes for proteins determined to be involved in Gram-positive bacteria stress responses were examined. The histidine kinase CroS (AEA95223.1) phosphosites Ser132 (newly identified here) and Thr346 (previously described) were increased in abundance in the wild type compared to Δ*ireK* following chlorhexidine and ceftriaxone treatments (Table 4.1, Figure S6, Lab_Members\Tony\ThesisSupp). CroS, along with CroR, are an *E. faecalis* TCS that is known to respond to diverse cell wall stresses. TCSs are common in bacteria and allow them to sense and adapt to changing environments, including cell wall stress[31,32]. When CroS detects a cell wall-stress signal, it autophosphorylates at its conserved His172 residue[63]. CroS then binds to the response regulator CroR and transfers the phosphoryl group to Asp52, allowing the phosphorylated CroR to bind to DNA and promote target gene expression[64,65].

The phosphorylation of CroS Thr346 is IreK-dependent through which IreK positively influences CroR-dependent gene expression to promote antimicrobial resistance[30]. Thr346 is

located within a structure of the CroS ATPase domain known as the ATP lid, which is involved in kinase autophosphorylation and makes contact with CroR during the phosphoryl group exchange[66,67]. The replacement of CroS Thr346 with Ala or the putative phosphomimetic Glu leads to the accumulation of phosphorylated CroR and a dramatic enhancement of ceftriaxone resistance, displaying that reversible phosphorylation is critical for the proper regulation of CroS activity[30]. The data presented here reinforces that IreK regulates CroS phosphorylation, which enhances *E. faecalis* TCS signaling towards antimicrobial resistance.

4.3.4 Validation of Direct IreK Substrates Further Implicates this Kinase in *E. faecalis* Cell Wall Homeostasis

To validate IreK-dependent phosphorylation of GpsB, two orthogonal approaches were adopted. First, His-tagged GpsB (GpsB-His_6) was retrieved from cell lysates of exponentially growing *E. faecalis* cells and the enriched GpsB was analyzed by immunoblot using anti-pThr antibody (multiple phosphosites on GpsB were found to be Thr). For this experiment, a Δ*ireP* mutant strain of *E. faecalis* was included as one of the host strains, as IreK is known to be highly active in the absence of its cognate phosphatase IreP regardless of the presence of antimicrobial stress[21,24]. GpsB-His_6 was successfully enriched from all host strains and a robust pThr signal was observed for GpsB enriched from the Δ*ireP* mutant, whereas no signal was found for GpsB enriched from the Δ*ireK* mutant (Figure 4.5B), consistent with IreK-dependent phosphorylation of GpsB in *E. faecalis* cells. To test if IreK can directly phosphorylate GpsB, recombinant GpsB, IreK-n (cytoplasmic kinase fragment of IreK that has been used in previous studies[24,25,28]), and a catalytically impaired IreK-n mutant (IreK-n K41R) were purified for analysis by *in vitro* kinase assays. Purified GpsB was phosphorylated by IreK-n, but not by IreK-n K41R (Figure 4.5C), demonstrating that IreK-n is specifically capable of directly phosphorylating *E. faecalis* GpsB.

The MltG (AEA94902.1) phosphosites Thr20 and Thr75 were increased in abundance in the wild type compared to Δ*ireK* within every condition, while MltG phosphosites Ser49 and Thr77 were only increased in the wild type compared to Δ*ireK* following chlorhexidine and ceftriaxone treatments (Table 4.1, Figure 4.6A). MltG is a transmembrane protein with an extracellular enzymatic domain that functions as a lytic transglycosylase (cleaves peptidoglycan strands) and is thought to terminate peptidoglycan synthesis[68]. In addition to the transmembrane and extracellular domains, the *E. faecalis* MltG homolog also encodes a ~100 amino acid segment located in the cytoplasm. The four potential sites of IreK-dependent phosphorylation in the differential phosphoproteomics dataset are located on this cytoplasmic tail of MltG. To validate these phosphoproteomics results, Phos-tag SDS-PAGE was used to analyze phosphorylation of MltG. Acrylamide-pendant Phos-tag is a dinuclear metal complex that is polymerized directly into polyacrylamide gels and acts as a selective phosphate-binding tag to retard the migration of phosphorylated protein proteoforms. Hence, a snapshot of the *in vivo* phosphorylation state of a protein of interest can be obtained by analyzing whole-cell lysates via immunoblotting after Phos-tag SDS-PAGE, where the least phosphorylated proteoform migrates furthest through the gel and more phosphorylated proteoform(s) migrate more slowly. MltG migration from exponentially growing wild type, Δ*ireK*, and Δ*ireP E. faecalis* cells were analyzed (Figure 4.6B). While the abundance of MltG was similar across all strains as judged by the immunoblot after standard SDS-PAGE, differences in migration were observed after Phos-tag SDS-PAGE. During Phos-tag SDS-PAGE, MltG from wild type cells fractionated as 2 main proteoforms. The more slowly migrating (phosphorylated) proteoform was absent in the Δ*ireK* mutant and was more abundant in the Δ*ireP* mutant (where IreK is highly active). Hence the more slowly migrating (phosphorylated) MltG proteoform in *E. faecalis* was dependent on IreK, and the abundance of

this phosphorylated proteoform correlated with the activity of IreK. Several additional minor bands were observed in the Δ*ireP* mutant. These bands could represent additional proteoforms of MltG resulting from phosphorylation at multiple sites by IreK, but more work is needed to explore this possibility. While multiple MltG phosphosites were identified in the phosphoproteomics data, only singly phosphorylated peptides were detected. Therefore, additional investigation is needed to determine if IreK can phosphorylate MltG at multiple sites of the same protomer. Given the predicted role for MltG as a lytic transglycosylase that participates in regulation of peptidoglycan synthesis, IreK-dependent phosphorylation of the cytoplasmic domain of MltG may serve as a mechanism by which IreK can directly regulate extracellular cell wall synthesis in enterococci.

4.4 Conclusion

In summary, a quantitative phosphoproteomics method was optimized, allowing for robust coverage of the *E. faecalis* phosphoproteome. Using this approach, *E. faecalis* protein phosphorylation was examined to better understand its involvement with resistance in response to cell wall-active antimicrobials. Stress-modulated phosphorylation events between untreated, chlorhexidine-, and ceftriaxone-treated wild type *E. faecalis* were examined, implicating phosphorylation with defense signaling and cell division under both antimicrobial treatments. This displays the conserved phosphorylation of cell division proteins by Gram-positive bacterial kinases to preserve cell wall homeostasis. Some distinct phosphorylation-mediated reactions to antimicrobial treatments were identified, with potential upregulation of the SOS regulatory network in response to chlorhexidine and trans-translation in response to ceftriaxone. To reveal the stress-modulated phosphorylation events caused by IreK, this analysis was extended to incorporate untreated, chlorhexidine-, and ceftriaxone-treated *E. faecalis* Δ*ireK* strains. This

revealed IreK substrates involved with the regulation of peptidoglycan biosynthesis and within the *E. faecalis* two-component system, reinforcing the association between these two pathways in antimicrobial resistance. Future studies will be focused on how the phosphorylation of these proteins mechanistically contributes to IreK-mediated *E. faecalis* virulence and resistance.

4.5 Tables

Table 4.1 Differential phosphoproteomics analysis for phosphopeptides discussed in text. All phosphopeptides are statistically significant (FDR-adjusted *p*-value < 0.05) except those denoted as non-significant (NS).

		Trend analysis	wild type vs. Δ*ireK*, $\log_2$(fold change)		
Protein	Phosphosites from identified peptide	Cluster	Untreated	Chlorhexidine	Ceftriaxone
IreB	T4	A	2.99	5.12	4.39
	T7	A	3.82	5.13	6.47
GpsB	S80	A	3.21	3.43	6.52
	T84	A	5.07	5.36	3.81
	T107	A	8.98	8.49	8.77
	T107, S109	A	4.25	7.21	6.79
	T107, T120, T133	A	2.31 (NS)	6.20	4.17
	T107, T133	A	6.91	6.28	5.46
	S109	NS	8.48	7.34	9.47
	S109, T110	NS	3.98	6.42	4.49
	T110	NS	11.1	10.4	10.5
	T110, T113	NS	5.60	6.30	4.19
	T113	A	4.12 (NS)	4.56	5.82
	T133	C	8.07	6.87	7.06
RecA	S161	B	0.39 (NS)	0.35 (NS)	-0.09 (NS)
SmpB	T98	C	4.74	1.56	8.96
CroS	S132	A	1.05 (NS)	2.02	2.43
	T346	C	5.42	5.79	5.23
MltG	T20	NS	7.29	7.31	8.42
	S49	NS	2.97 (NS)	5.84	4.20
	T75	A	6.00	9.04	5.64
	T77	A	3.72 (NS)	7.98	7.25

4.6 Figures

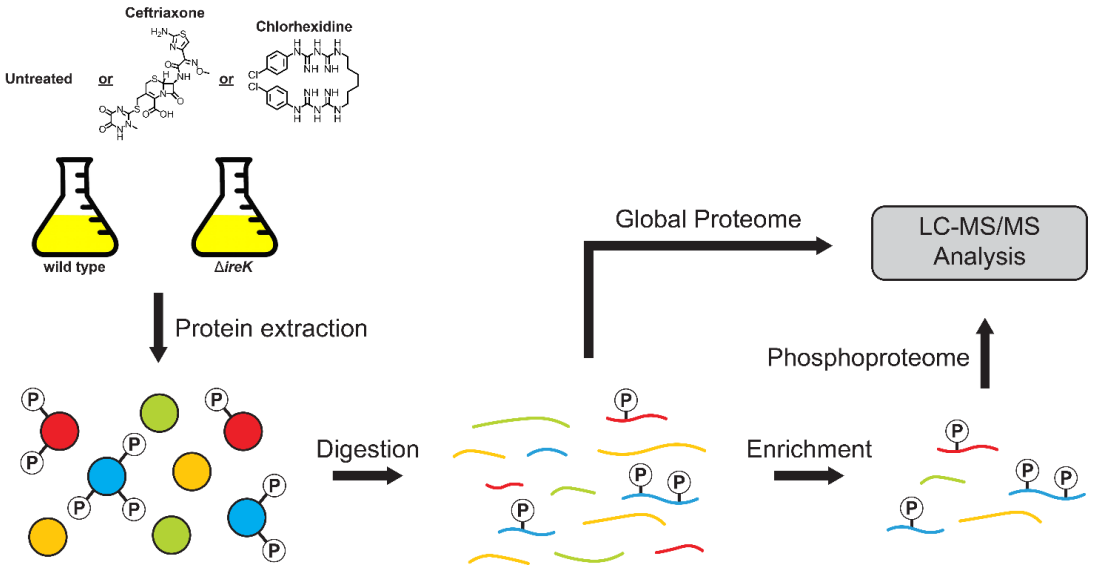


Figure 4.1 General workflow for label-free quantitative phosphoproteomics. Proteins were extracted from untreated, ceftriaxone-, or chlorhexidine-treated *E. faecalis* wild type and *ΔireK* strains. Following extraction, proteins were reduced with DTT, alkylated with IAM, and trypsin digested. Before TiO_2 phosphopeptide enrichment, an aliquot was taken from each sample for global proteome analysis and phosphopeptide abundance normalization. Global proteome and phosphopeptide-enriched samples were analyzing using liquid chromatography-tandem mass spectrometry (LC-MS/MS).

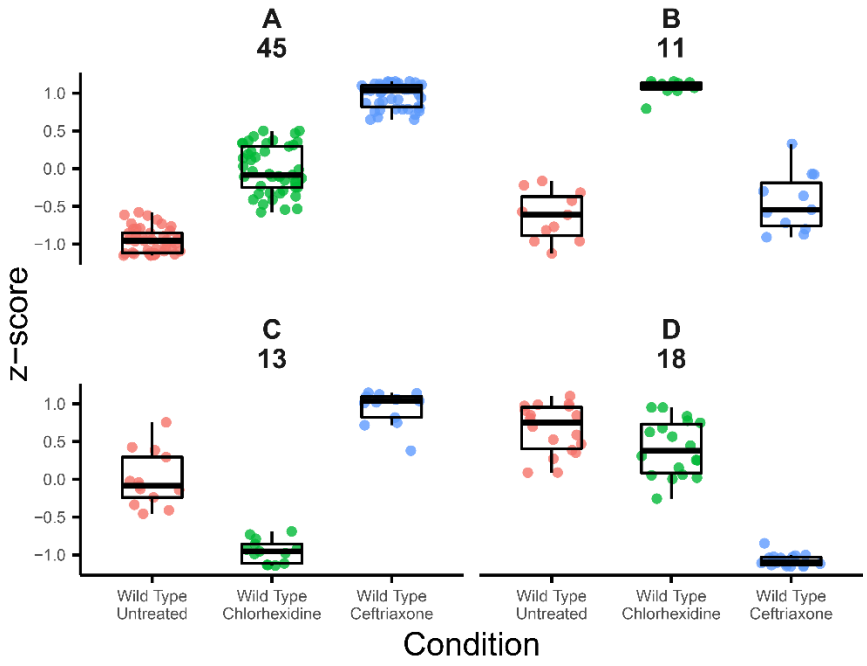


Figure 4.2 Unsupervised hierarchical clustering of the 87 phosphopeptides significantly changing across untreated, chlorhexidine-, and ceftriaxone-treated wild type *E. faecalis* after a one-way ANOVA (FDR-adjusted *p*-value < 0.05), displaying cell wall-active antimicrobial-modulated *E. faecalis* phosphorylation events. Cluster A contains peptides with increased phosphorylation within both the chlorhexidine- and ceftriaxone-treated samples compared to the untreated samples. Cluster B and C contain peptides with increased phosphorylation specifically after chlorhexidine or ceftriaxone treatment, respectively. Cluster D contains peptides with decreased phosphorylation within both the chlorhexidine- and ceftriaxone-treated samples compared to the untreated samples.

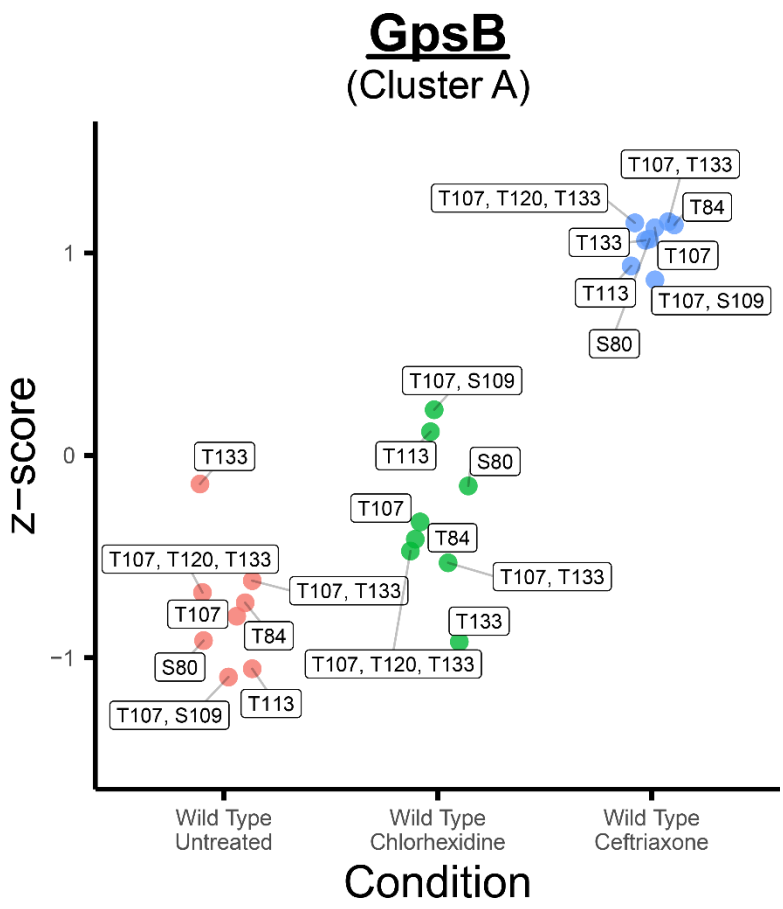


Figure 4.3 The seven GpsB phosphopeptides clustered with the peptides increasing in abundance in the chlorhexidine- and ceftriaxone-treated wild type *E. faecalis* samples compared to the untreated samples after a one-way ANOVA (FDR-adjusted *p*-value < 0.05). The phosphosites within each phosphopeptide are labeled.

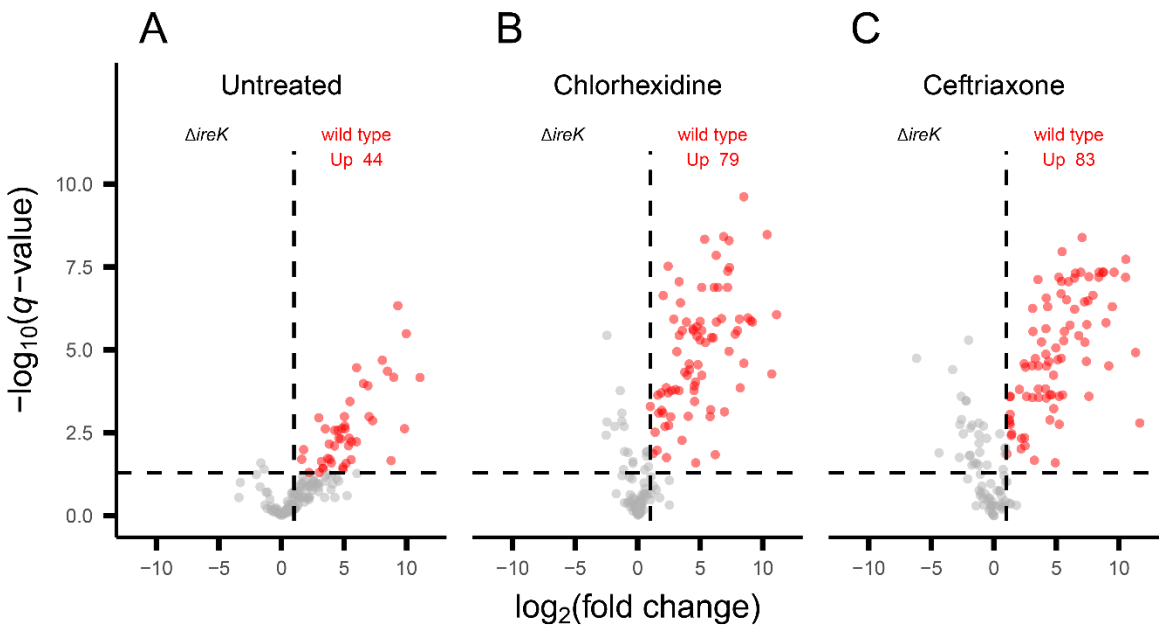


Figure 4.4 Differential analysis of the *E. faecalis* phosphoproteome. This phosphoproteome data was normalized by dividing the phosphopeptide abundances with their protein abundances from the global proteome data in each replicate. Red circles represent significantly changing phosphopeptides (FC ≥ 2, FDR-adjusted *p*-value < 0.05) in the wild type strain after a two-sided, equal variance *t*-test, representing potential IreK phosphorylation substrates. (A) Comparison of untreated strains. (B) Comparison of chlorhexidine-treated strains. (C) Comparison of ceftriaxone-treated strains.

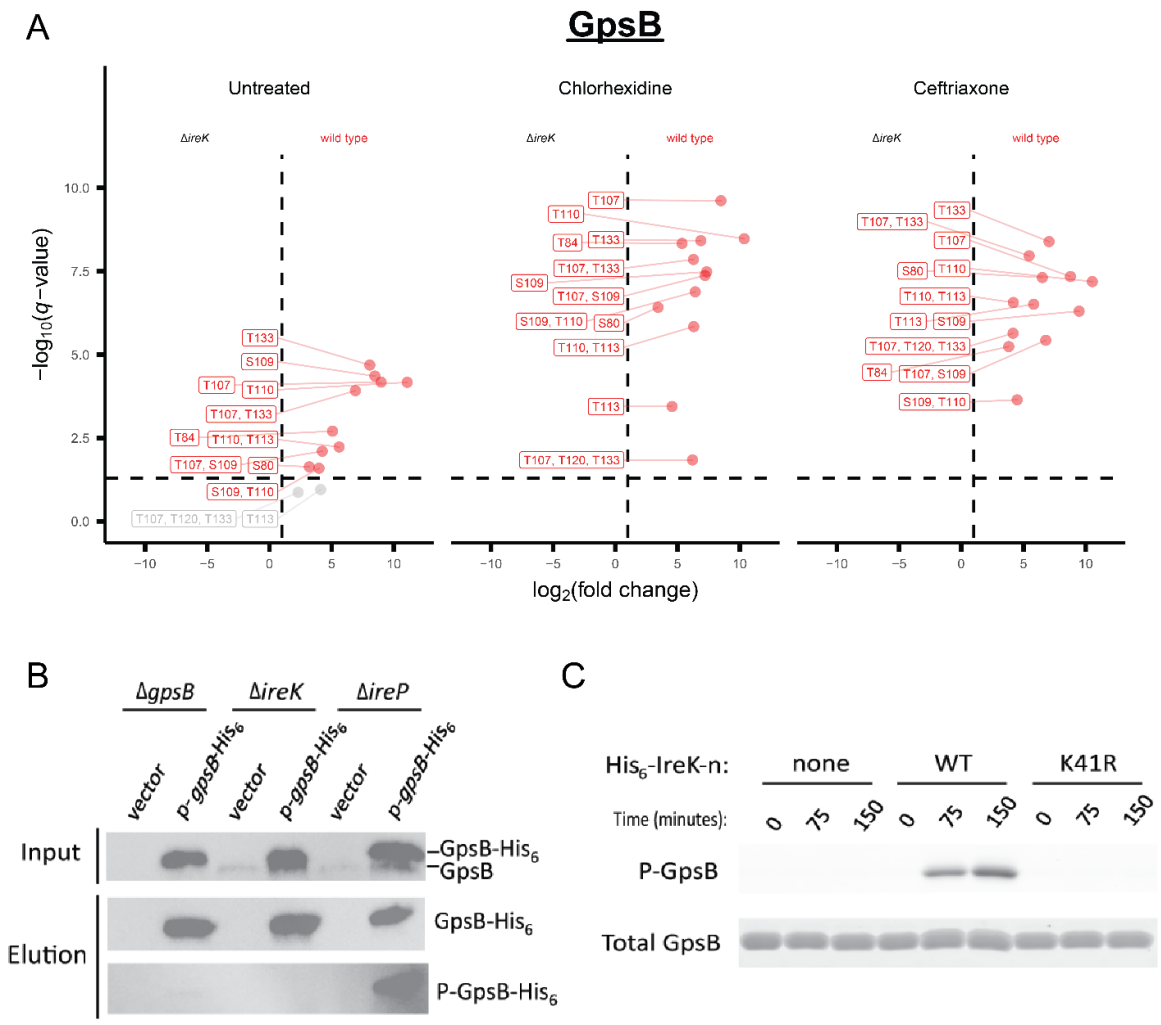


Figure 4.5 IreK-directed phosphorylation of GpsB. (A) Differential analysis of GpsB phosphopeptides. Red circles represent significantly changing GpsB phosphopeptides (FC ≥ 2, FDR-adjusted *p*-value < 0.05) in the wild type strain after a two-sided, equal variance *t*-test, representing potential IreK phosphorylation substrates. The phosphosites within each phosphopeptide are labeled. From left to right: comparison of untreated, chlorhexidine-, and ceftriaxone-treated *E. faecalis* strains. (B) GpsB-His_6 was enriched from lysates of indicated exponentially growing *E. faecalis* cells ("input") using immobilized metal affinity chromatography ("elution"). SDS-PAGE and immunoblotting were performed using anti-GpsB antisera (which detect GpsB and GpsB-His_6) or anti-pThreonine antibody (to detect phosphorylated GpsB, [P-GpsB-His_6]). (C) *In vitro* kinase assays contained purified recombinant GpsB, ATP, and either no kinase (none), wild type His_6-IreK-n catalytic domain (WT), or catalytically impaired His_6-IreK-n K41R mutant (K41R). Data is representative of a minimum of 3 independent replicates.

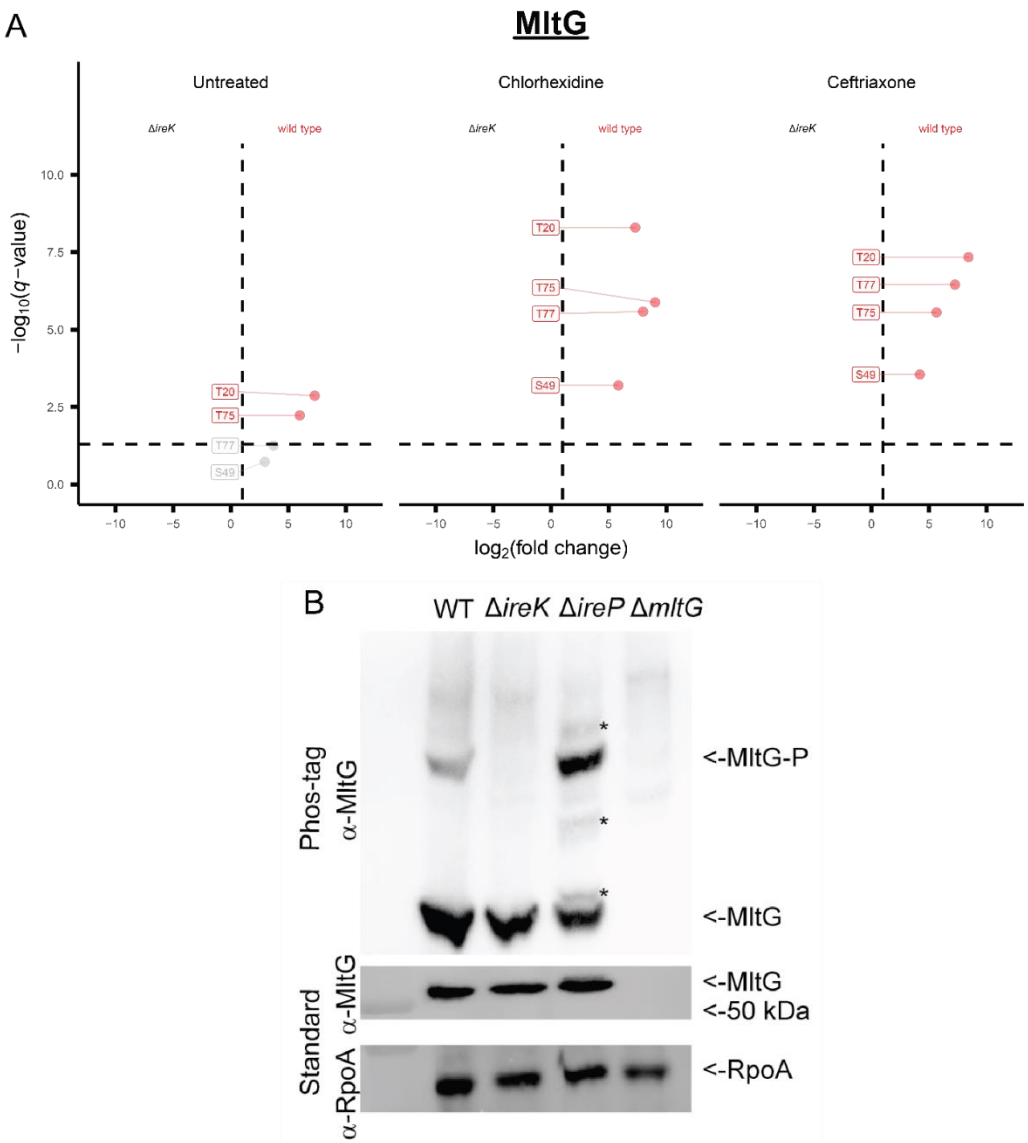


Figure 4.6 Effect of IreK on MltG phosphorylation *in vivo*. (A) Differential analysis of MltG phosphopeptides. Red circles represent significantly changing MltG phosphopeptides (FC ≥ 2, FDR-adjusted *p*-value < 0.05) in the wild type strain after a two-sided, equal variance *t*-test, representing potential IreK phosphorylation substrates. The phosphosites within each phosphopeptide are labeled. From left to right: comparison of untreated, chlorhexidine-, and ceftriaxone-treated *E. faecalis* strains. (B) Immunoblot analysis was performed to analyze phosphorylation of MltG in exponentially growing *E. faecalis* cells. RpoA was used as a loading control. Asterisks indicate potential multiply phosphorylated MltG proteoforms present in the Δ*ireP* mutant but not wild type. Each image is representative of three independent biological replicates.

REFERENCES

1. Gilmore, M. S.; Lebreton, F.; van Schaik, W. Genomic transition of enterococci from gut commensals to leading causes of multidrug-resistant hospital infection in the antibiotic era. *Curr. Opin. Microbiol.* **2013**, *16*(1), 10–16.

2. Hollenbeck, B. L.; Rice, L. B. Intrinsic and acquired resistance mechanisms in *Enterococcus*. *Virulence* **2012**, *3*(5), 421–569.

3. Ubeda, C.; Taur, Y.; Jenq, R. R.; Equinda, M. J.; Son, T.; Samstein, M.; Viale, A.; Socci, N. D.; Van Den Brink, M. R. M.; Kamboj, M.; Pamer, E. G. Vancomycin-resistant *Enterococcus* domination of intestinal microbiota is enabled by antibiotic treatment in mice and precedes bloodstream invasion in humans. *J. Clin. Invest.* **2010**, *120*(12), 4332–4341.

4. Bhardwaj, P.; Ziegler, E.; Palmer, K. L. Chlorhexidine induces VanA-type Vancomycin resistance genes in enterococci. *Antimicrob. Agents Chemother.* **2016**, *60*(4), 2209–2221.

5. Wells, C. L.; Jechorek, R. P.; Erlandsen, S. L. Evidence for the translocation of *Enterococcus faecalis* across the mouse intestinal tract. *J. Infect. Dis.* **1990**, *162*(1), 82–90.

6. Arias, C. A.; Murray, B. E. The rise of the *Enterococcus*: Beyond vancomycin resistance. *Nat. Rev. Microbiol.* **2012**, *10*(4), 266–278.

7. Kristich, C. J.; Wells, C. L.; Dunny, G. M. A eukaryotic-type Ser/Thr kinase in *Enterococcus faecalis* mediates antimicrobial resistance and intestinal persistance. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*(9), 3508–3513.

8. Dahl, A.; Lauridsen, T. K.; Arpi, M.; Sørensen, L. L.; Østergaard, C.; Sogaard, P.; Bruun, N. E. Risk Factors of Endocarditis in Patients with *Enterococcus faecalis* Bacteremia: External Validation of the NOVA Score. *Clin. Infect. Dis.* **2016**, *63*(6), 771–775.

9. Débarbouillé, M.; Dramsi, S.; Dussurget, O.; Nahori, M. A.; Vaganay, E.; Jouvion, G.; Cozzone, A.; Msadek, T.; Duclos, B. Characterization of a serine/threonine kinase involved in virulence of *Staphylococcus aureus*. *J. Bacteriol.* **2009**, *191*(13), 4070–4081.

10. Echenique, J.; Kadioglu, A.; Romao, S.; Andrew, P. W.; Trombe, M. C. Protein Serine/Threonine Kinase StkP Positively Controls Virulence and Competence in *Streptococcus pneumoniae*. *Infect. Immun.* **2004**, *72*(4), 2434–2437.

11. Pensinger, D. A.; Boldon, K. M.; Chen, G. Y.; Vincent, W. J. B.; Sherman, K.; Xiong, M.; Schaenzer, A. J.; Forster, E. R.; Coers, J.; Striker, R.; Sauer, J. D. The *Listeria monocytogenes* PASTA Kinase PrkA and Its Substrate YvcK Are Required for Cell Wall Homeostasis, Metabolism, and Virulence. *PLoS Pathog.* **2016**, *12*(11), e1006001.

12. Banu, L. D.; Conrads, G.; Rehrauer, H.; Hussain, H.; Allan, E.; Van Der Ploeg, J. R. The *Streptococcus mutans* serine/threonine kinase, PknB, regulates competence development,

bacteriocin production, and cell wall metabolism. *Infect. Immun.* **2010**, *78*(5), 2209–2220.

13. Pensinger, D. A.; Schaenzer, A. J.; Sauer, J. D. Do Shoot the Messenger: PASTA Kinases as Virulence Determinants and Antibiotic Targets. *Trends Microbiol.* **2018**, *26*(1), 56–69.

14. Pensinger, D. A.; Aliota, M. T.; Schaenzer, A. J.; Boldon, K. M.; Ansari, I. U. H.; Vincent, W. J. B.; Knight, B.; Reniere, M. L.; Striker, R.; Sauer, J. D. Selective pharmacologic inhibition of a PASTA kinase increases *Listeria monocytogenes* susceptibility to β-lactam antibiotics. *Antimicrob. Agents Chemother.* **2014**, *58*(8), 4486–4494.

15. Beilharz, K.; Nováková, L.; Fadda, D.; Branny, P.; Massidda, O.; Veening, J. W. Control of cell division in *Streptococcus pneumoniae* by the conserved Ser/Thr protein kinase StkP. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*(15), E905–E913.

16. Burnside, K.; Lembo, A.; de los Reyes, M.; Iliuk, A.; BinhTran, N.-T.; Connelly, J. E.; Lin, W.-J.; Schmidt, B. Z.; Richardson, A. R.; Fang, F. C.; Tao, W. A.; Rajagopal, L. Regulation of Hemolysin Expression and Virulence of *Staphylococcus aureus* by a Serine/Threonine Kinase and Phosphatase. *PLoS One* **2010**, *5*(6), e11071.

17. Rajagopal, L.; Clancy, A.; Rubens, C. E. A eukaryotic type serine/threonine kinase and phosphatase in *Streptococcus agalactiae* reversibly phosphorylate an inorganic pyrophosphatase and affect growth, cell segregation, and virulence. *J. Biol. Chem.* **2003**, *278*(16), 14429–14441.

18. Barthe, P.; Mukamolova, G. V.; Roumestand, C.; Cohen-Gonsaud, M. The structure of PknB extracellular PASTA domain from *Mycobacterium tuberculosis* suggests a ligand-dependent kinase activation. *Structure* **2010**, *18*(5), 606–615.

19. Pereira, S. F. F.; Goss, L.; Dworkin, J. Eukaryote-Like Serine/Threonine Kinases and Phosphatases in Bacteria. *Microbiol. Mol. Biol. Rev.* **2011**, *75*(1), 192–212.

20. Lombana, T. N.; Echols, N.; Good, M. C.; Thomsen, N. D.; Ng, H. L.; Greenstein, A. E.; Falick, A. M.; King, D. S.; Alber, T. Allosteric activation mechanism of the *Mycobacterium tuberculosis* receptor Ser/Thr protein Kinase, PknB. *Structure* **2010**, *18*(12), 1667–1677.

21. Labbe, B. D.; Kristich, C. J. Growth- and stress-induced PASTA kinase phosphorylation in *Enterococcus faecalis*. *J. Bacteriol.* **2017**, *199*(21), e00363-17.

22. Boudreau, M. A.; Fishovitz, J.; Llarrull, L. I.; Xiao, Q.; Mobashery, S. Phosphorylation of BlaR1 in Manifestation of Antibiotic Resistance in Methicillin-Resistant *Staphylococcus aureus* and Its Abrogation by Small Molecules. *ACS Infect. Dis.* **2016**, *1*(10), 454–459.

23. Banla, I. L.; Kommineni, S.; Hayward, M.; Rodrigues, M.; Palmer, K. L.; Salzman, N. H.; Kristich, C. J. Modulators of Enterococcus faecalis cell envelope integrity and antimicrobial resistance influence stable colonization of the mammalian gastrointestinal tract. *Infect. Immun.* **2018**, *86*(1), e00381-17.

24. Kristich, C. J.; Little, J. L.; Hall, C. L.; Hoff, J. S. Reciprocal regulation of cephalosporin resistance in *Enterococcus faecalis*. *MBio* **2011**, *2*(6), 199–210.

25. Labbe, B. D.; Hall, C. L.; Kellogg, S. L.; Chen, Y.; Koehn, O.; Pickrum, A. M.; Mirza, S. P.; Kristich, C. J. Reciprocal regulation of PASTA kinase signaling by differential modification. *J. Bacteriol.* **2019**, *209*(10), e00016-19.

26. Hall, C. L.; Lytle, B. L.; Jensen, D.; Hoff, J. S.; Peterson, F. C.; Volkman, B. F.; Kristich, C. J. Structure and Dimerization of IreB, a Negative Regulator of Cephalosporin Resistance in *Enterococcus faecalis*. *J. Mol. Biol.* **2017**, *429*(15), 2324–2336.

27. Wamp, S.; Rutter, Z. J.; Rismondo, J.; Jennings, C. E.; Möller, L.; Lewis, R. J.; Halbedel, S. PrkA controls peptidoglycan biosynthesis through the essential phosphorylation of ReoM. *Elife* **2020**, *9*, e56048.

28. Hall, C. L.; Tschannen, M.; Worthey, E. A.; Kristich, C. J. IreB, a Ser/Thr kinase substrate, influences antimicrobial resistance in *Enterococcus faecalis*. *Antimicrob. Agents Chemother.* **2013**, *57*(12), 6179–6186.

29. Silvestroni, A.; Jewell, K. A.; Lin, W. J.; Connelly, J. E.; Ivancic, M. M.; Tao, W. A.; Rajagopal, L. Identification of serine/threonine kinase substrates in the human pathogen group B *Streptococcus*. *J. Proteome Res.* **2009**, *8*(5), 2563–2574.

30. Kellogg, S. L.; Kristich, C. J. Convergence of PASTA Kinase and Two-Component Signaling in Response to Cell Wall Stress in *Enterococcus faecalis*. *J. Bacteriol.* **2018**, *200*(12), e00086-18.

31. Kellogg, S. L.; Kristich, C. J. Functional dissection of the CroRS two-component system required for resistance to cell wall stressors in *Enterococcus faecalis*. *J. Bacteriol.* **2016**, *198*(8), 1326–1336.

32. Kellogg, S. L.; Little, J. L.; Hoff, J. S.; Kristich, C. J. Requirement of the CroRS two-component system for resistance to cell wall-targeting antimicrobials in Enterococcus faecium. *Antimicrob. Agents Chemother.* **2017**, *61*(5), e02461-16.

33. Kristich, C. J.; Chandler, J. R.; Dunny, G. M. Development of a host-genotype-independent counterselectable marker and a high-frequency conjugative delivery system and their use in genetic analysis of *Enterococcus faecalis*. *Plasmid* **2007**, *57*(2), 131–144.

34. Werth, E. G.; McConnell, E. W.; Gilbert, T. S. K.; Couso Lianez, I.; Perez, C. A.; Manley, C. K.; Graves, L. M.; Umen, J. G.; Hicks, L. M. Probing the global kinome and phosphoproteome in *Chlamydomonas reinhardtii* via sequential enrichment and quantitative proteomics. *Plant J.* **2017**, *89*(2), 416–426.

35. Käll, L.; Canterbury, J. D.; Weston, J.; Noble, W. S.; MacCoss, M. J. Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat. Methods* **2007**, *4*(11), 923–925.

36. Gianetto, Q. G. imp4p: Imputation for Proteomics https://cran.r-project.org/web/packages/imp4p/index.html (accessed Aug 6, 2019).

37. Wu, R.; Dephoure, N.; Haas, W.; Huttlin, E. L.; Zhai, B.; Sowa, M. E.; Gygi, S. P. Correct Interpretation of Comprehensive Phosphorylation Dynamics Requires Normalization by Protein Expression Changes. *Mol. Cell. Proteomics* **2011**, *10*(8), M111.009654.

38. Krahmer, N.; Najafi, B.; Schueder, F.; Quagliarini, F.; Steger, M.; Seitz, S.; Kasper, R.; Salinas, F.; Cox, J.; Uhlenhaut, N. H.; Walther, T. C.; Jungmann, R.; Zeigerer, A.; Borner, G. H. H.; Mann, M. Organellar Proteomics and Phospho-Proteomics Reveal Subcellular Reorganization in Diet-Induced Hepatic Steatosis. *Dev. Cell* **2018**, *47*(2), 205-221.e7.

39. Wang, Z.; Ma, J.; Miyoshi, C.; Li, Y.; Sato, M.; Ogawa, Y.; Lou, T.; Ma, C.; Gao, X.; Lee, C.; Fujiyama, T.; Yang, X.; Zhou, S.; Hotta-Hirashima, N.; Klewe-Nebenius, D.; Ikkyu, A.; Kakizaki, M.; Kanno, S.; Cao, L.; Takahashi, S.; Peng, J.; Yu, Y.; Funato, H.; Yanagisawa, M.; Liu, Q. Quantitative phosphoproteomic analysis of the molecular substrates of sleep need. *Nature* **2018**, *558*(7710), 435–439.

40. Zhang, H.; Zhou, H.; Berke, L.; Heck, A. J. R.; Mohammed, S.; Scheres, B.; Menke, F. L. H. Quantitative Phosphoproteomics after Auxin-stimulated Lateral Root Induction Identifies an SNX1 Protein Phosphorylation Site Required for Growth. *Mol. Cell. Proteomics* **2013**, *12*(5), 1158–1169.

41. Savitski, M. M.; Lemeer, S.; Boesche, M.; Lang, M.; Mathieson, T.; Bantscheff, M.; Kuster, B. Confident phosphorylation site localization using the mascot delta score. *Mol. Cell. Proteomics* **2011**, *10*(2), S1–S12.

42. Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B* **1995**, *57*(1), 289–300.

43. Perez-Riverol, Y.; Csordas, A.; Bai, J.; Bernal-Llinares, M.; Hewapathirana, S.; Kundu, D. J.; Inuganti, A.; Griss, J.; Mayer, G.; Eisenacher, M.; Erez, E. P.; Uszkoreit, J.; Pfeuffer, J.; Sachsenberg, T.; Yılmaz, S.; Tiwary, S.; Cox, J.; Audain, E.; Walzer, M.; Jarnuczak, A. F.; Ternent, T.; Brazma, A.; Vizcaíno, J. A. The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* **2019**, *47*(D1), D442–D450.

44. Lin, M. H.; Sugiyama, N.; Ishihama, Y. Systematic profiling of the bacterial phosphoproteome reveals bacterium-specific features of phosphorylation. *Sci. Signal.* **2015**, *8*(394), rs10.

45. Faca, V. M.; Sanford, E. J.; Tieu, J.; Comstock, W.; Gupta, S.; Marshall, S.; Yu, H.; Smolka, M. B. Maximized quantitative phosphoproteomics allows high confidence dissection of the DNA damage signaling network. *Sci. Rep.* **2020**, *10*(1), 18056.

46. Sharma, K.; D'Souza, R. C. J.; Tyanova, S.; Schaab, C.; Wiśniewski, J. R.; Cox, J.; Mann, M. Ultradeep Human Phosphoproteome Reveals a Distinct Regulatory Nature of Tyr and

Ser/Thr-Based Signaling. *Cell Rep.* **2014**, *8*(5), 1583–1594.

47. Henry, C.; Haller, L.; Blein-Nicolas, M.; Zivy, M.; Canette, A.; Verbrugghe, M.; Mézange, C.; Boulay, M.; Gardan, R.; Samson, S.; Martin, V.; André-Leroux, G.; Monnet, V. Identification of Hanks-Type Kinase PknB-Specific Targets in the *Streptococcus thermophilus* Phosphoproteome. *Front. Microbiol.* **2019**, *10*, 1329.

48. Rioseras, B.; Shliaha, P. V.; Gorshkov, V.; Yagüe, P.; López-García, M. T.; Gonzalez-Quiñonez, N.; Kovalchuk, S.; Rogowska-Wrzesinska, A.; Jensen, O. N.; Manteca, A. Quantitative proteome and phosphoproteome analyses of *Streptomyces coelicolor* reveal proteins and phosphoproteins modulating differentiation and secondary metabolism. *Mol. Cell. Proteomics* **2018**, *17*(8), 1591–1611.

49. Rosenberg, A.; Soufi, B.; Ravikumar, V.; Soares, N. C.; Krug, K.; Smith, Y.; Macek, B.; Ben-Yehuda, S. Phosphoproteome dynamics mediate revival of bacterial spores. *BMC Biol.* **2015**, *13*(1), 1–19.

50. Prust, N.; Laarse, S. van der; Toorn, H. W. P. van den; Sorge, N. M. van; Lemeer, S. In-Depth Characterization of the *Staphylococcus aureus* Phosphoproteome Reveals New Targets of Stk1. *Mol. Cell. Proteomics* **2021**, *20*, 100034.

51. Halbedel, S.; Lewis, R. J. Structural basis for interaction of DivIVA/GpsB proteins with their ligands. *Mol. Microbiol.* **2019**, *111*(6), 1404–1415.

52. Cleverley, R. M.; Rutter, Z. J.; Rismondo, J.; Corona, F.; Tsui, H.-C. T.; Alatawi, F. A.; Daniel, R. A.; Halbedel, S.; Massidda, O.; Winkler, M. E.; Lewis, R. J. The cell cycle regulator GpsB functions as cytosolic adaptor for multiple cell wall enzymes. *Nat. Commun.* **2019**, *10*(1), 261.

53. Pompeo, F.; Byrne, D.; Mengin-Lecreulx, D.; Galinier, A. Dual regulation of activity and intracellular localization of the PASTA kinase PrkC during *Bacillus subtilis* growth. *Sci. Rep.* **2018**, *8*(1), 1660.

54. Pompeo, F.; Foulquier, E.; Serrano, B.; Grangeasse, C.; Galinier, A. Phosphorylation of the cell division protein GpsB regulates PrkC kinase activity through a negative feedback loop in *Bacillus subtilis*. *Mol. Microbiol.* **2015**, *97*(1), 139–150.

55. Erill, I.; Campoy, S.; Barbé, J. Aeons of distress: An evolutionary perspective on the bacterial SOS response. *FEMS Microbiol. Rev.* **2007**, *31*(6), 637–656.

56. Martin-Verstraete, I.; Peltier, J.; Dupuy, B. The Regulatory Networks That Control *Clostridium difficile* Toxin Synthesis. *Toxins (Basel).* **2016**, *8*(5), 153.

57. Thompson, S. A.; Blaser, M. J. Isolation of the *Helicobacter pylori* recA Gene and Involvement of the recA Region in Resistance to Low pH. *Infect. Immun.* **1995**, *63*(6), 2185–2193.

58. Barbin, L. E.; Saquy, P. C.; Guedes, D. F. C.; Sousa-Neto, M. D.; Estrela, C.; Pécora, J. D.

Determination of para-Chloroaniline and Reactive Oxygen Species in Chlorhexidine and Chlorhexidine Associated with Calcium Hydroxide. *J. Endod.* **2008**, *34*(12), 1508–1514.

59. Li, J.; Ji, L.; Shi, W.; Xie, J.; Zhang, Y. Trans-translation mediates tolerance to multiple antibiotics and stresses in *Escherichia coli*. *J. Antimicrob. Chemother.* **2013**, *68*(11), 2477–2481.

60. Amato, S. M.; Brynildsen, M. P. Persister heterogeneity arising from a single metabolic stress. *Curr. Biol.* **2015**, *25*(16), 2090–2098.

61. Darnell, R. L.; Knottenbelt, M. K.; Todd Rose, F. O.; Monk, I. R.; Stinear, T. P.; Cook, G. M. Genomewide Profiling of the *Enterococcus faecalis* Transcriptional Response to Teixobactin Reveals CroRS as an Essential Regulator of Antimicrobial Tolerance. *mSphere* **2019**, *4*(3), e00228-19.

62. Abranches, J.; Tijerina, P.; Avilés-Reyes, A.; Gaca, A. O.; Kajfasz, J. K.; Lemos, J. A. The Cell Wall-Targeting Antibiotic Stimulon of *Enterococcus faecalis*. *PLoS One* **2013**, *8*(6), e64875.

63. Bhate, M. A. P.; Molnar, K. A. S.; Goulian, M.; Degrado, W. F. Signal Transduction in Histidine Kinases: Insights from New Structures. *Structure* **2015**, *23*(6), 981–994.

64. Gao, R.; Stock, A. M. Biological insights from structures of two-component proteins. *Annu. Rev. Microbiol.* **2009**, *63*, 133–154.

65. Comenge, Y.; Quintiliani, R.; Li, L.; Dubost, L.; Brouard, J. P.; Hugonnet, J. E.; Arthur, M. The CroRS Two-Component Regulatory System Is Required for Intrinsic β-Lactam Resistance in *Enterococcus faecalis*. *J. Bacteriol.* **2003**, *185*(24), 7184–7192.

66. Stewart, R. C. Protein histidine kinases: assembly of active sites and their regulation in signaling pathways. *Curr. Opin. Microbiol.* **2010**, *13*(2), 133–141.

67. Casino, P.; Rubio, V.; Marina, A. Structural Insight into Partner Specificity and Phosphoryl Transfer in Two-Component Signal Transduction. *Cell* **2009**, *139*(2), 325–336.

68. Yunck, R.; Cho, H.; Bernhardt, T. G. Identification of MltG as a potential terminase for peptidoglycan polymerization in bacteria. *Mol. Microbiol.* **2016**, *99*(4), 700–718.

CHAPTER 5: Standard Operating Procedures for Crosslinking- and Immunoprecipitation-Mass Spectrometry

5.1 Introduction

Novel protein-protein interactions can be identified by crosslinking-mass spectrometry (XL-MS) and/or immunoprecipitation-mass spectrometry (IP-MS). Proteome-wide XL-MS studies are employed to elucidate the global protein interactome, while IP-MS experiments lead to the identification of protein-protein interactions for a specific protein of interest. Both approaches share overlapping processing steps: organism growth and harvest, native protein extraction, proteolytic digestion, clean up, and LC-MS/MS analysis, though order of operation can differ based on the experiment. This chapter focuses on the standard operating procedures optimized and utilized in XL- and IP-MS experiments to discover protein-protein interactions in the model alga species *Chlamydomonas reinhardtii*. Some of these procedures are implemented in Chapter 6, where XL-MS was performed to reveal the Chlamydomonas global protein interactome.

5.2 Algal Culturing

Strains: The Chlamydomonas Resource Center (https://www.chlamycollection.org/) at the University of Minnesota, Twin Cities maintains thousands of Chlamydomonas research strains that are available for purchase. The cell-walled Chlamydomonas strain CC-2895 is the strain of choice for experiments in the Hicks Laboratory and typically serves as the wild type strain[1–5]. For IP-MS experiments involving lethal with SEC13 protein 8 (LST8), the Hicks Laboratory received a Chlamydomonas CC-4533 *lst8-1* strain complemented with LST8 tagged with the

Escherichia coli OmpF Linker and mouse Langerin fusion sequence (OLLAS) on its C-terminus[6]. The parent strain CC-4533 was used as the wild type strain and negative control.

All culturing steps are performed in Kenan B328 either in the growth chamber or on the bench-top shaker. The growth chamber provides more consistent light and temperature, but space is limited to twelve 250 mL flasks. The bench-top shaker can support more cultures if an experiment requires additional replicates or larger culture sizes (as is the case for the immunoprecipitation of LST8-OLLAS), but this area is exposed to inconsistent overhead lighting and temperature fluctuations. Both areas are continuously illuminated by cool, white fluorescent bulbs that provide photosynthetically active radiation (PAR) between 400-700 nm at 50-100 µmol m^{-2} s^{-1} as measured by a PAR meter (Sun Systems) and constantly mixed on a platform shaker at 120 rpm.

Chlamydomonas strains in the Hicks Laboratory are maintained on Tris-acetate-phosphate (TAP)-agar plates (https://www.chlamycollection.org/methods/media-recipes/tap-and-tris-minimal/), located in the growth chamber. To ensure reproducible growth between biological replicates, all experimental cultures are grown simultaneously and inoculated from a primary culture that is prepared in advance. Liquid cultures are mixotrophically grown with acetate as a carbon source provided in the TAP media along with constant light exposure to maximize growth rate[7]. The primary culture is prepared by scraping a single colony of Chlamydomonas cells using an inoculating loop from its plate and transferring to a 100 mL liquid TAP culture. After inoculation with the primary culture, experimental cultures are grown to mid-log phase (OD_{750} ~0.4 AU) before harvesting. Each culture represents a biological replicate in subsequent experiments.

5.2.1 Typical Algae Cultures

Day 1:

1. Measure 100 mL of liquid TAP medium with a graduated cylinder and pour into an autoclaved 250 mL Erlenmeyer flask with an aluminum foil lid.
2. Inoculate the cultures with 1 mL of the primary culture (grown to an OD_{750} of 0.4 – 0.5 AU) and label flasks with a small piece of tape with identifying details (*e.g.*, strain, date, and initials).
3. Secure cultures on the shaker platform in the growth chamber and shake at 120 rpm under constant light. Ensure constant rotation and light conditions throughout the growth period.

Day 4:

4. After three days of growth, check that cultures have reached an OD_{750} 0.4-0.5 AU, corresponding approximately to mid-log phase growth.
5. Harvest cultures by pouring into 50 mL centrifuge tubes and centrifuging for 5 min at 3,220 x *g* and 4 °C.
6. Decant the supernatant and invert centrifuge tubes on a paper towel to dry (~30 s) before replacing the cap and weighing the cell pellets.
7. Flash-freeze samples in liquid nitrogen and store pellets at -80 °C until ready for extraction.

5.2.2 Large Algae Cultures

Day 1:

1. Measure 450 mL of liquid TAP medium with a graduated cylinder and pour into an autoclaved 1 L Erlenmeyer flask with an aluminum foil lid.
2. Inoculate the cultures with 4.5 mL of the primary culture and label flasks with a small piece of tape with identifying details (*e.g.*, strain, date, and initials).
3. Secure cultures on the shaker platform on the bench-top and shake at 120 rpm. Ensure constant rotation and conditions throughout the growth period.

Day 4:

4. After three days of growth, check that cultures have reached an OD_{750} 0.4-0.5 AU, corresponding approximately to mid-log phase growth.
5. Harvest cultures by pouring into 1 L centrifuge bottles and centrifuging for 5 min at 2,300 *x* g and 4 °C before decanting the supernatant.
6. Resuspend the cell pellets in 5 mL of fresh TAP, transfer to pre-weighted 15 mL centrifuge tubes, and centrifuge again for 5 min at 3,220 x *g* and 4 °C.
7. Decant the supernatant and invert centrifuge tubes on a paper towel to dry (~30 s) before replacing the cap and weighing the cell pellets.
8. Flash-freeze samples in liquid nitrogen and store pellets at -80 °C until ready for extraction.

5.3 Non-Denaturing Protein Extraction

Preserving native protein structures and protein-protein interactions before proteins are covalently bound during crosslinking is vital to both XL- and IP-MS analyses. Ionic detergents (*e.g.*, sodium dodecyl sulfate) are commonly used in protein extraction buffers, but these detergents denature proteins and disrupt protein-protein interactions so they cannot be used for

XL- and IP-MS experiments. Instead, alternative protein extraction methods that preserve protein-protein interactions while maintaining optimal extraction efficiencies are required. For the XL-MS experiments described here, detergent-free lysis via freeze-thaw cycles is used to lyse cells and release proteins in their native conformations. For the IP-MS experiments described here, non-ionic detergents (*e.g.*, Triton and Tween series) are included in protein extraction buffers. These detergents disturb lipid-lipid and protein-lipid interactions, but preserve native protein structures and protein-protein interactions. For experiments that include amine reactive crosslinkers, it is vital that Tris or other primary amine-containing chemicals are omitted from the extraction buffer. These reagents quench the crosslinking reaction and prevent protein crosslinking.

5.3.1 Freeze-Thaw Gentle Lysis

Procedure:

1. Remove cell pellets from the -80 °C freezer and thaw on ice for 30 min.
2. Resuspend cell pellets in 10 mL lysis buffer for every 0.6 g of wet cell pellet weight and mix gently with a 5 mL pipette.
    a. Lysis buffer is 0.5 mM dithiothreitol (DTT), 30 mM NaCl, 1.5 mM $MgCl_2$ and 1x cOmplete, EDTA-free protease inhibitor cocktail (Roche) in 20 mM HEPES, pH 7.8.
3. Lyse cells via five rounds of freeze-thaw at -80 °C for 1 h. Each time, thaw cells on ice for 30 min.
4. Clarify lysates by centrifuging for 20 min at 3,200 x *g* and 4 °C.
5. Transfer the clarified lysates into 50 mL centrifuge tubes and discard the cell pellet.

6. Estimate the protein concentrations using the CB-X protein assay (G-Biosciences) following the manufacturer's protocols.
7. Aliquot 3 mg protein for each replicate.
8. Store samples at 4 °C until ready for disuccinimidyl sulfoxide (DSSO) crosslinking.

5.3.2 Non-Ionic Detergent

Procedure:

1. Remove cell pellets from the -80 °C freezer and let them thaw at room temperature (RT) for 5 min.
2. Resuspend cell pellets in 5 mL lysis buffer for every 1 g of wet cell pellet weight and mix with a 5 mL pipette.
    a. Lysis buffer is 0.5% Triton X-100, 1 mM EDTA, 150 mM NaCl, and 1x cOmplete, EDTA-free protease inhibitor cocktail (Roche) in 50 mM HEPES, pH 7.5.
3. Incubate samples with end-over-end rotation for 1.5 h at 4 °C.
4. Transfer the samples to 2 mL Eppendorf tubes and clarify lysates by centrifuging for 10 min at 16,000 x *g* and 4 °C.
5. Transfer the clarified lysates into 50 mL centrifuge tubes.
6. Estimate the protein concentrations using the CB-X protein assay (G-Biosciences) following the manufacturer's protocols.
7. Aliquot necessary protein amount for robust immunoprecipitation of protein of interest (based on SDS-PAGE and western blot results from optimization experiments, Section 5.6) for each replicate.

a. For the immunoprecipitation of LST8-OLLAS, the optimal amount was 35 mg protein.

8. Store samples at 4 °C until ready for dithiobis(succinimidyl propionate) (DSP) crosslinking.

5.4 Chemical Crosslinking

Chemical crosslinking is implemented to stabilize protein complexes through the covalent linkage of two protein residues that are in close proximity. Crosslinking *in vivo* has proven to be difficult as crosslinker permeability into algae cells is hampered by the cell wall, so protein-protein interactions preserved by native protein extractions can be crosslinked *in vitro*. Crosslinker reagents vary in reactive groups (*e.g.*, NHS ester for amine reactivity, maleimide for cysteine reactivity), spacer arm length, and additional features (*e.g.*, MS-cleavable, reversible, enrichable). There are many commercially available crosslinkers and choice of crosslinker should be tailored to the experiment and analysis being performed. Amine-reactive crosslinkers are commonly used due to the prevalence of lysine residues in proteins that offer more positions for potential crosslinking. Crosslinkers with longer spacer arms can allow for a more comprehensive analysis of protein structures and the covalent attachment of transient protein-protein interactions, but their use can introduce more false positives, especially when crosslinking *in vitro*[8].

5.4.1 DSSO Crosslinking

DSSO is a homo-bifunctional, amine-reactive, MS-cleavable crosslinker that can covalently bind proteins at lysine residue side chains and N-termini[9]. It is water-insoluble, so DSSO must first be solubilized in an organic solvent such as dimethyl sulfoxide (DMSO) or dimethylformamide (DMF) before adding to a native protein lysate. During LC-MS/MS analysis,

selection and fragmentation of a peptide pair crosslinked with DSSO induces cleavage of the spacer arm. This generates characteristic crosslink reporter ion doublets (*i.e.,* observed mass shifts from the short or long end of the cleaved crosslinker) that facilitate the identification of crosslinked peptides. The cleavage of DSSO also permits more efficient peptide fragmentation and less complex MS/MS spectra of increased quality, which simplifies crosslink identification in database searching and lends confidence to site assignment[10]. If a mass spectrometer has the capabilities, the characteristic crosslinker reporter ions can be selected for MS^3 fragmentation to analyze each peptide of the crosslinked pair individually, which decreases fragment spectra complexity and improves confidence in crosslinked peptide identification.

Procedure:

1. Remove samples from 4 °C and warm to RT on benchtop.
    a. Adding concentrated solution of DSSO in DMSO to cold protein lysate can result in DSSO precipitation.
2. Create stock solution of 50 mM DSSO in DMSO.
3. Slowly add DSSO stock solution to obtain final concentration of 2 mM DSSO and 1% DMSO.
    a. Quickly adding concentrated solution of DSSO in DMSO to aqueous protein lysate can result in DSSO precipitation.
    b. High concentrations of DMSO can denature proteins and disrupt protein-protein interactions.
    c. The optimal crosslinker concentration is determined by aliquoting the protein lysate and crosslinking these aliquots with a range of reagent concentrations. Results are visualized by SDS-PAGE (Section 5.6.1) and Coomassie staining. The

optimal concentration is the lowest concentration that provides a mixture of crosslinked, higher molecular weight and non-crosslinked, lower molecular weight species. When low concentrations of crosslinker are added, there will be a lack of proteins in the high molecular weight region representing low crosslinking efficiency. With high crosslinker concentrations, there will be a lack of signal intensity in the lower molecular weight regions. Introducing high crosslinker concentrations are problematic as the crosslinker will block all protein lysines, leading to incomplete digestion with trypsin, and introduce false positive protein-protein interactions when performing crosslinking *in vitro*.

4. Incubate samples with end-over-end rotation for 30 min at RT.
5. Add 1 M Tris, pH 8 to a final concentration of 20 mM to quench crosslinking.
6. Incubate samples with end-over-end rotation for 30 min at RT.
7. Proceed with protein reduction, alkylation, and digestion.

5.4.2 DSP Crosslinking

DSP is a homo-bifunctional, amine-reactive, thiol-cleavable crosslinker that can reversibly crosslink proteins at lysine residue side chains and N-termini[11]. Since it is water-insoluble, DSP must first be solubilized in an organic solvent (DMSO or DMF) before adding to a native protein lysate. The disulfide bond within the DSP spacer arm can be cleaved by DTT or tris(2-carboxyethyl)phosphine (TCEP), which allows for the separation of crosslinked proteins prior to digestion. The reversible nature of DSP produces non-crosslinked tryptic peptides that are available for selection and fragmentation in data-dependent acquisition LC-MS/MS methods. This permits the use of the typical Hicks Laboratory data analysis pipeline (Progenesis/Mascot/*QuantifyR*), as this prevents the need for database searching software with

crosslinked peptide database searching capabilities. The only difference is that protocols using DSP require the inclusion of the fixed modification “CAMthiopropanoyl (K)” to account for the reduced DSP crosslinker modified with iodoacetamide (IAM). This crosslinker should only be used in experiments that include IP and therefore, cannot be used to profile the global protein interactome. For global protein interactome profiling, selection and fragmentation of a peptide pair with the crosslinker intact is required to differentiate between intralinks and interlinks, and understand the two proteins involved in the protein-protein interaction for detected interlinks. On the other hand, IP enriches the protein of interest and its interacting proteins so peptides identified with the “CAMthiopropanoyl (K)” fixed modification are theoretically derived from a crosslink between the protein of interest and one of its interacting proteins.

Procedure:

1. Remove samples from 4 °C and let them warm to RT on benchtop.
    a. Adding concentrated solution of DSP in DMSO to cold protein lysate can result in DSP precipitation.
2. Create stock solution of 100 mM DSP in DMSO.
3. Slowly add DSP stock solution to obtain final concentration of 1 mM DSP and 1% DMSO.
    a. Quickly adding concentrated solution of DSP in DMSO to aqueous protein lysate can result in DSP precipitation.
    b. High concentrations of DMSO can denature proteins and disrupt protein-protein interactions.
4. Incubate samples with end-over-end rotation for 1 h at RT.
5. Add 1 M Tris, pH 7.5 to a final concentration of 20 mM to quench crosslinking.

6. Incubate samples with end-over-end rotation for 15 min at RT.
7. Proceed with IP.

5.5 Immunoprecipitation

A protein of interest and its DSP-crosslinked, interacting proteins can be enriched through the affinity purification technique IP. To pulldown a target protein, epitope tags with commercially available antibodies, such as hemagglutinin antigen (HA), FLAG, polyHis, and OLLAS, can be fused to the protein of interest using recombinant DNA methods[12]. For the IP-MS experiment described here, the OLLAS epitope tag is chosen for the IP of LST8 due to its higher sensitivity compared to other conventional tags in western blot and IP analyses[13]. Besides the epitope-tagged protein and antibody, an IP support (*e.g.*, Protein A/G agarose) is necessary to immunoprecipitate the protein-antibody construct from solution by binding to the antibody Fc (constant) region. Selection of Protein A or Protein G depends on the host species of the antibody, as Protein A has a stronger binding affinity for rabbit and pig antibodies while Protein G has a stronger binding affinity to mouse and human antibodies[14].

There are multiple considerations when developing and implementing an IP method. All steps should be performed in Eppendorf tubes or 15 mL falcon tubes if increased protein loading amounts are needed. The IP support does not settle at the bottom of 50 mL falcon tubes following centrifugation, which will lead to sample loss. The IP support is typically shipped in a storage buffer, so this resin should be equilibrated in the IP wash buffer before adding it to samples. Non-specific binding proteins that bind to the IP support increase the potential for false positives, thus samples should be incubated (*i.e.*, precleared) with the resin before the addition of the antibody to remove these non-specific binding proteins. During wash steps, stringent wash buffers can be used to further eliminate potential false positives without the fear of losing true

positive, interacting proteins since crosslinking is performed before the IP. There are many choices for the IP elution buffer including acidic glycine or SDS-PAGE buffers, but a urea-based elution buffer is chosen here due to its compatibility with downstream bottom-up proteomics methods.

Day 1:

1. Add 20 µL of Protein A/G agarose resin (Thermo Fisher Scientific) to preclear the samples.
2. Incubate samples with end-over-end rotation for 1 h at 4 °C.
3. Clarify lysates by centrifuging for 1 min at 2,000 x *g* and 4 °C, transferring supernatant to new centrifuge tubes, and discarding resin.
4. Add protein tag-specific antibody to precleared lysates.
    a. The amount of antibody needed for IP should be listed in the manufacturer's product details.
5. Incubate samples with end-over-end rotation overnight (> 16 h) at 4 °C.

Day 2:

6. Add Protein A/G agarose resin to the samples.
    a. The amount of Protein A/G agarose resin required is dependent on the resin binding capacity and should be listed in the manufacturer's product details.
7. Incubate samples with end-over-end rotation for 3 h at 4 °C.
8. Collect the resin and immunoprecipitated complex by centrifuging for 1 min at 2,000 x *g* and 4 °C and discard the supernatant.
    a. Remove as much supernatant as possible without disrupting the resin.

9. Wash the resin with 1 mL of wash buffer, collect the resin and immunoprecipitated complex by centrifuging for 1 min at 2,000 x *g* and 4 °C, and discard the supernatant.
    a. Wash buffer is 50 mM NaCl, 10 mM Tris, pH 7.5
    b. Remove as much supernatant as possible without disrupting the resin.
    c. Repeat this process for a total of five times.
10. Elute the immunoprecipitated complex with 40 µL of 4 M urea, incubate for 5 min at RT, clarify the eluate by centrifuging for 1 min at 2,000 x *g* and 4 °C, and transfer to a fresh 2 mL Eppendorf tube.
    a. Repeat the elution and combine supernatants for a total volume of 80 µL.
11. Proceed with SDS-PAGE/western blotting if optimizing aspects of immunoprecipitation procedure or continue with protein reduction, alkylation, and digestion if performing final, optimized method with downstream LC-MS/MS analysis.

5.6 SDS-PAGE and Western Blot

Immunoprecipitation procedures must be optimized to ensure robust enrichment of the protein of interest and its interacting partners. The bulk of the IP-MS optimization within this chapter involves adjusting the protein loading amounts in the IP. A series of experiments can be performed simultaneously, where each replicate contains a different protein loading amount. The elutions from these IPs can be collected and analyzed with western blotting following protein separation by SDS-PAGE. This procedure permits sensitive and selective recognition of the protein of interest, with the optimal protein loading amount determined by the detection of robust signal provided by the western blot secondary antibody.

5.6.1 SDS-PAGE

Procedure:

1. Prepare SDS-PAGE sample buffer.
    a. SDS-PAGE sample buffer is 5% β-mercaptoethnaol in 4x Laemmli buffer (Bio-Rad).
    b. Prepare enough buffer for the number of samples and blanks required (need 10 µL SDS-PAGE sample buffer for each lane).
2. Aliquot 30 µL of each sample into a fresh Eppendorf tube and add 10 µL of SDS-PAGE sample buffer.
    a. Prepare blanks by adding ratio of 3:1 water:SDS-PAGE sample buffer to obtain total volume needed for blank lanes.
3. Heat samples on the heat block for 5 min at 95 °C.
4. Remove comb and wrapping from a 4-20% Criterion™ TGX™ Precast Midi Protein Gel, 12+2 well, 45 µL (Bio-Rad). Rinse with milli-Q water and load into the gel box.
5. Fill the upper and lower gel compartments with SDS-PAGE running buffer.
    a. SDS-PAGE running buffer is 0.25 M Tris base, 1.92 M glycine, 1% sodium dodecyl sulfate.
6. Add 30 µL of either sample or blank into each of the larger, middle lanes.
    a. Adding blanks to empty wells ensures even sample migration down the gel.
7. Add 5 µL of Precision Plus Protein Dual Color Standards (Bio-Rad) to the 2 smaller, outer ladder lanes.
8. Place the gel box lid onto the gel box, matching up the contacts of the same color on both the gel box lid and the power supply.
9. Run the gel for 10 min at 50 V to allow the samples to migrate into the gel. Continue at 150 V for 45 min until the dye front reaches about 1 cm from the bottom of the gel.

10. Ensure that the power source is off, take the gel box lid off of the gel box, remove gel from the compartment, crack open plastic container to obtain the gel, and rinse gel with milli-Q water 3 times.
11. Proceed with western blotting.

5.6.2 Western Blot

Day 1:

1. Wash gel in transfer buffer for 10 min, shaking at RT.
    a. Transfer buffer is 0.025 M Tris base, 0.192 M glycine, 20% methanol
2. Cut polyvinylidene fluoride (PVDF) membrane to obtain slice that is same size of gel.
3. Activate PVDF membrane by rinsing in methanol for 1 min, and then transfer buffer for 1 min.
4. Assemble western blot stack in transfer buffer.
    a. Order of stack: black side of gel holder cassette, foam pad, filter paper, gel, PVDF membrane, filter paper, foam pad, red side of gel holder cassette.
5. Roll out the stack in transfer buffer to remove any bubbles that can interfere with the transfer.
6. Insert western blot stack into the western blot box, ensuring that the colors of the cassette and the electrode assembly match.
7. Place an ice pack/stir bar in the western blot box and add transfer buffer to the western blot box fill line.
8. Fix the western blot box lid onto the gel box, matching up the contacts of the same color on the western blot box lid.

9. Take the western blot box to the cold room, place it on the magnetic stir plate next to the power supply, and match the colors of the connects from the western blot box to the colors on the power supply.
10. Set the magnetic stir plate at a medium spin rate, turn the power supply on, and ensure that the power supply is set to a constant current of 10 mA.
11. Run the transfer overnight in the cold room

Day 2:

12. Disassemble stack and rinse PVDF membrane in TBST.
    a. TBST is 0.1% Tween 20 in 1x Tris-buffered saline (TBS, diluted from 20x TBS stock in chemical room).
13. Block membrane with 15 mL of blocking solution for 1 h, shaking at RT.
    a. Blocking solution is 3% bovine serum albumin (do not use MS grade) in TBST.
14. Incubate membrane with 15 mL of primary antibody solution in the cold room, shaking overnight.
    a. Primary antibody solution is the antibody used in the immunoprecipitation added to blocking solution.
    b. The amount of primary antibody needed for this solution should be listed in the manufacturer's product details.

Day 3:

15. Wash membrane 3 times with 15 mL of TBST for 5 min, shaking at RT.
16. Incubate membrane with 15 mL of secondary antibody solution for 1 h, shaking at RT.

a. Secondary antibody solution is the secondary antibody added to blocking solution.
b. The secondary antibody is an antibody specific to the primary antibody's Fc region and provides some form of measurable signal (*e.g.*, fluorescence)
c. The amount of secondary antibody needed for this solution should be listed in the manufacturer's product details.

17. Wash membrane 3 times with 15 mL of TBST for 5 min, shaking at RT.
18. Detect presence of secondary antibody.
    a. If using secondary antibody that contains a fluorophore, the Typhoon 9400 scanner (GE Life Sciences) in the Lockett lab can be used to image the membrane after gaining permission.
    b. The excitation and emission wavelengths for the secondary antibody fluorophore should be listed in the manufacturer's product details.
    c. The optimal protein loading amount for the immunoprecipitation can be determined by robust indirect detection of pulled-down protein of interest through the secondary antibody at the correct molecular weight on the membrane.

5.7 Protein Reduction, Alkylation, and Digestion

In these XL- and IP-MS procedures, a bottom-up proteomics approach is implemented, where protein cysteines are reduced and alkylated before trypsin digestion. The reductant DTT and the alkylating agent IAM are typically used in Hicks Laboratory and are the most commonly used reagents in bottom-up proteomics. For the XL-MS workflows described here, the reductant TCEP and alkylating agent chloroacetamide (CAA) are substituted for these processes as they allow for simultaneous protein reduction and alkylation at elevated temperatures, which results

in increased efficiency[15]. In the IP-MS protocol, the typical reduction and alkylation reagents are used as the disulfide bond in the DSP spacer arm is readily cleaved with 10-50 mM DTT at 37 °C.

5.7.1 Protein Reduction, Alkylation, and Digestion for XL-MS

Day 1:

1. Precipitate crosslinked proteins with 5 volumes of cold acetone.
2. Pellet proteins by centrifuging for 20 min at 3,200 x *g* and 4 °C, and discard the supernatant.
3. Resuspend proteins with 200 µL of 8 M urea in 100 mM Tris, pH 8.
4. Add 10 mM TCEP and 40 mM CAA. Incubate samples for 1 h at 37 °C covered from light.
5. Dilute samples to 1.6 M urea with 100 mM Tris, pH 8.
    a. Effective protein digestion with trypsin requires urea concentration < 2 M
6. Add 60 µg of Trypsin Gold (Promega) and incubate overnight (> 16 h) at 37 °C.
    a. Trypsin is prepared by dissolving in 50 mM acetic acid to 0.5 µg/µL and storing aliquots in the -80 °C freezer.

Day 2:

7. Quench the protein digestion by adding 10% trifluoroacetic acid (TFA) to approximately pH 3.
8. Proceed with solid-phase extraction (SPE) for desalting.

5.7.2 Protein Reduction, Alkylation, and Digestion for IP-MS

Day 1:

1. Add 20 mM DTT and incubate for 30 min at 37 °C.
2. Add 60 mM IAM and incubate for 45 min at RT covered from light.
3. Dilute samples to 1.6 M urea with 100 mM Tris, pH 8.
    a. Effective protein digestion with trypsin requires urea concentration < 2 M
4. Add 1 μg of Trypsin Gold (Promega) and incubate overnight (> 16 h) at RT.

Day 2:

5. Quench the protein digestion by adding 10% TFA to approximately pH 3.
6. Proceed with ethyl acetate washes for removal of non-ionic detergent.

5.8 Ethyl Acetate Washes

Non-ionic detergents preserve native protein structures and protein-protein interactions during protein extraction for IP- and XL-MS experiments, but detergents significantly interfere with the reversed-phase LC separation and MS ionization efficiency of peptides. Unfortunately, common methods for ionic detergent removal such as SPE and protein precipitation are ineffective for non-ionic detergents. Liquid-liquid extraction using water-saturated ethyl acetate is an alternative method and efficient in removing non-ionic detergents from peptide samples[16]. This clean up step should be performed before reversed-phase SPE, as non-ionic detergents are extremely hydrophobic and outcompete peptides for binding to C18 resin.

Procedure:

1. Mix equal volumes of ethyl acetate and milli-Q water in a screw cap glass bottle.
2. Invert immiscible mixture multiple times to create water-saturated ethyl acetate.
    a. After inverting, wait ~1 h to let layers separate. After settling, the top layer is ethyl acetate, while the bottom layer is water.

3. Add 4 volumes of water-saturated ethyl acetate to each sample and vortex for 10 s.
4. Centrifuge samples for 30 s at 15,000 x *g* and 4 °C.
5. Remove and discard the top ethyl acetate layer using a pipette.
    a. Ensure that no sample loss occurs during this step. Try to remove as much ethyl acetate as possible, but it is better to leave a little ethyl acetate rather than accidently discard some of the aqueous layer.
6. Repeat this process of addition and removal of ethyl acetate 5 times.
7. Proceed with reversed-phase SPE for desalting and LC-MS/MS analysis

5.9 Solid-Phase Extraction

Detergents, salts, and chaotropes in extraction buffers facilitate cell lysis and protein solubilization, but are not compatible with LC-MS. Reversed-phase SPE is commonly included in bottom-up proteomics workflows prior to LC-MS/MS analysis to isolate tryptic peptides of interest away from these interferents. In the Hicks Laboratory, reversed-phase SPE is performed using C18 ZipTips (Millipore Sigma) when samples contain < 10 μg peptides or Sep-Pak C18 cartridges (Waters) when larger amounts of peptides are present. Reversed-phase SPE using Sep-Pak C18 cartridges is employed in IP- and XL-MS experiments, and is described in detail in the following section. The sample loading capacity for Sep-Pak C18 cartridges is estimated to be 1 mg of trypsin-digested proteome per 50 mg sorbent, so sorbent weight and solvent volumes should be adjusted based on the samples being processed. The Hicks Laboratory owns a 24-position vacuum manifold (Phenomenex) that features 2-way stopcocks to secure Sep-Pak C18 cartridges and allows for batch sample processing to minimize technical variability.

Procedure:

1. Install a 50 mg/1.0 mL Sep-Pak C18 cartridge into the stopcock secured on the vacuum manifold.
    a. Do not exceed a flow rate of 1 drop/s in any step.
    b. Do not let the cartridge dry out in any step except for the elution step.
2. Pre-elute the cartridge with 1 mL of 80% acetonitrile/0.1% TFA.
3. Condition the cartridge with 1 mL of 0.1% TFA.
4. Ensure that samples are acidified to pH < 3, centrifuge for 5 min at 15,000 x *g* and 4 °C to pellet any precipitated protein, and load onto the cartridges in two passes.
5. Wash the cartridge with 2 mL of 0.1% TFA.
6. Elute peptides from the cartridge with 1 mL of 80% acetonitrile/0.1% TFA. When the sample stops eluting, apply increasing vacuum to the manifold until the cartridge bed is completely dried and all sample is eluted.
7. Freeze eluate in -80 °C freezer before vacuum centrifuging samples to dryness.
8. Proceed with SCX fractionation in XL-MS experiments or directly to LC-MS/MS analysis in IP-MS experiments.

5.10 Strong Cation Exchange Fractionation

Even with continuous MS hardware advances, the dynamic range of peptide abundances in a one-dimensional separation of a bottom-up proteomics sample precludes identification of low abundance peptides[17]. This is an issue for XL-MS, where non-crosslinked peptides greatly outnumber and suppress the ionization of the low abundant crosslinked peptides. Orthogonal separations integrated into proteomics workflows can decrease sample complexity and allow for improved peptide coverage. To this end, crosslinked peptides are enriched using SCX fractionation prior to LC-MS/MS analysis. SCX leverages the highly positively charged

crosslinked peptides to separate them from the less charged linear, non-crosslinked peptides. SCX fractionation simultaneously enriches for crosslinked peptides and decreases sample complexity, thus increasing the depth of coverage for crosslinked peptides.

Dried samples are resuspended in 200 µL of 10 mM potassium phosphate monobasic, 20% acetonitrile, pH 2.7. Configurations for the Prominence HPLC System (Shimadzu) are listed in Table 5.1. Mobile phase A consists of 10 mM potassium phosphate monobasic, 20% acetonitrile, pH 2.7, mobile phase B is 10 mM potassium phosphate monobasic, 250 mM potassium chloride, 20% acetonitrile, pH 2.7, and mobile phase C is 10 mM potassium phosphate monobasic, 600 mM potassium chloride, 20% acetonitrile, pH 2.7. Peptides are fractionated on a PolySulfoethyl A column (100 mm x 4.6 mm, 3 µm particles; PolyLC) using a linear gradient (Table 5.2) of increasing mobile phases B and C at a flow rate of 0.5 mL/min. After 10 min of 100% mobile phase A, mobile phase B is increased from 0% to 15% in 9.3 min, where it is held for 8.7 min before ramping to 30% in 8 min, where it is held for 11 min and then ramping to 100% in 5 min, where it is held for 5 min. After, mobile phase C is increased from 0% to 100% in 5 min before returning to 100% mobile phase A in 5 min and re-equilibrating for 25 min. After 10 min into the gradient, fractions are collected every 1 min. The fractions are desalted using reversed-phase SPE as described in Section 5.9 before vacuum centrifuging fractions to dryness.

5.11 LC-MS/MS Analysis

Dried samples are resuspended 0.1% TFA, where the necessary volume largely depends on the experiment type. A loading level of 1 µg digest is ideal, but can be hard to predict following enrichment or fractionation. For these experiments, 15-20 µL was found to be the proper resuspension volume, but if someone is performing these procedures for the first time it is strongly recommended that samples are resuspended in a greater volume. Following

resuspension, a test injection (1 µL) should be performed to ensure appropriate loading levels (*i.e.*, TIC intensity does not exceed 2E10 on the Thermo Q Exactive HF-X).

Configurations for the ACQUITY UPLC M-Class System (Waters) are listed in Table 5.3. Mobile phase A consists of water with 0.1% formic acid (Thermo Fisher Scientific) and mobile phase B is acetonitrile with 0.1% formic acid. Injections are made to a Symmetry C18 trap column (100 Å, 5µm, 180µm x 20 mm; Waters) with a flow rate of 5 µL/min for 3 min using 99% A and 1% B. Peptides are then separated on a HSS T3 C18 column (100 Å, 1.8µm, 75µm x 250 mm; Waters) using a linear gradient (Table 5.4) of increasing mobile phase B at a flow rate of 300 nL/min. Mobile phase B is increased from 5% to 35% in 90 min before ramping to 85% in 5 min, where it is held for 10 min before returning to 5% in 2 min and re-equilibrating for 13 min.

The Q Exactive HF-X mass spectrometer (Thermo Fisher Scientific) is operated (Table 5.5) in positive polarity and the Nanospray Flex source (Thermo Fisher Scientific) has spray voltage floating at 2.1 kV, capillary temperature at 325 °C, and funnel RF level at 40. Lock masses of background polysiloxane ions are included. During acquisition (Table 5.6), MS survey scans are collected with a scan range of 350 – 2000 *m/z* at a resolving power of 120,000 and an AGC target of 3×10^6 with a maximum injection time of 50 ms. A top 20 data-dependent acquisition is used where HCD fragmentation of precursor ions having +2 to +7 charge state is performed using a normalized collision energy setting of 28. MS/MS scans are performed at a resolving power of 30,000 and an AGC target of 1×10^5 with a maximum injection time of 100 ms. Dynamic exclusion for precursor *m/z* is set to a 10 s window.

5.12 Data Processing and Analysis for IP-MS

For the IP-MS experiments described here, label-free quantification (LFQ), database searching, data processing, and differential analysis can be carried out using the typical Hicks Laboratory data analysis pipeline. Retention time alignment, peak feature detection, peptide mapping, and protein quantification of LC-MS/MS runs is performed using Progenesis and Mascot. The only change from the Hicks Laboratory global proteome SOPs is the inclusion of the fixed modification "CAMthiopropanoyl (K)" in the IP-MS Mascot database searching to account for the reduced DSP crosslinker modified with IAM.

After protein measurements are exported from Progenesis, the Global LFQ workflows from the *QuantifyR* package (https://github.com/hickslab/QuantifyR) can be applied to the data for processing, statistical analysis, and result visualization. To determine interacting proteins for the protein of interest, two thresholds are applied to identify proteins that are significantly increased in abundance in the strain containing the epitope-tagged protein of interest compared to the wild type strain. The first threshold is a fold change in protein abundance between strains and the second threshold is statistical significance following a two-sided, equal-variance *t*-test. Commonly in the Hicks Laboratory, thresholds of fold change ≥ 2, and false discovery rate (FDR)-adjusted p-value < 0.05 using the method of Benjamini and Hochberg are applied[18], but these can be adjusted to provide a more stringent or lenient analysis.

5.13 Data Processing and Analysis for XL-MS

The current Mascot software (version 2.5.0) in the Hicks Laboratory does not have the capabilities to perform crosslinked peptide database searching, so the typical data analysis pipeline used in the Hicks Laboratory (Progenesis/Mascot) cannot be implemented for workflows involving MS-cleavable crosslinkers. In comparisons analyzing DSSO-crosslinked,

SCX-fractionated Chlamydomonas protein digest, Proteome Discoverer (version 2.5, Thermo Fisher Scientific, Waltham, MA) with XlinkX[19] (version 2.5) provided a greater depth of crosslinked peptide coverage for global interactomics over various freeware (*e.g.*, pLink2[20], MetaMorpheus[21], MeroX[22]). Proteome Discoverer is a proprietary software, but a 60-day free trial can be downloaded at https://thermo.flexnetoperations.com/control/thmo/login after registering an account. Proteome Discoverer differs from Progenesis/Mascot in that data processing and database searching settings are configured altogether as nodes, and then applied to a sample set at once as workflows. The workflows outlined in Section 5.13.1 are adapted from Klykov *et al.* and Proteome Discoverer node settings that are different from the default are listed[15].

Crosslinks identified through Proteome Discoverer are grouped into two types: intralinks (crosslinks between two tryptic peptides within the same protein), and interlinks (crosslinks between two tryptic peptides from two different proteins). Detected intralinks that can be overlayed onto known protein structures or homology models can be used to evaluate the intralink dataset by comparing the Euclidean distance between the crosslinked residues to the maximum crosslinker distance[23]. In cases where few proteins of a species of interest have existing structures in the Protein Data Bank (PDB), the Integrative Modeling Platform within XLinkDB 3.0 can used for homology modeling and this process is outlined in Section 5.13.2[24,25]. Intralinks can also be used as constraints to guide computational modeling for novel protein structures using the iterative threading assembly refinement (I-TASSER) server[26]. This computational protein structure modeling using I-TASSER is described in Section 5.13.3.

Detected interlinks can be used to generate protein-protein interaction models when known structures are unavailable from other structural approaches. Known protein structures from PDB

or protein structure models from I-TASSER are imported into High Ambiguity Driven protein-protein Docking (HADDOCK) for protein complex docking using identified interlinks as constraints[27]. The workflows for HADDOCK protein complex docking are adapted from Orbán-Németh *et al.* and outlined in Section 5.13.4[28]. These protein-protein interactions models and their corresponding detected interlinks can be visualized using ChimeraX and the functionalities of this software are described in Section 5.13.5[29,30]. Furthermore, these detected interlinks inform the protein-protein interactome for a particular species. The protein interactome can be mapped using xiNET and this process is explained in Section 5.13.6[31].

5.13.1 Proteome Discoverer

1. Open Thermo Proteome Discoverer (version 2.5) and create a new experiment by selecting the "New Study/Analysis…" button.
    a. Enter a "Study Name" with identifying details (*e.g.*, date, initials, and study).
    b. Choose a folder in "Study Root Directory" to select where study/analysis files will be saved.
    c. If Processing and Consensus Workflows have already been created, these can be input at this time.
2. Select "Add Files" and import all acquired data from fractions.
3. If the necessary protein database has not been imported into Proteome Discoverer yet, select the "Administration" tab, and then select "Maintain FASTA Files"
    a. Select "Add", find/select the necessary protein database, and click "Open".
    b. Wait until the protein database is imported into Proteome Discoverer.
4. Select "New Analysis" and then "Edit" in the Processing Step box from the Analysis section that appears.

a. This will automatically switch to the “Workflows” tab.

b. Finished Processing Workflow can be found in Figure 5.1

5. Drag the “Spectrum Files” node onto the Workflow Tree.

    a. This node has no associated settings.

6. Drag the “Spectrum Selector” node onto the Workflow Tree.

    a. Select “Use MS(n-1) with Parent Precursors” for the “Precursor Selection” setting.

    b. Select “HCD” for the “Unrecognized Activation Type Replacements” setting.

7. Drag the “XlinkX/PD Detect” node onto the Workflow Tree.

    a. Select “MS2” for the “Acquisition strategy” setting.

    b. Select “True” for the “Enable protein N-terminus linkage” setting.

    c. Select “DSSO / +158.004 Da (K)” for the “Crosslink Modification” setting.

8. Drag the “XlinkX/PD Filter” node onto the Workflow Tree.

    a. Select “Crosslinks” for the “Select” setting.

9. Drag the “XlinkX/PD Search” node onto the Workflow Tree.

    a. Select “Trypsin (Full)” for the “Enzyme” setting.

    b. Select the necessary protein database for the “Protein Database” setting.

    c. Set the “Maximum Missed Cleavages” setting to “2”.

    d. Set the “Fragment Mass Tolerance” setting to “0.02 Da”.

    e. Set the “Precursor Mass Tolerance” setting to “15 ppm”.

    f. Select “Carbamidomethyl / +57.021 Da (C)” for the “Static Modification” setting.

    g. Select “Oxidation / +15.995 Da (M)” for the “Dynamic Modification” setting.

h. Select “Acetyl / +42.011 Da (N-Terminus)” for the “Dynamic Protein N-term Modification” setting.

10. Drag the “XlinkX/PD Validator” node onto the Workflow Tree.
    a. Set the “FDR Threshold” setting to “0.01”.
11. Drag another “XlinkX/PD Filter” node onto the Workflow Tree and ensure that the node is connected to the “XlinkX/PD Detect” node.
    a. Select “Peptides” for the “Select” setting.
12. Drag the “Sequest HT” node onto the Workflow Tree and ensure that the node is connected to the second “XlinkX/PD Filter” node that is filtering for “Peptides”.
    a. Select “Trypsin (Full)” for the “Enzyme” setting.
    b. Select the necessary protein database for the “Protein Database” setting.
    c. Set the “Maximum Missed Cleavages” setting to “2”.
    d. Set the “Fragment Mass Tolerance” setting to “0.02 Da”.
    e. Set the “Precursor Mass Tolerance” setting to “15 ppm”.
    f. Set the “Weight of a Ions”, “Weight of b Ions”, and “Weight of y Ions” settings to “1”.
    g. Set the “Maximum Equal Modifications Per Peptide” setting to “3”.
    h. Select “Carbamidomethyl / +57.021 Da (C)” for the “Static Modification” setting.
    i. Select “Oxidation / +15.995 Da (M)”, “DSSO Tris / +279.078 Da (K)”, and “DSSO Hydrolyzed / +176.014 Da (K)” for the “Dynamic Modification” setting.
    j. Select “Acetyl / +42.011 Da (N-Terminus)” for the “Dynamic Protein N-term Modification” setting.

13. Drag the “Percolator” node onto the Workflow Tree and ensure that the node is connected to the “Sequest HT” node.
    a. Select “Concatenated” for the “Target/Decoy Selection” setting.
    b. Set the “Target FDR (Strict)” setting to “0.01”.
    c. Set the “Target FDR (Relaxed)” setting to “0.05”.
14. Name the Processing Workflow in the “Workflow” address bar and save by selecting “Save”.
15. Select “Edit” in the Consensus Step box from the Analysis section.
    a. Finished Consensus Workflow can be found in Figure 5.2
16. Drag the “MSF Files” node onto the Workflow Tree.
    a. Select “Identified” for the “Spectra to Store” setting.
17. Drag the “PSM Grouper” node onto the Workflow Tree and use the default settings.
18. Drag the “Peptide Validator” node onto the Workflow Tree and use the default settings.
19. Drag the “Peptide and Protein Filter” node onto the Workflow Tree and use the default settings.
20. Drag the “Protein Scorer” node onto the Workflow Tree.
    a. This node has no associated settings.
21. Drag the “Protein FDR Validator” node onto the Workflow Tree and use the default settings.
22. Drag the “Protein Grouping” node onto the Workflow Tree, ensure that the node is connected to the “Protein Scorer” node, and use the default settings.
23. Drag the “XlinkX/PD Crosslink Grouping” node onto the Workflow Tree and ensure that the node is connected to the “Protein Grouping” node, and use the default settings.

24. Drag the “XlinkX/PD Crosslink Export” node onto the Workflow Tree and ensure that the node is connected to the “XlinkX/PD Crosslink Grouping” node.
    a. Choose a folder in “xiNET Output Files Directory” to select where output files will be saved.
25. Name the Consensus Workflow in the “Workflow” address bar and save by selecting “Save”.
26. Select the “Input Files” tab, highlight all the files that will be analyzed, and drag them into the “Files for Analysis” section in the Processing Step box in the Analysis section.
27. Name the output files in the “Result File” address bars found in the Analysis section for both the Processing (.msf file) and Consensus (.pdResult file) Steps.
28. Save the Analysis Template in the Analysis section, and click “Run” to proceed with the data processing and database searching.
29. Once the analysis is complete (analysis of ~60 fractions can take multiple days to finish), select the “Analysis Results” tab, and double click the row containing the corresponding pdResult file.
30. In the pdResult file, there are multiple tabs that contain tables of results (*e.g.*, crosslinks, CSMs) available for export and further analysis.
    a. Select all rows in a given table, right click, select “Copy with Headers”, and paste into Excel to export results.
    b. Selecting “Export” in the “File” tab seems to only export database search results from Sequest HT (non-crosslinked peptides).

5.13.2 Intralink Euclidean Distances using XlinkDB

Procedure:

1. Prepare identified crosslinks for input into XlinkDB by arranging data into six columns as follows: Peptide A, Protein Accession A, Crosslink position A, Peptide B, Protein Accession B, Crosslink position B. Save this as a tab delimited text file.
    a. Use Uniprot identifiers for protein accessions.
    b. The crosslink position is the position of the crosslinked site within the peptide starting from 0.
    c. Remove all column header names (*i.e.*, the first row should contain data from the first identified crosslink).
2. On the XlinkDB homepage (http://xlinkdb.gs.washington.edu/xlinkdb/), scroll to the "Upload New Data" section.
3. Upload the tab delimited text file prepared in step 1.
4. Choose the organism studied in this experiment.
5. Choose the crosslinker used in this experiment.
6. Provide information on "Experiment name", "Description", "Lab name", and "Email address".
7. Keep this data private, by keeping the box labeled "I want my published data to be available to the public!" unchecked.
    a. Private data can be accessed if the "Experiment name" is known and therefore this should be included in Experimental sections of manuscripts.
8. Upload the data by selecting "Upload". This process, including the calculation of Euclidean distances of intralinks that are mapped onto known structures or homology models, should take about one day.

9. Once uploaded, scroll to the “Generate an XL-MS Interaction Table” section, input the “Experiment name” into the “Network Name” address bar, and select “Generate Network Table”.
10. Filter the table for intralinks by highlighting “intra” in the “Cross-link type” section and selecting “Refresh Network Table”.
11. For each intralink, view the mapped known structures or homology models by selecting the hyperlink number in the column labeled “Available Structures”.
    a. When there are multiple structures for a given intralinks, use tiebreakers in the following order to select one for its corresponding calculated Euclidean distance (labeled “XL Distance” in table):
        i. Use known structures over homology models.
        ii. Lowest “Maximum expect” score.
        iii. Highest number of “Mapped Crosslinks”.
        iv. Lowest “XL Distance”.
12. Manually tabulate Euclidean distance for each mapped intralink.
13. View the mapped intralinks on the known structure or homology model in NGL Viewer by selecting “View Structure” in the column labeled “NGL Viewer”.
    a. Left click rotates the protein structure.
    b. Right click moves the protein structure translationally.
14. Remove atoms of crosslinked residues by selecting the button featuring three bars in the “spacefill” tab and slide the “opacity” setting to “0”.
15. Adjust visualized crosslink settings by selecting the button featuring three bars in the “distance” tab.

16. Adjust protein ribbon structure color by selecting the button featuring three bars in the "cartoon" tab and selecting an option in the "colorScale" dropbox.
17. Export image of protein structure by selecting "File", "Export image…", and "export".

5.13.3 Protein Structure Modeling using I-TASSER

Procedure:

1. Go to the I-TASSER homepage (https://zhanglab.dcmb.med.umich.edu/I-TASSER/).
2. Create an account by selecting "please click here if you do not have a password" in the "Password section" and fill out the form.
3. Paste the amino acid sequence of the protein to be modeled into the provided form. Ensure the sequence is in FASTA format.
4. If adding restraints based on identified intralinks, click on "Option I: Assign additional restraints & templates to guide I-TASSER modeling."
    a. Prepare a .txt file containing the list of all detected intralinks, along with the maximum crosslinking distance based on the crosslinker used (30 Å for DSSO). Use the following syntax in each row to define these crosslinks: DIST (tab) <1st residue position> (tab) CA (tab) <2nd residue position> (tab) CA (tab) < maximum crosslinking distance>
    b. Example for intralink between Lys25 and Lys50 when using DSSO as the crosslinker: DIST (tab) 25 (tab) CA (tab) 50 (tab) 30
    c. Upload the restraint .txt file by selecting "Assign contact/distance restraints" in the Option 1 submenu and selecting the .txt file.
5. Enter account email/password and enter an ID for the modeling job.
6. Keep the box labeled "Keep my results public…" checked.

    a. When making results private, I-TASSER is supposed to send an email containing a code to access the final results, but sometimes this email is never sent.

7. Select “Run I-TASSER” to start protein modeling.
    a. Run time is dependent on protein length. Proteins close to 1000 amino acids in length can take a week to finish modeling. Proteins with a couple hundred amino acids can take one day or two to finish modeling.
    b. I-TASSER prevents more than one job running from the same computer/email.
8. Monitor modeling job by selecting “Queue” above the form used to provide the protein sequence and locating the inputted ID for the specific job (listed as “Protein Name”).
9. Once the job is completed, select the ID. The top 10 templates used for modeling can be found in the section labeled “Top 10 threading templates used by I-TASSER” and the top 5 protein structure models can be found in the section labeled “Top 5 final models predicted by I-TASSER”.
    a. The confidence of each model is reported as a “C-score”, in a range of -5 to 2, where 2 is the best.
10. Download the .pdb file for the top model (Model 1) by selecting “Download Model 1”.

5.13.4 Protein Complex Docking using HADDOCK

Procedure:

1. Visit the HADDOCK homepage (https://wenmr.science.uu.nl/haddock2.4/).
2. Register for a new account by selecting “Register” and entering the listed information.
3. Once confirmed, login and go to your account information. Request Expert Access Level by selecting “Request Elevated Access” in the “Access Level” tab and fill out the form.

4. Calculate the active and passive residues of each protein structure model with CPORT (http://milou.science.uu.nl/services/CPORT/).
    a. Unfold the “Protein structure to predict” section.
    b. Select “I am submitting it” in the “Where is the structure provided” tab.
    c. If the protein being submitted only has 1 subunit, select “A” in the “Which chain of the structure must be used”. If the protein has multiple subunits, select the subunit of interest in this section.
    d. Select “Choose File” in the “PDB structure to submit” and upload the .pdb file from the I-TASSER protein structure modeling.
    e. Select “Very sensitive (recommended for HADDOCK)” in the “Threshold” tab.
    f. After ~10 min, a web page will open providing a link to the results. Select that link and note the calculated active and passive residues for the given protein model.
    g. Repeat this process for the other protein in the proposed interaction.
5. On the HADDOCK homepage, select “Submit a new job”.
6. Set the settings for the “Input data” tab.
    a. Name the job in the “Job name” address bar.
    b. Unfold the “Molecule 1 – input” tab.
    c. If the protein being submitted only has 1 subunit, select “A” in the “Which chain of the structure must be used”. If the protein has multiple subunits, select the subunit of interest in this section.
    d. Select “Choose File” in the “PDB structure to submit” and upload the .pdb file for the first protein from the I-TASSER protein structure modeling.

e. Use the default settings for everything else.

f. Repeat steps b-e for the second protein in the “Molecule 2 – input” tab.

7. Select “Next” and set the settings for the “Input parameters” tab.

    a. Unfold the “Molecule 1 – parameters” and “Active/Passive residues – Selection #1” tabs.

    b. List the active residues for the first protein obtained from CPORT in the “Active residues (directly involved in the interaction)” address bar.

    c. Set the “Automatically define passive residues around the active residues” slider to “Off” (left).

    d. List the passive residues for the first protein obtained from CPORT in the “Passive residues (surrounding surface residues)” address bar.

    e. Use the default settings for everything else.

    f. Repeat steps a-e for the second protein in the “Molecule 2 – parameters” tab.

8. Select “Next”, select “OK” if a warning about the number of active residues appears, and set the settings for the “Docking parameters” tab.

    a. If adding restraints based on identified interlinks, open Excel and prepare a file containing the list of all detected interlinks, along with the maximum crosslinking distance based on the crosslinker used (30 Å for DSSO). Use the following syntax in each row (each input separated by commas should be in its own column) to define these crosslinks: assign, (resid <protein 1 residue position> and segid <protein 1 molecule chain>), (resid <protein 2 residue position> and segid <protein 2 molecule chain>), <target distance>, <lower distance margin>, <higher distance margin>. Save this as a .txt file.

b. Example for interlink between protein 1, Lys100 and protein 2, Lys200 when using DSSO as the crosslinker: assign, (resid 100 and segid A), (resid 200 and segid A), 30, 30, 0.
c. Upload the restraint .txt file by selecting “Choose File” in the “You can supply a HADDOCK restraints TBL file with restraints that will always be enforced (unambiguous restraints)” tab and selecting the .txt file.
d. Use the default settings for everything else.

9. Select “Submit” and wait until the job is done processing (typically takes ~6-8 hours to complete). An email will be sent when the job is complete.
10. Click the link in the email confirming the completion of the protein complex docking job.
11. Scroll down to find download links for structure .pdb files. Select the buttons for particular structures to download their corresponding files.
    a. Structures are grouped into different clusters, where the clusters are ordered starting with the most reliable on top.

5.13.5 Visualizing Interaction Models using ChimeraX

Procedure:

1. Open the ChimeraX software (version 1.2.5).
2. Select “Open” and find the .pdb file from the HADDOCK protein complex modeling to view the protein-protein interaction model.
    a. Left click rotates the protein structure.
    b. Right click moves the protein structure translationally.
    c. Control key and left click selects a residue.
    d. Control + Shift key and left click allows the selection of multiple residues.

e. Control key and left click in blank space unselects all residues/protein chains.

3. If adding crosslinks to the model based on identified intralinks, prepare a .txt file containing a list of all detected intralinks.
    a. The first four lines each contain a semicolon, a space after the semicolon, and a particular setting for the crosslinks. All settings require an input. These settings include “halfbond” (if set to true, the crosslink color will be split in two, where the color of each half is dictated by the protein color of that side of the crosslink, if set to false, the whole crosslink will be colored based on the input of the color setting), “color” (if halfbond is set to false, this will be the color of the crosslink), “radius” (dictates the size of the crosslink), and “dashes” (amount of dashes in the crosslink, setting of 0 outputs a solid crosslink).
    b. Example for these settings: “; halfbond = false(next line); color = red(next line); radius = 0.3(next line); dashes = 0”
    c. Use this syntax in the following rows to define these crosslinks: /a:< protein 1 residue position>@ca /b:< protein 2 residue position>@ca
    d. Example for interlink between protein 1, Lys100 and protein 2, Lys200: /a:100@ca /b:200@ca
    e. If particular crosslinks require different settings, create a new set of four lines including those new settings (*e.g.*, halfbond, color, radius, dashes) and include those crosslinks underneath those settings.
    f. Save this as a .pb file.
4. Set the working directory in ChimeraX to where the .pb file is located by typing “cd <file path>” in the command line at the bottom of the window.

5. Place the crosslinks onto the model by typing “open <.pb file including the “.pb” at the end>” in the command line at the bottom of the window.
6. Label the crosslinks with their Euclidean distances by typing “label pbonds size <number> height fixed” in the command line at the bottom of the window.
7. Change the color of each protein, by selecting “A” or “B” in the box labeled “Chain information for <file name>” from the log, and then selecting “Action”, “Color”, and a color from the provided list.
8. Export the image by selecting “File”, “Save”, and saving as an image file format.

5.13.6 Visualizing Interlinks using xiNET

Procedure:

1. Prepare identified crosslinks for input into xiNET by arranging data into eight columns as follows: PepPos1, PepPos2, Protein1, Protein2, PepSeq1, LinkPos1, PepSeq2, LinkPos2. Save this as a .csv file.
    a. PepPos refers to the position of the most N-terminal amino acid in the identified peptides within their original protein sequences.
    b. Protein refers to the protein accessions (does not have to be UniProt identifiers)
    c. PepSeq refers to the sequence of the identified peptides.
    d. LinkPos refers to the position of the crosslinked site within the peptides (0 means that the crosslink was identified on the N-terminus of the peptide).
    e. Ensure that columns are labeled with the listed header names.
2. Go to the xiNET homepage (http://crosslinkviewer.org/).
3. Upload the .csv file prepared in step 1 and its corresponding FASTA file (ensuring that the accessions match between the two files). Select “Upload”.

4. After being directed to the uploaded data, filter for interlinks by unchecking every type of link on the bottom of the webpage except for “Heteromeric”.
5. Adjust the layout of proteins containing detected interlinks to better visualize the data.
    a. Drag individual protein by left clicking on a specific protein.
    b. Translationally move across the screen by left clicking in an area where there are no proteins.
    c. Zoom using the mouse wheel.
    d. Select multiple proteins by right clicking and drawing a rectangle around multiple proteins, and drag all proteins by left clicking on one of the highlighted proteins.
6. Save the updated protein positions by typing a file name in the “Enter Save Layout Name” address bar and selecting “Save”. This save file can be viewed later by selecting “Load”.
7. Download an image of the updated protein positions by selecting “Download Image as SVG” in the top right corner.
8. View a Circos plot of the data by selecting the “Views” tab, and then selecting “Circular”.
    a. The proteins displayed in the Circos plot are based on the types of links checked on the bottom of the page (*e.g.*, Heteromeric, Self).
    b. Download an image of the Circos plot by selecting “Download Image as SVG”.
9. The web address can be shared as a link to allow others to view the uploaded data.

5.14 Tables

Table 5.1 Configuration for Shimadzu Prominence HPLC System operation.

HPLC system	Shimadzu Prominence HPLC
Mobile phase A	10 mM potassium phosphate monobasic, 20% acetonitrile, pH 2.7
Mobile phase B	10 mM potassium phosphate monobasic, 250 mM potassium chloride, 20% acetonitrile, pH 2.7
Mobile phase C	10 mM potassium phosphate monobasic, 600 mM potassium chloride, 20% acetonitrile, pH 2.7
Analytical column	PolyLC PolySulfoethyl A (100 mm x 4.6 mm, 3 μm particles)
Flow rate	0.5 mL/min
Run time	92 min

Table 5.2 Gradient conditions for Shimadzu Prominence HPLC System operation.

Time (min)	Mobile Phase A (%)	Mobile Phase B (%)	Mobile Phase C (%)
0	100	0	0
10	100	0	0
19.3	85	15	0
28	85	15	0
36	70	30	0
47	70	30	0
52	0	100	0
57	0	100	0
62	0	0	100
67	100	0	0
92	100	0	0

Table 5.3 Configuration for Waters ACQUITY M-Class UPLC System operation.

UPLC system	Waters ACQUITY M-Class UPLC
Mobile phase A	0.1% formic acid in water
Mobile phase B	0.1% formic acid in acetonitrile
Trap column	Waters Symmetry C18 (100 Å, 5 µm, 180 µm x 20 mm)
Trapping conditions	5 µL/min for 3 min 99% A and 1% B
Analytical column	Waters HSS T3 C18 (100 Å, 1.8 µm, 75 µm x 250 mm)
Flow rate	300 nL/min
Run time	120 min

Table 5.4 Gradient conditions for Waters ACQUITY M-Class UPLC System operation.

Time (min)	Mobile Phase A (%)	Mobile Phase B (%)
0	95	5
90	65	35
95	15	85
105	15	85
107	95	5
120	95	5

Table 5.5 Hardware configuration for Thermo Fisher Scientific Q Exactive HF-X operation.

MS system	Thermo Fisher Scientific Q Exactive HF-X
Ion source	Thermo Fisher Scientific Nanospray Flex source
Polarity	Positive
Spray voltage	2.1 kV
Capillary temperature	320 °C
RF level	40

Table 5.6 Method parameters for Thermo Fisher Scientific Q Exactive HF-X operation.

Software	Xcalibur
MS1 survey scan range	350 to 2000 *m/z*
MS1 resolution	120,000
MS1 maximum IT	50 ms
MS1 AGC target	3e6
DDA selection criteria	Top 20 features +2 to +7 charge state 5e4 intensity threshold
MS2 resolution	30,000
Normalized collision energy	28
MS2 maximum IT	100 ms
MS2 AGC target	1e5
Dynamic exclusion window	10 s
Isolation window	1.5 *m/z*

5.15 Figures

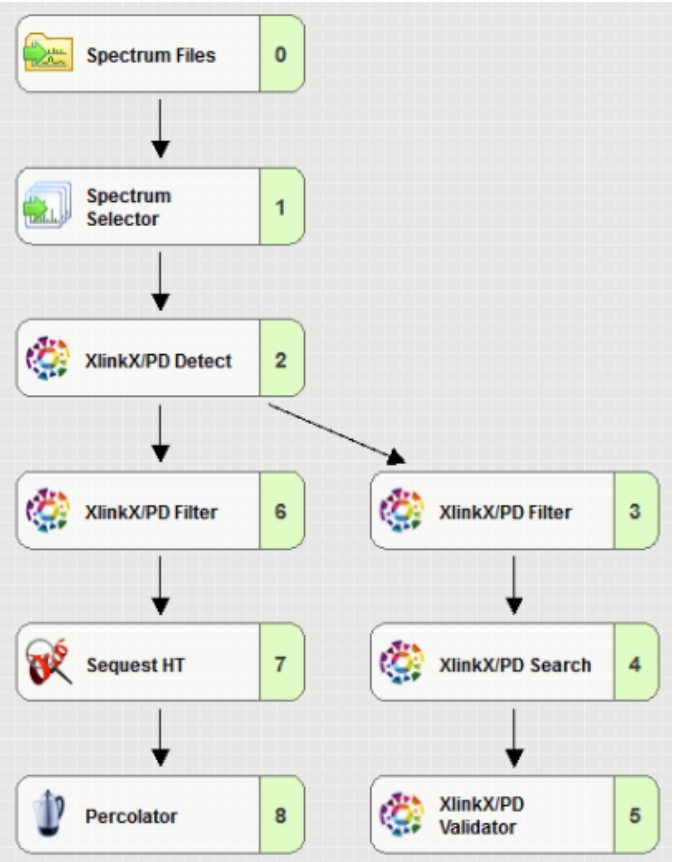


Figure 5.1 Proteome Discoverer processing workflow for XL-MS analysis.

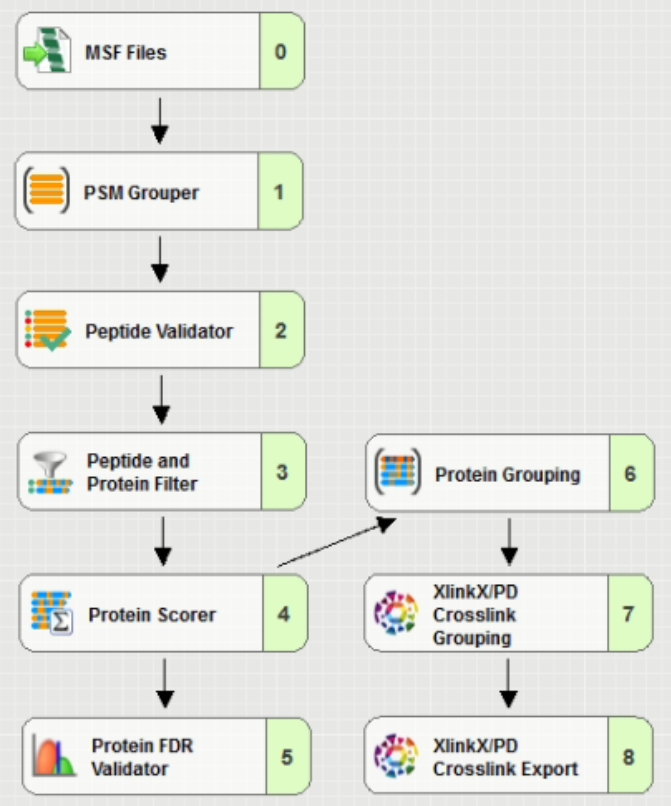


Figure 5.2 Proteome Discoverer consensus workflow for XL-MS analysis.

REFERENCES

1. Smythers, A. L.; McConnell, E. W.; Lewis, H. C.; Mubarek, S. N.; Hicks, L. M. Photosynthetic Metabolism and Nitrogen Reshuffling Are Regulated by Reversible Cysteine Thiol Oxidation Following Nitrogen Deprivation in *Chlamydomonas*. *Plants* **2020**, *9*(6), 784.

2. Ford, M. M.; Smythers, A. L.; McConnell, E. W.; Lowery, S. C.; Kolling, D. R. J.; Hicks, L. M. Inhibition of TOR in *Chlamydomonas reinhardtii* Leads to Rapid Cysteine Oxidation Reflecting Sustained Physiological Changes. *Cells* **2019**, *8*(10).

3. McConnell, E. W.; Smythers, A. L.; Hicks, L. M. Maleimide-Based Chemical Proteomics for Quantitative Analysis of Cysteine Reactivity. *J. Am. Soc. Mass Spectrom.* **2020**, *31*(8), 1697–1705.

4. McConnell, E. W.; Werth, E. G.; Hicks, L. M. The phosphorylated redox proteome of *Chlamydomonas reinhardtii*: Revealing novel means for regulation of protein structure and function. *Redox Biol.* **2018**, *17*, 35–46.

5. Werth, E. G.; McConnell, E. W.; Gilbert, T. S. K.; Couso Lianez, I.; Perez, C. A.; Manley, C. K.; Graves, L. M.; Umen, J. G.; Hicks, L. M. Probing the global kinome and phosphoproteome in *Chlamydomonas reinhardtii* via sequential enrichment and quantitative proteomics. *Plant J.* **2017**, *89*(2), 416–426.

6. Couso, I.; Pérez-Pérez, M. E.; Ford, M. M.; Martínez-Force, E.; Hicks, L. M.; Umen, J. G.; Crespo, J. L. Phosphorus Availability Regulates TORC1 Signaling via LST8 in *Chlamydomonas*. *Plant Cell* **2020**, *32*(1), 69–80.

7. Gorman, D. S.; Levine, R. P. Cytochrome f and plastocyanin: their sequence in the photosynthetic electron transport chain of *Chlamydomonas reinhardtii*. *Proc. Natl. Acad. Sci. U. S. A.* **1965**, *54*(6), 1665.

8. Miernyk, J. A.; Thelen, J. J. Biochemical approaches for discovering protein–protein interactions. *Plant J.* **2008**, *53*(4), 597–609.

9. Kao, A.; Chiu, C.; Vellucci, D.; Yang, Y.; Patel, V. R.; Guan, S.; Randall, A.; Baldi, P.; Rychnovsky, S. D.; Huang, L. Development of a Novel Cross-linking Strategy for Fast and Accurate Identification of Cross-linked Peptides of Protein Complexes. *Mol. Cell. Proteomics* **2011**, *10*(1), M110.002170.

10. Sinz, A. Divide and conquer: cleavable cross-linkers to study protein conformation and protein-protein interactions. *Anal. Bioanal. Chem.* **2017**, *409*(1), 33–44.

11. Lomant, A. J.; Fairbanks, G. Chemical probes of extended biological structures: Synthesis and properties of the cleavable protein cross-linking reagent [35S]dithiobis(succinimidyl propionate). *J. Mol. Biol.* **1976**, *104*(1), 243–261.

12. Young, C. L.; Britton, Z. T.; Robinson, A. S. Recombinant protein expression and

purification: A comprehensive review of affinity tags and microbial applications. *Biotechnol. J.* **2012**, *7*(5), 620–634.

13. Park, S. H.; Cheong, C.; Idoyaga, J.; Kim, J. Y.; Choi, J. H.; Do, Y.; Lee, H.; Jo, J. H.; Oh, Y. S.; Im, W.; Steinman, R. M.; Park, C. G. Generation and application of new rat monoclonal antibodies against synthetic FLAG and OLLAS tags for improved immunodetection. *J. Immunol. Methods* **2008**, *331*(1–2), 27–38.

14. Grodzki, A. C.; Berenstein, E. Antibody Purification: Affinity Chromatography – Protein A and Protein G Sepharose. *Methods Mol. Biol.* **2010**, *588*, 33–41.

15. Klykov, O.; Steigenberger, B.; Pektaş, S.; Fasci, D.; Heck, A. J. R.; Scheltema, R. A. Efficient and robust proteome-wide approaches for cross-linking mass spectrometry. *Nat. Protoc.* **2018**, *13*(12), 2964–2990.

16. Yeung, Y.-G.; Stanley, E. R. Rapid Detergent Removal From Peptide Samples With Ethyl Acetate For Mass Spectrometry Analysis. *Curr. Protoc. protein Sci.* **2010**, *16*(12).

17. Nielsen, M. L.; Savitski, M. M.; Zubarev, R. A. Extent of modifications in human proteome samples and their effect on dynamic range of analysis in shotgun proteomics. *Mol. Cell. Proteomics* **2006**, *5*(12), 2384–2391.

18. Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B* **1995**, *57*(1), 289–300.

19. Liu, F.; Rijkers, D. T. S.; Post, H.; Heck, A. J. R. Proteome-wide profiling of protein assemblies by cross-linking mass spectrometry. *Nat. Methods* **2015**, *12*(12), 1179–1184.

20. Chen, Z.-L.; Meng, J.-M.; Cao, Y.; Yin, J.-L.; Fang, R.-Q.; Fan, S.-B.; Liu, C.; Zeng, W.-F.; Ding, Y.-H.; Tan, D.; Wu, L.; Zhou, W.-J.; Chi, H.; Sun, R.-X.; Dong, M.-Q.; He, S.-M. A high-speed search engine pLink 2 with systematic evaluation for proteome-scale identification of cross-linked peptides. *Nat. Commun.* **2019**, *10*(1), 1–12.

21. Lu, L.; Millikin, R. J.; Solntsev, S. K.; Rolfs, Z.; Scalf, M.; Shortreed, M. R.; Smith, L. M. Identification of MS-Cleavable and Noncleavable Chemically Cross-Linked Peptides with MetaMorpheus. *J. Proteome Res.* **2018**, *17*(7), 2370–2376.

22. Götze, M.; Pettelkau, J.; Fritzsche, R.; Ihling, C. H.; Schäfer, M.; Sinz, A. Automated assignment of MS/MS cleavable cross-links in protein 3d-structure analysis. *J. Am. Soc. Mass Spectrom.* **2014**, *26*(1), 83–97.

23. Bartolec, T. K.; Smith, D.-L.; Pang, C. N. I.; Xu, Y. D.; Hamey, J. J.; Wilkins, M. R. Cross-linking Mass Spectrometry Analysis of the Yeast Nucleus Reveals Extensive Protein–Protein Interactions Not Detected by Systematic Two-Hybrid or Affinity Purification-Mass Spectrometry. *Anal. Chem.* **2019**, *92*(2), 1874–1882.

24. Russel, D.; Lasker, K.; Webb, B.; Velázquez-Muriel, J.; Tjioe, E.; Schneidman-Duhovny, D.; Peterson, B.; Sali, A. Putting the Pieces Together: Integrative Modeling Platform

Software for Structure Determination of Macromolecular Assemblies. *PLOS Biol.* **2012**, *10*(1), e1001244.

25. Schweppe, D. K.; Zheng, C.; Chavez, J. D.; Navare, A. T.; Wu, X.; Eng, J. K.; Bruce, J. E. XLinkDB 2.0: integrated, large-scale structural analysis of protein crosslinking data. *Bioinformatics* **2016**, *32*(17), 2716–2718.

26. Yang, J.; Zhang, Y. I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Res.* **2015**, *43*(W1), W174-81.

27. Van Zundert, G. C. P.; Rodrigues, J. P. G. L. M.; Trellet, M.; Schmitz, C.; Kastritis, P. L.; Karaca, E.; Melquiond, A. S. J.; Van Dijk, M.; De Vries, S. J.; Bonvin, A. M. J. J. The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J. Mol. Biol.* **2016**, *428*(4), 720–725.

28. Orbán-Németh, Z.; Beveridge, R.; Hollenstein, D. M.; Rampler, E.; Stranzl, T.; Hudecz, O.; Doblmann, J.; Schlögelhofer, P.; Mechtler, K. Structural prediction of protein models using distance restraints derived from cross-linking mass spectrometry data. *Nat. Protoc.* **2018**, *13*(3), 478.

29. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Meng, E. C.; Couch, G. S.; Croll, T. I.; Morris, J. H.; Ferrin, T. E. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* **2021**, *30*(1), 70–82.

30. Goddard, T. D.; Huang, C. C.; Meng, E. C.; Pettersen, E. F.; Couch, G. S.; Morris, J. H.; Ferrin, T. E. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci.* **2018**, *27*(1), 14–25.

31. Combe, C. W.; Fischer, L.; Rappsilber, J. XiNET: Cross-link Network Maps With Residue Resolution. *Mol. Cell. Proteomics* **2015**, *14*(4), 1137.

CHAPTER 6: Crosslinking Mass Spectrometry Unveils Novel Interactions and Structural Distinction in the Model Green Alga *Chlamydomonas reinhardtii*[1]

6.1 Introduction

Crosslinking mass spectrometry (XL-MS) is a powerful approach that couples biochemistry with molecular and structural biology through simultaneous analysis of protein-protein interactions (PPIs), conformations, and structure[1–5]. While alternative methods to investigate PPIs rely on genetic transformation (*e.g.*, the yeast two-hybrid assay) and/or the use of highly specific antibodies (*e.g.*, affinity-purification mass spectrometry), XL-MS is limited only by the reactivity and specificity of the chemical crosslinker, enabling global unbiased delineation of protein networks[6–10]. Further, while other structural approaches often require homogenous protein samples, large sample abundance, and/or crystallization, XL-MS can capture dynamic protein conformations in their native environment, without isolation or purification[11,12].

Chemical crosslinking involves the covalent linkage of two protein residues that are in close proximity. Crosslinker reagents vary in reactive groups (*e.g.*, NHS ester for amine reactivity, maleimide for cysteine reactivity) and spacer arm length (*e.g.*, 10.3 Å for DSSO). The maximum distance between crosslinked residues is limited by the spacer arm length, but it is accepted that this constraint can be exceeded due to dynamic protein conformational changes (*e.g.*, 30 Å limit for DSSO despite a 10 Å spacer arm)[13]. Crosslinking can be coupled with conventional bottom-up workflows (*i.e.*, proteolytic cleavage) for the identification of

[1] Reprinted with permission from Smythers, A. L.; Iannetta, A. A.; Hicks, L. M. Crosslinking mass spectrometry unveils novel interactions and structural distinctions in the model green alga *Chlamydomonas reinhardtii*. *Mol. Omics* **2021**, https://doi.org/10.1039/D1MO00197C.

crosslinked peptides. These detected crosslinks produce three-dimensional information for proteins; crosslinks between two tryptic peptides from two different proteins (interlinks) inform PPIs while crosslinks between two tryptic peptides within the same protein (intralinks) enhance structural understanding of that protein.

XL-MS is an increasingly powerful approach to examine PPIs, yet the requisite data processing is immensely challenging, as linking two peptides increases the proteome search space by n^2 and creates challenging MS/MS fragment spectra that is difficult to deconvolute. This has been partly addressed via a portfolio of search algorithms and bioinformatics platforms, yet still poses several ongoing challenges including: 1) increased missed cleavages due to crosslinks blocking potential cleavage sites, 2) altered ionization, 3) more complex MS/MS fragmentation, and 4) low abundance of crosslinked peptides compared to linear peptides[14–16]. MS-cleavable crosslinker development has greatly increased the applicability of XL-MS to systems-wide studies, as observed mass shifts from the short or long end of the cleaved crosslinker (*i.e.*, crosslink reporter ions), more efficient peptide fragmentation, and MS/MS spectra of increased quality simplify crosslink identification in database searching and lend confidence to site assignment[17].

Chlamydomonas reinhardtii is a unicellular green alga and one of the most widely studied models for photosynthesis, attributed in part to its rapid growth rate, large singular chloroplast, and well annotated genome[18,19]. *C. reinhardtii* is a beneficial model organism for studying fundamental biochemical processes, including autophagy, signal transduction, and nitrogen flux, among others[20–25]. Despite its long-term interest among plant scientists and cell biologists, its complex interactome has yet to be thoroughly investigated. In the STRING database, there are 1,278 experimentally confirmed (medium confidence score >400) unique

protein-protein interactions in *C. reinhardtii*, compared to 31,283 in *Arabidopsis thaliana* and 51,599 in *Saccharomyces cerevisiae*[26]. Herein, the amine-reactive, MS-cleavable crosslinker disuccinimidyl sulfoxide (DSSO) was leveraged to analyze the *C. reinhardtii* interactome via XL-MS (Figure 6.1). Extraction via freeze-thaw enabled global detection of interacting proteins throughout the cell, including 612 intralinks and 157 interlinks at 1% FDR.

6.2 Materials and Methods

6.2.1 Cell Growth

Wild-type *Chlamydomonas reinhardtii* strain CC-2895 6145c mt- was purchased from the Chlamydomonas Resource Center (St. Paul, MN, USA) and batch cultures were maintained photoheterotrophically on Tris-acetate-phosphate (TAP) agar plates. *C. reinhardtii* was inoculated into 100 mL of TAP medium using a 1 mL inoculum in a foil-covered 250 mL Erlenmeyer flask and grown photoheterotrophically[27]. Cultures were maintained at 22 °C on an Innova 2000 platform shaker (New Brunswick Scientific, Enfield, CT, USA) at 120 rpm under constant 100 μmol m^{-2} s^{-1} illumination. Cells were grown to mid-log phase (OD_{750} 0.4-0.5), harvested by centrifuging for 5 min at 3220 *g* and discarding the supernatant, and flash-frozen using liquid nitrogen. Cell pellets were stored at -80 °C until use.

6.2.2 Protein Extraction

Three frozen cell pellets (~0.6 g each) were thawed on ice before combining and adding 10 mL of 20 mM HEPES buffer, pH 7.8 containing 30 mM sodium chloride, 1.5 mM magnesium chloride, 0.5 mM dithiothreitol, and 1x cOmplete EDTA-free protease inhibitor cocktail (Roche, Basal, Switzerland). Soft lysis via five rounds of freeze-thaw at -80 °C for 60 min was performed to ensure the extraction of intact protein interactions. Cellular debris was cleared by

centrifugation at 3,200 *g* for 20 min at 4 °C before protein concentrations were estimated using the CB-X assay (G-Biosciences, St. Louis, MO, USA) according to the manufacturer's protocol.

6.2.3 Protein Crosslinking

A 50 mM stock solution of disuccinimidyl sulfoxide (DSSO, Thermo Fisher Scientific, Waltham, MA, USA) in DMSO was prepared and added to 3 mg of protein lysate at a working concentration of 2 mM. Samples were incubated with end-over-end rotation for 30 min at RT, before quenching the crosslinking reaction with 20 mM Tris, pH 8 for 30 min at RT.

6.2.4 Protein Digestion

Crosslinked proteins (3 mg) were precipitated with 5x cold acetone and centrifuged at 3200 *g* and 4 °C for 20 min. The supernatant was removed and proteins were resuspended in 200 µL of 100 mM Tris-HCl, pH 8 containing 8 M urea. Samples were reduced and alkylated simultaneously using 10 mM tris(2-carboxyethyl)phosphine and 40 mM chloroacetamide for 1 h at 37 °C. Samples were diluted four-fold to 2 M urea with additional 100 mM Tris-HCl, pH 8 before digestion was performed with Trypsin Gold (Promega, Madison, WI, USA) at 37 °C for 16 h using a protease:protein ratio of 1:50 (w/w).

6.2.5 Reversed-Phase Solid-Phase Extraction

Samples were desalted with 100 mg/1.0 mL Sep-Pak C18 cartridges (Waters, Milford, MA, USA) using a 24-position vacuum manifold (Phenomenex, Torrance, CA, USA) at a flow rate of 1 drop/s. Resin was first pre-eluted using 1 mL of 80% acetonitrile/0.1% trifluoroacetic acid (TFA) before equilibration with 2 mL of 0.1% TFA. Samples were acidified to pH < 3 using 10% TFA, loaded onto the cartridges in two passes, and then washed using 2 mL of 0.1% TFA. Peptides were eluted using 1 mL of 80% acetonitrile/0.1% TFA and concentrated by vacuum

centrifugation. Peptides were resuspended in 200 μL of 10 mM potassium phosphate monobasic, 20% acetonitrile, pH 2.7.

6.2.6 Strong Cation Exchange Fractionation

The crosslinked peptides were fractionated with strong cation exchange (SCX) chromatography using a Shimadzu Prominence HPLC equipped with a UV-vis detector (220 nm) (Shimadzu, Kyoto, Japan) following the method described in Makepeace *et al.*[28]. Mobile phase A consisted of 10 mM potassium phosphate monobasic, 20% acetonitrile, pH 2.7, mobile phase B was 10 mM potassium phosphate monobasic, 250 mM potassium chloride, 20% acetonitrile, pH 2.7, and mobile phase C was 10 mM potassium phosphate monobasic, 600 mM potassium chloride, 20% acetonitrile, pH 2.7. Peptides were fractionated on a PolySulfoethyl A column (100 mm x 4.6 mm, 3 μm particles; PolyLC) using a linear gradient of increasing mobile phases B and C at a flow rate of 0.5 mL/min. After 10 min of 100% mobile phase A, mobile phase B increased from 0% to 15% in 9.3 min, where it was held for 8.7 min before ramping to 30% in 8 min, where it was held for 11 min and then ramping to 100% in 5 min, where it was held for 5 min. After, mobile phase C increased from 0% to 100% in 5 min before returning to 100% mobile phase A in 5 min and re-equilibrating for 25 min. After 10 min into the gradient, fractions were collected every 1 min. These were desalted using reversed-phase solid phase extraction as described above, concentrated under vacuum centrifugation, and resuspended in 15 μL of 0.1% TFA.

6.2.7 LC-MS/MS Analysis

Fractions were analyzed using an Acquity UPLC M-Class System (Waters) coupled to a Q Exactive HF-X mass spectrometer (Thermo Fisher Scientific). Mobile phase A consisted of water with 0.1% formic acid (Thermo Fisher Scientific) and mobile phase B was acetonitrile

with 0.1% formic acid. Injections (4 µL) were made to a Symmetry C18 trap column (100 Å, 5µm, 180µm x 20 mm; Waters) with a flow rate of 5 µL/min for 3 min using 99% A and 1% B. Peptides were then separated on a HSS T3 C18 column (100 Å, 1.8µm, 75µm x 250 mm; Waters) using a linear gradient of increasing mobile phase B at a flow rate of 300 nL/min. Mobile phase B increased from 5% to 35% in 90 min before ramping to 85% in 5 min, where it was held for 10 min before returning to 5% in 2 min and re-equilibrating for 13 min. The mass spectrometer was operated in positive polarity and the Nanospray Flex source had spray voltage floating at 2.1 kV, capillary temperature at 320 °C, and funnel RF level at 40. MS survey scans were collected with a scan range of 350 – 2000 *m/z* at a resolving power of 120,000 and an AGC target of 3×10^6 with a maximum injection time of 50 ms. A top 20 data-dependent acquisition was used where HCD fragmentation of precursor ions having +2 to +7 charge state was performed using a normalized collision energy setting of 28. MS/MS scans were performed at a resolving power of 30,000 and an AGC target of 1×10^5 with a maximum injection time of 100 ms. Dynamic exclusion for precursor *m/z* was set to a 10 s window.

6.2.8 Data Analysis

Acquired spectral files (*.raw) from the 61 fractions were analyzed using Proteome Discoverer v.2.5 with the incorporated XLinkX nodes and searched against the Joint Genome Institute's v.5.6 database (https://phytozome-next.jgi.doe.gov/info/Creinhardtii_v5_6, 19,523 entries, accessed 02/2020) appended with the NCBI chloroplast and mitochondrial databases (chloroplastic-NCBI: BK000554; mitochondrial-NCBI: NC_001638.1; 77 entries, accessed 02/2020) and sequences for common laboratory contaminants (https://www.thegpm.org/cRAP/, 116 entries, accessed 02/2020). For crosslinked peptides, target-decoy searches of MS/MS data used a trypsin protease specificity with the possibility of two missed cleavages, peptide/fragment

mass tolerances of 15 ppm/0.02 Da, fixed modification of cysteine carbamidomethylation, and variable modifications of N-terminus acetylation and methionine oxidation. Identified crosslinks are reported at 1% FDR, controlled at the CSM level using XlinkX and its target-decoy database searching strategy. Detailed settings for Proteome Discoverer nodes are found in Supplemental Information 1.

Initial analysis of protein interlinks was achieved using the STRING database as well as KEGG mapper[29–32]. In the STRING database, active interaction sources were limited to experiments and co-expression with a minimum required interaction score of 0.400. Following STRING and KEGG annotation, the proteins were manually annotated for location and function using a combination of UniProt, PANTHER functional analysis, and the PlaPPIsite (http://zzdlab.com/plappisite/index.php)[33–35].

To generate protein-protein interaction models, an established protocol[36] was applied, where protein structure modeling was achieved through I-TASSER[37] and protein complex docking was performed using HADDOCK[38]. These models were visualized and Euclidean distances of mapped crosslinks were measured using ChimeraX[39–42]. Mapping of detected interlinks was performed using xiNET[43]. XlinkDB 3.0 was used to automatically calculate Euclidean distances of intralinks that were mapped onto known structures or homology models generated by the Integrative Modeling Platform[44–46]. Most intralinks that were mapped were visualized using NGL Viewer[47]. For eukaryotic translation initiation factor 1 alpha, proteins were visualized by mapping to the homology model in SWISS-MODEL, downloading the pdb file, visualizing in Chimera, and manually annotating for detected crosslinks[39,48]. Detected crosslinks from this study were made private on the XLinkDB 3.0 database

(http://xlinkdb.gs.washington.edu/xlinkdb/). These data can be accessed by using the filename "Crcrosslinking_Lesliehicks" in any of the sections labeled "Network Name".

6.2.9 Data Availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository and can be accessed with the dataset identifier PXD026433[49].

6.3 Results and Discussion

6.3.1 Crosslinked Proteome Coverage

Although XL-MS is a powerful tool for the global analysis of PPIs, its success is highly dependent on the permeability and reactivity of the crosslinker. In preliminary trials with DSSO, the crosslinker did not permeate the *Chlamydomonas* cell wall effectively and did not enable global analysis of the interactome (data not shown). Therefore, gentle, detergent-free lysis via freeze-thaw was used to release proteins from the cell in near-native conformations. Intact protein-protein interactions were crosslinked *in vitro* with DSSO and crosslinked peptides were enriched using strong cation exchange (SCX) fractionation prior to LC-MS/MS analysis (Figure 6.1C). SCX leverages the highly positively charged crosslinked peptides to separate them from the less charged linear, non-crosslinked peptides. This is essential as non-crosslinked peptides greatly outnumber and suppress the ionization of the low abundant crosslinked peptides. SCX fractionation simultaneously enriches for crosslinked peptides and decreases sample complexity, thus increasing the depth of coverage for crosslinked peptides.

The analysis of 61 SCX fractions resulted in the identification of 56,595 crosslink reporter ion peaks, yielding 1,705 crosslink-spectrum matches (CSMs, representative CSM can be found

in Figure S1, Supporting Information, 10.1039/D1MO00197C) at 1% FDR grouped into 769 unique crosslinks (Table S1, Supporting Information, 10.1039/D1MO00197C). We also identified a total of 116,086 non-crosslinked peptides from these SCX fractions (Table S2, Supporting Information, 10.1039/D1MO00197C), corresponding to the identification of 7,482 proteins (≥2 unique peptides, master proteins, 1% FDR, Table S3, Supporting Information, 10.1039/D1MO00197C). The full suite of identified crosslinks is depicted as a circos plot in Figure 6.2A, while the identified *C. reinhardtii* interactome is displayed in Figure 6.2B (interactive versions of both are located at https://tinyurl.com/HicksXLMS). The majority of the CSMs are contained in later SCX fractions (49-65), supporting the use of SCX as an enrichment technique for more positively charged crosslinked peptides (Figure 6.2C). Among the 769 detected crosslinks, 157 are between two tryptic peptides derived from two different proteins (interlinks) and 612 are between two tryptic peptides within the same protein (intralinks) (Figure 6.2D). A majority of the detected proteins with at least one identified crosslink contain one crosslink (63%), while 13% contain two crosslinks and 24% contain three or more (Figure 6.2E). Additionally, of the 157 identified interlinks, only 12 are currently reported in *C. reinhardtii* in the plant PPI database (Table S4, Supporting Information, 10.1039/D1MO00197C); therefore 92% of the crosslinks observed herein represent the first empirical evidence of interactions between the proteins[35].

6.3.2 Intralinks

Overlaying detected intralinks onto known protein structures and comparing the Euclidean distance between the crosslinked residues to the DSSO maximum crosslinking distance of 30 Å can be used to evaluate the intralink dataset[13,50–52]. Few proteins with identified intralinks have existing structures for *C. reinhardtii* in the Protein Data Bank (PDB), so the Integrative

Modeling Platform in XLinkDB 3.0 was used for homology modeling for the other proteins[44,46]. In this platform, a structural homology model is identified by multiple sequence alignment and used to predict protein structure. Out of all detected intralinks, 44% were mapped onto known structures or homologous proteins to obtain structural information (Table S5, Supporting Information, 10.1039/D1MO00197C). Of these, 96% featured residues within the theoretical distance of DSSO (Figure 6.3A); this is not surprising as the majority of the observed proteins are highly conserved across taxons and thus have similar structural features in distinct species.

The crosslinking data for Photosystem II Oxygen Evolution Enhancer protein 3 (Cre08.g372450.t1.2) confirms the structure established via cryogenic electron microscopy[53]. Of the 10 intralinks observed across this protein, none exceeded the 30 Å cutoff, with an average predicted link distance of 12 Å. However, there are relatively few proteins from *C. reinhardtii* with structural data in PDB; therefore, many of the observed intralinks were mapped to proteins in other organisms. For example, 6 intralinks derived from phosphoglycerate kinase (Cre11.g467770.t1.1) were mapped onto phosphoglycerate kinase from *Thermotoga maritima* (PDB ID: 1VPE), a thermophilic bacterium, and the predicted Euclidean distances were compared with the maximum distance enabled by DSSO. All of the crosslinks on *T. maritima* phosphoglycerate kinase contained residues < 30 Å apart, thus indicating that the DSSO crosslinked peptides from *C. reinhardtii* identified in this study conform to the structural predictions of the known protein structure from *T. maritima* (Figure 6.3B).

The lack of structural data for *C. reinhardtii* proteins creates challenges when using crosslinking to confirm tertiary structure. For example, eukaryotic translation elongation factor 1 alpha (Cre12.g498600.t1.2) had the second highest abundance of intralinks with 18 total, 14 of which were mapped within the distance restraints for DSSO (Figure 6.4). However, there are no

structures from *C. reinhardtii* in PDB and the crosslinks from *C. reinhardtii* could not be mapped to a singular homologous protein; rather, the *C. reinhardtii* sequence was mapped to proteins with high sequence similarity in *Homo sapiens, Oryctolagus cuniculus, Pyrococcus horikoshii,* and *Aeropyrum pernix* (Figure 6.4A). Additionally, three intralinks (K358-K293, K387-K413, K293-K230) were not able to be mapped to proteins with known structures, likely due to differences in the primary sequences in the protein models when compared to that of *C. reinhardtii* (Figure 6.4B). Therefore, while the average predicted crosslink distance for eukaryotic translation elongation factor 1 alpha was 17 Å (Figure 6.4C), well within the distance restraint of DSSO, complementary experiments would need to be conducted to fully understand how unique the protein structure is in *C. reinhardtii*.

Intralinks that include residues spaced above 30 Å could indicate differences from the homologous structure and/or false positive identifications. The high percentage of intralinks that are within the 30 Å distance limit provides high confidence in the dataset, therefore suggesting that those outside the limit are potential distinctions from the homologous protein structures. One notable example is ribosomal protein S5/elongation factor G/III/V family protein (Cre12.g516200.t1.2), from which 10 intralinks were observed. Of those intralinks, seven had predicted distances less than 30 Å. However, one intralink was not mapped to a known structure and two were outside the DSSO distance restraints; K498-K429 had a predicted distance of 40 Å while K498-K484 had a predicted distance of 36 Å, thus suggesting deviations from the mapped protein structures. Like eukaryotic translation elongation factor 1 alpha, this protein required several organisms' known protein structures to map the crosslinks, with both intralinks that surpassed the DSSO restraint being mapped to proteins from *H. sapiens*. However, further work

will need to be conducted to determine the extent to which the structure in *C. reinhardtii* differs from that of other known models.

Experimentally determined intralinks can be used as constraints to guide computational modeling for novel protein structures. This approach was implemented to predict the structure of elongation factor 3 and ABC transporter (Cre04.g222700.t1.2) by incorporating 11 identified intralinks (5 were not mapped onto a known structure or homology model) to computational model its structure using the iterative threading assembly refinement (I-TASSER) server[37]. From this, the top five structure models were output from I-TASSER (Figure S2, Supporting Information, 10.1039/D1MO00197C). The top template for these models derived from elongation factor 3A in *S. cerevisiae* (PDB: 2IWH), suggesting that Cre04.g222700.t1.2 is structurally similar to this protein. To test the agreement between these structures and the crosslink data, the detected intralinks were mapped onto the models and the distances between the crosslinked residues were calculated. In each case, all crosslinks had measured distances within the maximum restraint for DSSO, providing confidence in the modeling. These data demonstrate the value of these identified intralinks in generating novel protein structures.

6.3.3 Cytosolic Interlinks

Overall, 76 cytosolic interlinked protein pairs were identified, including 83 unique proteins (Figure 6.2B, Table S1, Supporting Information, 10.1039/D1MO00197C). Of these interlinked protein pairs, 63 included at least one protein from the cytosolic ribosome, reflecting the known high abundance of this large protein complex. KeggMapper revealed that 42% of the small ribosomal subunit and 38% of the large ribosomal subunit were identified in this study (Figure 6.5)[30]. This includes 34 proteins with at least one detected interlink and 29 proteins with at least one detected intralink, thus showing remarkable coverage of the ribosome from both a

protein-protein relationship perspective as well as a structural perspective, without any attempt at ribosomal enrichment. There were 84 intralinks identified on cytosolic ribosomal subunits, of which 67 were successfully mapped to a homology model and analyzed for Euclidian distance. Of these measurements, only three intralinks were determined to be greater than the maximum distance allowed by DSSO. Crosslinks between K498-K429 and K498-K484 from ribosomal protein S5 (Cre12.g516200.t1.2) had distances of 40 and 36 Å, respectively, and the crosslink between K209-K183 from cytosolic 80S ribosomal protein L8 (Cre12.g535851.t1.1) had a distance of 38 Å. This likely represents small changes in the structure of the subunits compared to the homology models, though further analysis is needed to fully characterize any deviations.

The extensive coverage of the ribosome indicates that crosslinking could be used to profile ribosomal changes in *C. reinhardtii,* which would be advantageous when combined with quantitative analysis for mapping changes to post-translational modifications and/or determining structural dynamics under stress, particularly as these dynamics are known to regulate substrate interaction and biological activity of protein translation[54–57]. Previous work leveraged MS to identify the order of subunit assembly along the ribosomal stalk proteins in *E. coli*[58], but cannot reveal structural changes occurring of well documented large-scale movements following substrate binding[59,60]. In contrast, solution-state NMR has been leveraged to understand the structure and motion of the ribosome in *E. coli*, but lacks the ability to identify the subunits responsible for this flexibility[61]. Recent work paired heat treatment with MS and bioinformatic analysis to identify 17 intrinsically disordered proteins within the cytosolic ribosome structure in *C. reinhardtii*; however, their contribution(s) to the dynamic ribosome structure and overall flexibility is currently unknown[62]. XL-MS enables simultaneous analysis of both the identification of subunits as well as their structural proximity, which could prove critical in

understanding ribosomal regulation in *C. reinhardtii*. Further enrichment of ribosomes paired with quantitative XL-MS may unveil how these proteins' flexibility contribute to ribosomal regulation.

Although the majority of detected interlinks related to the ribosome, several other identified interactions point to the existence of novel regulatory points in essential metabolic processes. Additionally, their interlink partners can help confirm or dismantle previous localization predictions. For example, glutaredoxin 6 (Cre01.g047800.t1.1) (Table S1, row 705, Supporting Information, 10.1039/D1MO00197C), which enzymatically deglutathionylates cysteine residues, has been predicted to be localized to the chloroplast despite being encoded in the nuclear genome[63–65]. However, here it was observed in an interlink with a cytosolic nucleotide-binding protein (Cre16.g685200.t1.1), thus contradicting the prediction that glutaredoxin 6 is localized to the chloroplast. Although validation is required, localization of glutatredoxin 6 in the cytosol could indicate a more substantial substrate list than previously predicted. Understanding the spectrum of glutaredoxin substrates is essential to understanding the impact of intracellular oxidative signaling and how this can be used as a regulatory mechanism in diverse cell processes.

Another intriguing result is the crosslink identified between starch phosphorylase (Cre07.g336950.t1.1) and a previously unannotated protein (Cre02.g142206.t1.1) (Table S1, row 742, Supporting Information, 10.1039/D1MO00197C) unique to *C. reinhardtii.* Conducting a protein BLAST search of this unidentified protein against the proteome of *A. thaliana* did not reveal any sequence similarity, but did result in sequence similarities of >50% to uncharacterized proteins in five green algae: *Volvox carteri f. nagariensis*, *Gonium pectoral*, *Haematococcus lacustris*, *Scenedesmus sp. NREL 46B-D3*, and *Polytomella parva* (Figure 6.6). Additionally,

searching the primary sequence against the NCBI conserved domain database (CDD) distinguished amino acids 66 – 173 as an oxidoreductase containing a GGXGXXG cofactor binding motif; this protein domain is present in many proteins related to the metabolism of steroids, cofactors, carbohydrates, lipids, aromatic compounds, and amino acids, as well as function in redox sensing, though it is not present in the five aligned, uncharacterized proteins with >50% sequence similarity[66]. Together, this data suggests this unannotated protein may be a unique regulatory protein along the starch biosynthesis pathway in *C. reinhardtii*, though its function needs to be further investigated.

Similarly, we identified a novel connection between inositol hexakisphosphate kinase (Cre01.g052650.t1.1) and sulfatase-domain containing protein (Cre01.g012126.t1.2) (Table S1, row 768, Supporting Information, 10.1039/D1MO00197C). Hexakisphosphate kinase catalyzes the conversion of hexakisphosphate ($InsP_6$) to disphosphoinositol pentakisphosphate ($InsP_7$). It is also an essential regulatory component of the target of rapamycin (TOR) pathway in *C. reinhardtii*, through which it engages in crosstalk with the phosphorylation network to modulate carbon metabolism and photosynthesis[67]. In *C. reinhardtii*, hexakisphosphate kinase is impacted by phosphorous deprivation, thus establishing the enzyme as a regulatory point in intracellular nutritional sensing. However, its interaction with a sulfatase-domain containing protein could suggest that it is also involved in sulfur recycling and/or sensing— not surprising due to TOR's longstanding association with many forms of nutritional deprivation[22,24,68,69]. Further experiments should explore the role of InsPs during sulfur deficiency to better understand this regulatory mechanism.

6.3.4 Chloroplast Interlinks

As the chloroplast typically comprises >40% of the cell's volume, it was not surprising that an abundance of crosslinked peptides were identified from chloroplast proteins[18]. In total, we identified 43 unique proteins involved in 35 interlinked peptide pairs, which were analyzed for known relationships using the STRING database (Figure 6.7)[29]. Several interlinks aligned with well-established protein relationships, such as between glyceraldehyde-3-phosphate dehydrogenase (GAPDH, Cre01.g010900.t1.2) and phosphoribulokinase (PRK, Cre12.g554800.t1.2) (Table S1, row 738, Supporting Information, 10.1039/D1MO00197C), which together form a complex that controls substrate availability for RuBisCO[70]. Previous work has used fluorescence correlation spectroscopy to distinguish the existence of the PRK-GAPDH complex, here confirmed via XL-MS[71]. GAPDH has also been shown to complex with CP12, an intrinsically disordered protein, to successfully recruit and complex with PRK, an essential regulatory step in the operation of the Calvin-Benson cycle[72,73]. However, CP12 was not identified among crosslinked peptides in this study. This could indicate that CP12's lysines are solvent inaccessible, or more likely that the protein was too low in abundance to identify in this dataset. The latter is suggested by recent X-ray diffraction work in *A. thaliana* that analyzed the GAPDH-CP12-PRK complex where the observed lysines were solvent accessible therein. This underscores the need for further targeted analysis via XL-MS and/or high resolution microscopy to delineate the complex structure in *C. reinhardtii*[74].

XL-MS also revealed a known relationship between photosystem II repair protein (Cre10.g430150.t1.2, REP27) and chloroplast ribosomal protein L11 (Cre10.g423650.t1.2, RPL11), detecting multiple crosslinks between REP27 at position K351 and RPL11 at positions K126, K129, and K146 (Table S1, rows 473, 474, and 598, Supporting Information,

10.1039/D1MO00197C). REP27 is a tetratricopeptide repeat protein that is encoded by the nuclear genome but localized to the chloroplast and is essential for the selective removal and replacement of photodamaged D1 proteins of photosystem II[75]. Although the structure of REP27 has not been elucidated via biochemical techniques, identification of two transmembrane domains facilitated the generation of a folding model of REP27, which predicts a single loop localized to the lumen due to transmembrane helices while both the N-terminus and the C-terminus are localized to the chloroplast stroma[76]. The highly charged C-terminus is essential for mRNA translation initiation and assembly of D1, with *rep27* knockouts possessing few intact D1 proteins when analyzed using western blotting. The data herein adds additional evidence to this mechanism, with the first empirical evidence of a direct linkage between REP27 and RPL1l that likely represents a docking point for the ribosome to the photosystem II repair complex during D1 repair and insertion[77].

While known protein interactions lend confidence to the acquired dataset, unique crosslinks presenting previously undescribed interactions are of particular interest for further examination. For example, a novel interaction was observed between GAPDH and an FKBP-type cis-trans isomerase (Cre10.g466850.t1.1) that facilitates protein disulfide bonds and is essential for the regulation/proliferation of oxidative signaling (Table S1, row 613, Supporting Information, 10.1039/D1MO00197C)[78,79]. GAPDH is highly redox regulated and we have previously observed reversible oxidation on C190 that remains unchanged following nitrogen deprivation[24]. While GAPDH has several documented regulatory redox partners, including thioredoxins[80–82], it has not previously been connected to FKBP-type cis-trans isomerase. This interaction could therefore facilitate the formation of regulatory disulfide bonds on GAPDH.

XL-MS also revealed Mg-chelatase subunit 1 (Cre06.g306300.t1.2), the first-committed enzyme of chlorophyll biosynthesis, to be a hub for protein interactions within the chloroplast. Mg-chelatase was shown to interact with flagellar associated protein 165 (Cre03.g211521.t1.1), a transcription initiation factor component of the TAF4 family (Cre02.g095800.t1.2), DNA polymerase theta (Cre08.g384390.t1.1), an IgA-specific serine endopeptidase (Cre09.g408350.t1.2), and two previously unannotated (predicted) *C. reinhardtii* proteins (Cre11.g477733.t1.2, Cre06.g306300.t1.2). Multiple interlinks were identified between one of the predicted proteins (Cre11.g477733.t1.2) at position K19 and Mg-chelatase at positions K70, K116, K157, K167, K198, K207, and K308 (Figure 6.8A, 6.8B). This small 14 kDa protein, referred herein as Mg-chelatase associated protein (MCAP), does not share sequence identity with any proteins in *A. thaliana*, suggesting that it may be a novel protein for the regulation of chlorophyll biosynthesis in green algae. While it shares 62.7% sequence overlap with a protein from the green alga *Gonium pectoral* and 59% overlap with a protein from the green alga *Chlamydomonas incerta*, these proteins also lack annotations.

To generate a protein-protein interaction model for Mg-chelatase and MCAP, an established protocol was applied, which employs I-TASSER for protein structure modeling and High Ambiguity Driven protein-protein Docking (HADDOCK) for protein complex docking[36–38]. The 9 detected intralinks for Mg-chelatase were incorporated in its I-TASSER structure modelling (Figure S3, Supporting Information, 10.1039/D1MO00197C) and the top structures for Mg-chelatase and MCAP were exported for interaction docking. All detected intralinks from Mg-chelatase were mapped onto its top structure and had measured distances within the maximum restraint for DSSO, thereby indicating that model refinement was unnecessary. In contrast, modeling MCAP was more complicated; all the modeled structures had low C-scores (-5 to -4,

scored on a range of -5–2, where 2 is best), where the top templates included a putative uncharacterized metacyclic invariant surface protein from *Trypanosoma brucei* (PDB: 5VTL) and *Mycobacterium tuberculosis* quinolinate phosphoribosyl transferase (PDB: 1QPN). This is likely the result of low sequence similarity between MCAP and proteins with known structures. The identified interlinks between Mg-chelatase and MCAP were incorporated into the preliminary complex modeling, top structures for each cluster were exported, and detected interlinks were mapped onto these structures to determine the alignment between the models and the experimental crosslink data (Figure S4, Supporting Information, 10.1039/D1MO00197C). The best overlap between one of the models and the identified interlinks resulted in 3 of 7 crosslinks (K167-K19, K207-K19, K308-K19) having measured distances within the maximum restraint for DSSO. This poor overlap between the detected interlinks and the protein complex docking could be attributed to the poor protein structure modeling obtained for MCAP. A refined complex model was created by repeating the protein complex docking and only incorporating the compatible interlinks (Figure 6.8C, S5, Supporting Information, 10.1039/D1MO00197C), and the same 3 of 7 crosslinks had measured distances within the maximum restraint for DSSO.

MCAP is likely low in abundance as it has not been detected in our previous global proteomics work in *C. reinhardtii*, whereas Mg-chelatase was detected in all studies[24,83,84]. This suggests that MCAP and Mg-chelatase may be present in different stoichiometric ratios. Chlorophyll synthesis is essential for ensuring efficient light capture and energy transfer, and investigating the function of MCAP should be a priority in order to validate the reported interaction results, and, if this interaction is confirmed, understand how this biosynthetic pathway differs in comparison to other phototrophs.

6.4 Figures

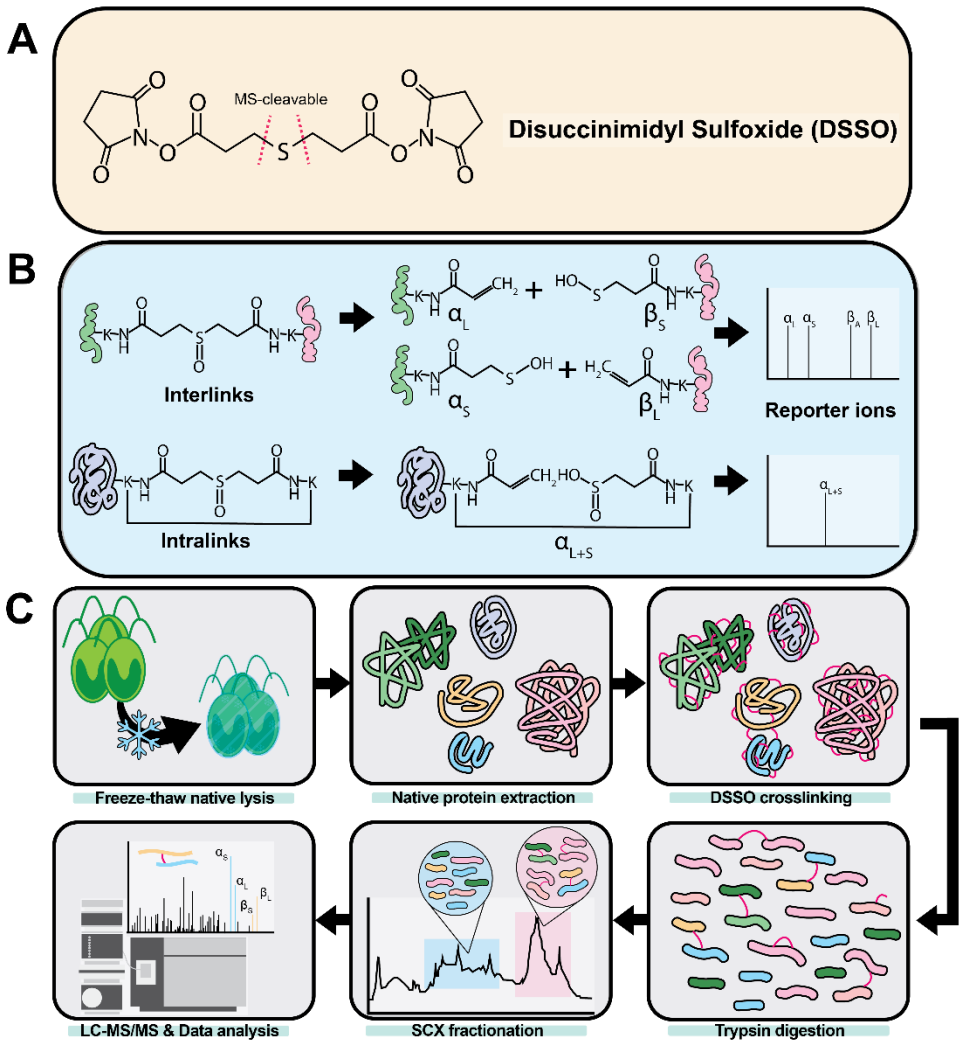


Figure 6.1 General workflow for *C. reinhardtii* interactomics. **(A)** The MS-cleavable crosslinker DSSO was used in this study. **(B)** DSSO generates reporter ions that enable delineation between crosslinks. These reporter ions are essential for minimizing false discovery rate in identifications. **(C)** Native protein structures and interactions were preserved and extracted using a gentle lysis via freeze-thaw. Proteins were crosslinked with DSSO and proteolyzed using trypsin. Crosslinked peptides were enriched and fractionated to increase depth of coverage for crosslink identifications. After LC-MS/MS analysis, the XLinkX nodes in Proteome Discoverer were implemented to recognize crosslinks by characteristic crosslink reporter ions (α_S, α_L, β_S, and β_L) in the MS2 spectra and identify crosslink-spectrum matches with database searching using the fragment ions from the MS2 spectra.

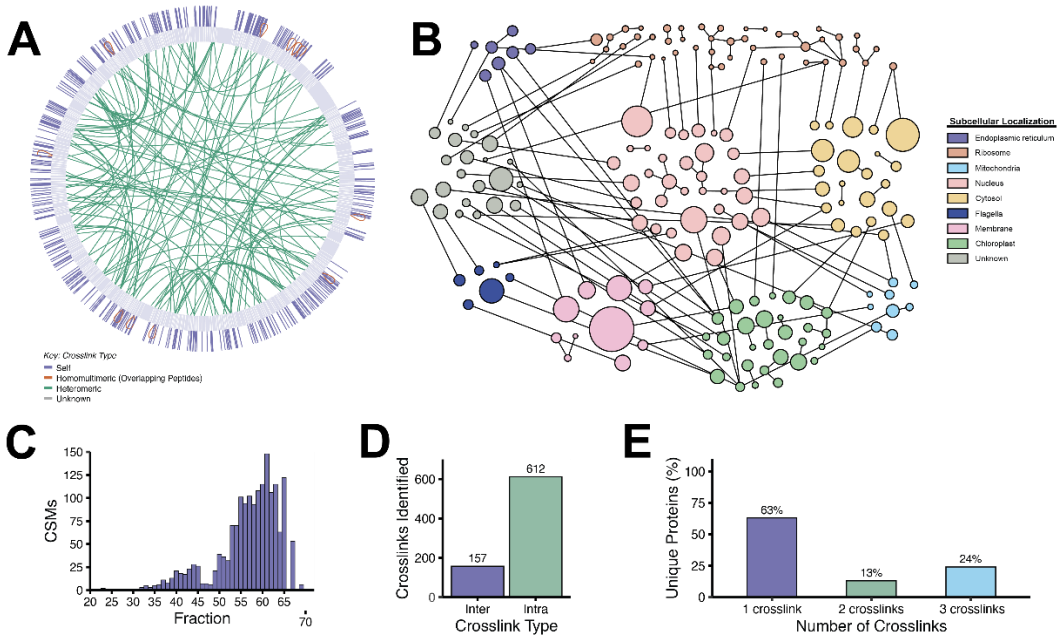


Figure 6.2 Description of identified *C. reinhardtii* crosslinked peptides following DSSO crosslinking of protein extracts. **(A)** Circos plot of all identified crosslinks. Inner, green curves represent interlinks between two different proteins, outer, orange curves represent interlinks within a homoligomer, and outer, purple curves represent intralinks within the same protein. **(B)** Protein interactome map featuring identified interlinks. Proteins are grouped by subcellular localization. **(C)** Histogram of the number of identified CSMs across each SCX fraction. **(D)** Types of detected crosslinks. **(E)** Distribution of identified unique proteins by number of detected crosslinks.

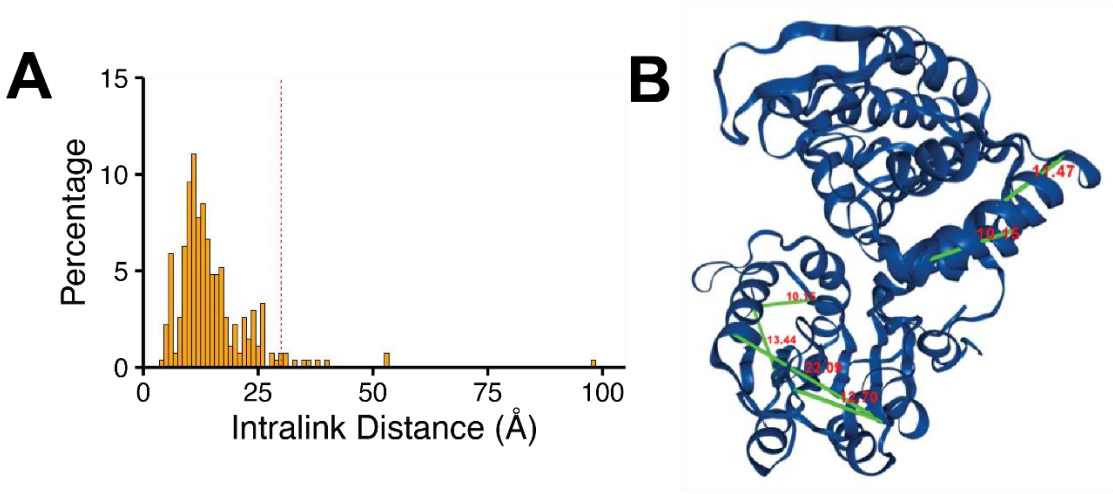
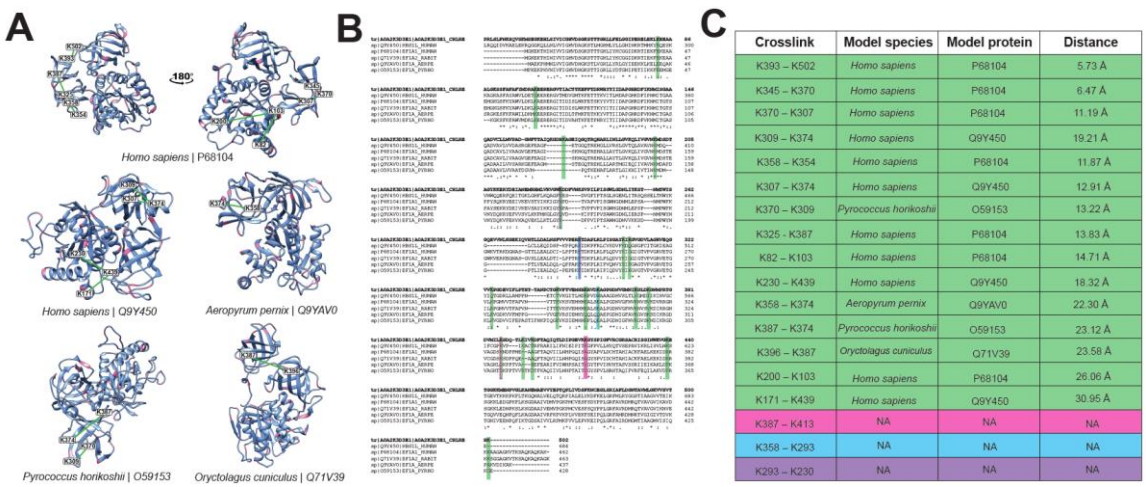


Figure 6.3 Intralinks were mapped onto known protein structures or homology models using the Integrative Modeling Platform in XLinkDB 3.0. **(A)** Histogram showing Euclidean distances between residues of detected intralinks. The red, dashed line represents the DSSO maximum crosslinking distance of 30 Å. Of intralinks that were able to be mapped, 96% featured residues within the theoretical distance of DSSO. **(B)** The 6 intralinks (in green) identified from *C. reinhardtii* phosphoglycerate kinase that mapped onto the structure of phosphoglycerate kinase from *T. maritima* (PDB ID: 1VPE). The Euclidean distances between the residues in each intralinks (in red) are less than 30 Å, the maximum crosslinking distance for DSSO.



Crosslink	Model species	Model protein	Distance
K393 – K502	*Homo sapiens*	P68104	5.73 Å
K345 – K370	*Homo sapiens*	P68104	6.47 Å
K370 – K307	*Homo sapiens*	P68104	11.19 Å
K309 – K374	*Homo sapiens*	Q9Y450	19.21 Å
K358 – K354	*Homo sapiens*	P68104	11.87 Å
K307 – K374	*Homo sapiens*	Q9Y450	12.91 Å
K370 – K309	*Pyrococcus horikoshii*	O59153	13.22 Å
K325 - K387	*Homo sapiens*	P68104	13.83 Å
K82 – K103	*Homo sapiens*	P68104	14.71 Å
K230 – K439	*Homo sapiens*	Q9Y450	18.32 Å
K358 – K374	*Aeropyrum pernix*	Q9YAV0	22.30 Å
K387 – K374	*Pyrococcus horikoshii*	O59153	23.12 Å
K396 – K387	*Oryctolagus cuniculus*	Q71V39	23.58 Å
K200 – K103	*Homo sapiens*	P68104	26.06 Å
K171 – K439	*Homo sapiens*	Q9Y450	30.95 Å
K387 – K413	NA	NA	NA
K358 – K293	NA	NA	NA
K293 – K230	NA	NA	NA

Figure 6.4 Intralinks of eukaryotic translation elongation factor 1 alpha (Cre12.g498600.t1.2) were mapped onto homology models using the Integrative Modeling Platform in XLinkDB 3.0 to measure theoretical distances between identified intralinks. **(A)** Visualization of structure maps used for distance determination, conducted via SWISS-MODEL. Lysine residues are shown in pink and crosslinks are shown via green lines. **(B)** Sequence alignment (conducted via Clustal Omega) of the five proteins used to determine the distances between crosslinked residues in eukaryotic translation elongation factor 1 alpha. Green highlight indicates residues that were successfully mapped to homology models, while pink, blue, and purple highlights indicate crosslinks (shown in C) that were not mapped to homology models. **(C)** Table of the intralinks observed on eukaryotic translation elongation factor 1 alpha, the species and protein accession the primary sequence was mapped to, and the distance between bound lysines according to each model.

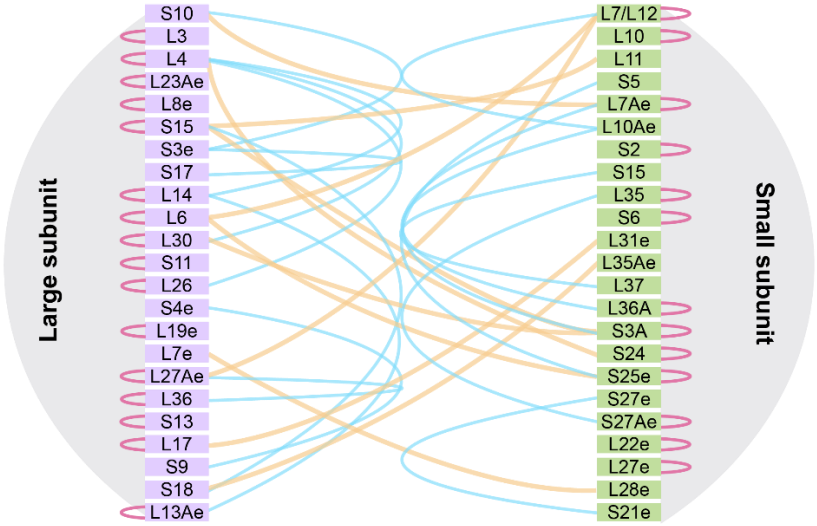


Figure 6.5 Visualization of all crosslinked subunits of the cytosolic ribosome. Intralinks are visualized in pink while interlinks within the large or small ribosomal subunits are visualized in blue. The orange interlinks represent interactions between proteins found in opposite subunits.

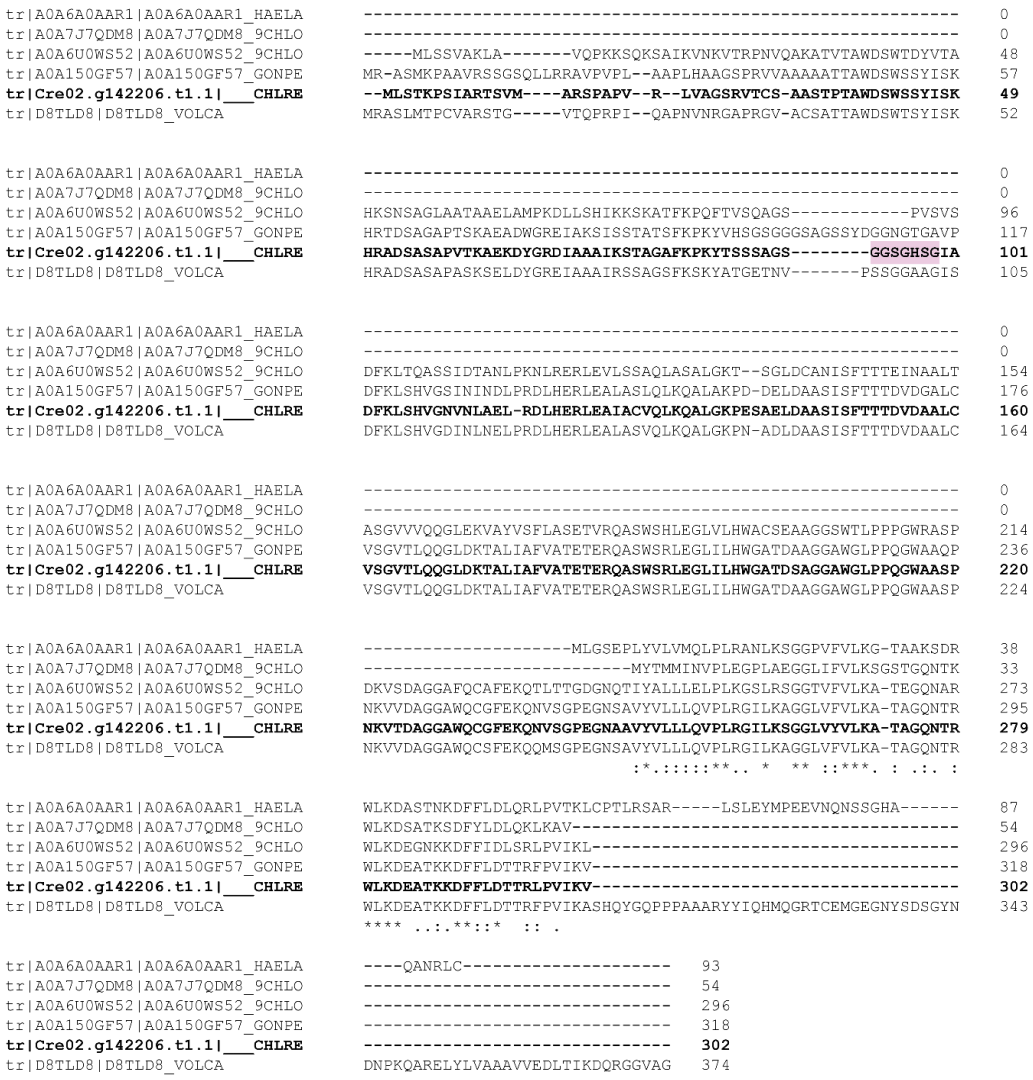


Figure 6.6 Sequence alignment between Cre02.g142206.t1.1, an uncharacterized protein observed to be crosslinked to starch phosphorylase (Cre07.g336950.t1.1), and the five proteins with sequence similarity >50%. Organisms include *Volvox carteri f. nagariensis*, *Gonium pectoral*, *Haematococcus lacustris*, *Scenedesmus sp. NREL 46B-D3*, and *Polytomella parva.* The oxidoreductase domain is highlighted in pink.

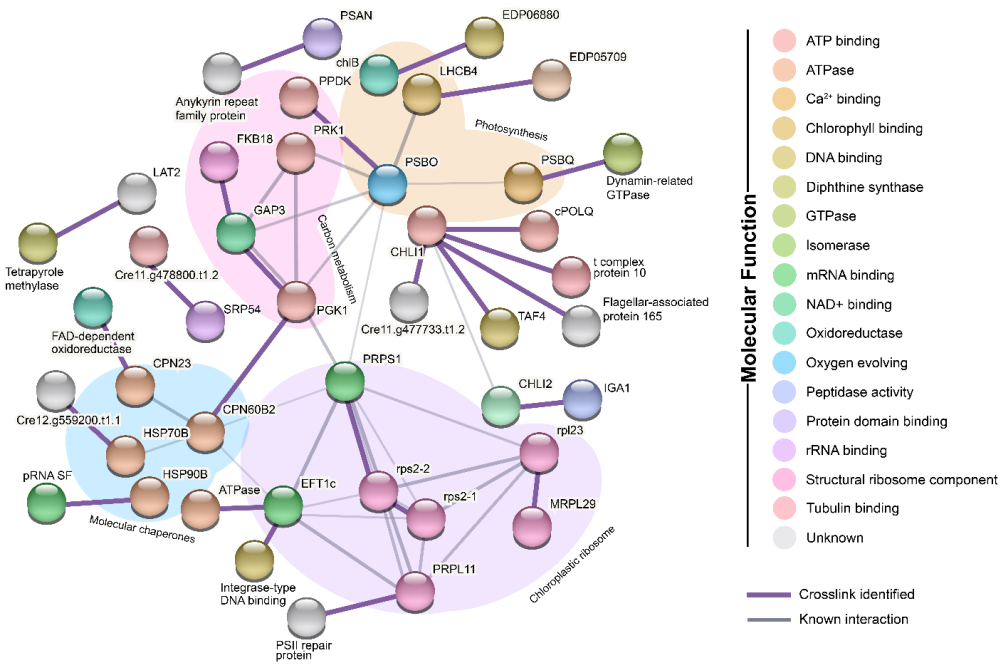


Figure 6.7 Protein interactome map featuring identified interlinks localized to the chloroplast. Proteins are color coded by molecular function and clustered biological processes are shaded and labeled. Purple lines denote observed interlinks in this study while grey lines indicate empirical evidence of interactions as compiled by the STRING database, where increasing line thickness indicates increased confidence in interaction.

A

```
        10         20         30         40         50
MAVAHVRAAA EHRRAEEVKA AEGRRAEGRR ALSIDFRTCA IILGIFLVGH
        60         70         80         90        100
PLRELFNKCC AALDAASRKV VGPIQTFIVD QAAPIAGGLT AFWIVARDWS
       110        120        130
PLLGSERRRL PRPCLALENG AINHCASRAC SF
```

B

Mg-chelatase site	MCAP site	Euclidean distance (Å)	Mg-chelatase sequence	MCAP sequence	Max XlinkX Score	# CSMs
K70	K19	46.2	TE[K]ELGQARPIFPFTAIVGQDEMK	RAEEV[K]AAEGR	122.98	10
K116	K19	33.1	GTG[K]STTIR	RAEEV[K]AAEGR	72.62	1
K157	K19	30.1	V[K]AGEQLPVSSK	RAEEV[K]AAEGR	82.39	1
K167	K19	18.2	AGEQLPVSS[K]K	RAEEV[K]AAEGR	101.21	2
K198	K19	36.7	ALTEGV[K]AFEPGLLAK	RAEEV[K]AAEGR	88.64	1
K207	K19	26.2	AFEPGLLA[K]ANR	RAEEV[K]AAEGR	94.4	6
K308	K19	24.9	STFDENPAAFR[K]DYEAGQMALTQR	RAEEV[K]AAEGR	114.59	3

C

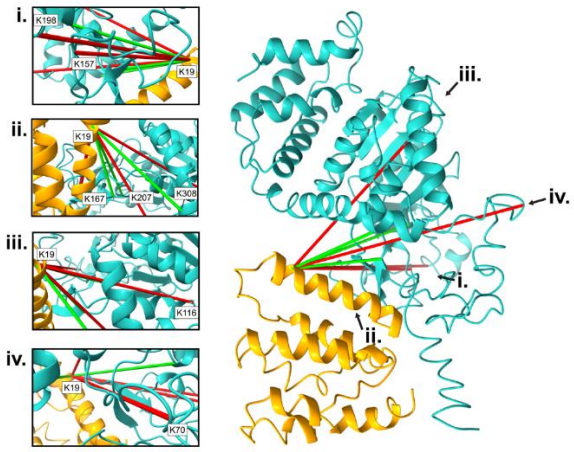


Figure 6.8 Crosslinking unveiled a novel complex between Mg-chelatase (Cre06.g306300.t1.2) and uncharacterized protein Cre11.g477733.t1.2, herein referred to as Mg-chelatase associated protein, or MCAP. **(A)** Sequence of MCAP, which is uncharacterized on UniProt. The crosslinked residue is shown in pink. **(B)** Table of observed crosslinks between Mg-chelatase and MCAP. Crosslinked resides are shown in brackets. Euclidean distance refers to that interlink mapped onto the refined protein complex shown in Figure 6.8C. **(C)** I-TASSER and HADDOCK were used to generate a protein-protein interaction model between Mg-chelatase (teal) and MCAP (orange). The structure with the most mapped interlinks within the maximum restraint for DSSO (in green) from the refined complex modeling is shown. The boxes labeled with roman numerals show the crosslinked lysine residues from the perspective of the arrows on the model complex structure.

REFERENCES

1. Leitner, A.; Faini, M.; Stengel, F.; Aebersold, R. Crosslinking and Mass Spectrometry: An Integrated Technology to Understand the Structure and Function of Molecular Machines. *Trends Biochem. Sci.* **2016**, *41*(1), 20–32.

2. Leitner, A.; Walzthoeni, T.; Kahraman, A.; Herzog, F.; Rinner, O.; Beck, M.; Aebersold, R. Probing native protein structures by chemical cross-linking, mass spectrometry, and bioinformatics. *Mol. Cell. Proteomics* **2010**, *9*(8), 1634–1649.

3. Sinz, A. The advancement of chemical cross-linking and mass spectrometry for structural proteomics: from single proteins to protein interaction networks. *Expert Rev. Proteomics* **2014**, *11*(6), 733–743.

4. Iacobucci, C.; Götze, M.; Sinz, A. Cross-linking/mass spectrometry to get a closer view on protein interaction networks. *Curr. Opin. Biotechnol.* **2020**, *63*, 48–53.

5. Liu, F.; Heck, A. J. R. Interrogating the architecture of protein assemblies and protein interaction networks by cross-linking mass spectrometry. *Curr. Opin. Struct. Biol.* **2015**, *35*, 100–108.

6. Smythers, A. L.; Hicks, L. M. Mapping the plant proteome: tools for surveying coordinating pathways. *Emerg. Top. Life Sci.* **2021**, *5*(2), 203.

7. Kaur, U.; Meng, H.; Lui, F.; Ma, R.; Ogburn, R. N.; Johnson, J. H. R.; Fitzgerald, M. C.; Jones, L. M. Proteome-Wide Structural Biology: An Emerging Field for the Structural Analysis of Proteins on the Proteomic Scale. *J. Proteome Res.* **2018**, *17*(11), 3614–3627.

8. Paiano, A.; Margiotta, A.; Luca, M. De; Bucci, C. Yeast Two-Hybrid Assay to Identify Interacting Proteins. *Curr. Protoc. Protein Sci.* **2019**, *95*(1), e70.

9. Peden, E. A.; Boehm, M.; Mulder, D. W.; Davis, R.; Old, W. M.; King, P. W.; Ghirardi, M. L.; Dubini, A. Identification of Global Ferredoxin Interaction Networks in *Chlamydomonas reinhardtii*. *J. Biol. Chem.* **2013**, *288*(49), 35192.

10. Mackinder, L. C. M. The *Chlamydomonas* CO2-concentrating mechanism and its potential for engineering photosynthesis in plants. *New Phytol.* **2018**, *217*(1), 54–61.

11. Schneider, M.; Belsom, A.; Rappsilber, J. Protein Tertiary Structure by Crosslinking/Mass Spectrometry. *Trends Biochem. Sci.* **2018**, *43*(3), 157.

12. O'Reilly, F. J.; Rappsilber, J. Cross-linking mass spectrometry: methods and applications in structural, molecular and systems biology. *Nat. Struct. Mol. Biol.* **2018**, *25*(11), 1000–1008.

13. Merkley, E. D.; Rysavy, S.; Kahraman, A.; Hafen, R. P.; Daggett, V.; Adkins, J. N. Distance restraints from crosslinking mass spectrometry: Mining a molecular dynamics simulation database to evaluate lysine–lysine distances. *Protein Sci.* **2014**, *23*(6), 747–

759.

14. Matzinger, M.; Mechtler, K. Cleavable Cross-Linkers and Mass Spectrometry for the Ultimate Task of Profiling Protein–Protein Interaction Networks *in Vivo*. *J. Proteome Res.* **2020**, *20*(1), 78–93.

15. Mendes, M. L.; Fischer, L.; Chen, Z. A.; Barbon, M.; O'Reilly, F. J.; Giese, S. H.; Bohlke-Schneider, M.; Belsom, A.; Dau, T.; Combe, C. W.; Graham, M.; Eisele, M. R.; Baumeister, W.; Speck, C.; Rappsilber, J. An integrated workflow for crosslinking mass spectrometry. *Mol. Syst. Biol.* **2019**, *15*(9), e8994.

16. Götze, M.; Iacobucci, C.; Ihling, C. H.; Sinz, A. A Simple Cross-Linking/Mass Spectrometry Workflow for Studying System-wide Protein Interactions. *Anal. Chem.* **2019**, *91*(15), 10236–10244.

17. Sinz, A. Divide and conquer: cleavable cross-linkers to study protein conformation and protein-protein interactions. *Anal. Bioanal. Chem.* **2017**, *409*(1), 33–44.

18. Harris, E. The Chlamydomonas Sourcebook: Introduction into Chlamydomonas and its laboratory use. *Elsevier Acad. Press. San Diego, CA* **2008**, 159–210.

19. Goodstein, D. M.; Shu, S.; Howson, R.; Neupane, R.; Hayes, R. D.; Fazo, J.; Mitros, T.; Dirks, W.; Hellsten, U.; Putnam, N.; Rokhsar, D. S. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* **2012**, *40*(D1), D1178–D1186.

20. Pérez-Pérez, M.; Couso, I.; Crespo, J. The TOR Signaling Network in the Model Unicellular Green Alga *Chlamydomonas reinhardtii*. *Biomolecules* **2017**, *7*(3), 54.

21. Pérez-Pérez, M. E.; Couso, I.; Heredia-Martínez, L. G.; Crespo, J. L. Monitoring Autophagy in the Model Green Microalga *Chlamydomonas reinhardtii*. *Cells* **2017**, *6*(4), 36.

22. Park, J. J.; Wang, H.; Gargouri, M.; Deshpande, R. R.; Skepper, J. N.; Holguin, F. O.; Juergens, M. T.; Shachar-Hill, Y.; Hicks, L. M.; Gang, D. R. The response of *Chlamydomonas reinhardtii* to nitrogen deprivation: a systems biology analysis. *Plant J.* **2015**, *81*(4), 611–624.

23. Juergens, M. T.; Deshpande, R. R.; Lucker, B. F.; Park, J.-J.; Wang, H.; Gargouri, M.; Holguin, F. O.; Disbrow, B.; Schaub, T.; Skepper, J. N.; Kramer, D. M.; Gang, D. R.; Hicks, L. M.; Shachar-Hill, Y. The Regulation of Photosynthetic Structure and Function during Nitrogen Deprivation in *Chlamydomonas reinhardtii*. *Plant Physiol.* **2015**, *167*(2), 558–573.

24. Smythers, A. L.; McConnell, E. W.; Lewis, H. C.; Mubarek, S. N.; Hicks, L. M. Photosynthetic Metabolism and Nitrogen Reshuffling Are Regulated by Reversible Cysteine Thiol Oxidation Following Nitrogen Deprivation in *Chlamydomonas*. *Plants* **2020**, *9*(6), 784.

25. Takeuchi, T.; Benning, C. Nitrogen-dependent coordination of cell cycle, quiescence and TAG accumulation in Chlamydomonas. *Biotechnol. Biofuels* **2019**, *12*(1), 1–20.

26. Szklarczyk, D.; Gable, A. L.; Nastou, K. C.; Lyon, D.; Kirsch, R.; Pyysalo, S.; Doncheva, N. T.; Legeay, M.; Fang, T.; Bork, P.; Jensen, L. J.; von Mering, C. The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* **2021**, *49*(D1), D605–D612.

27. Ramanan, R.; Kim, B.-H.; Cho, D.-H.; Ko, S.-R.; Oh, H.-M.; Kim, H.-S. Lipid droplet synthesis is limited by acetate availability in starchless mutant of *Chlamydomonas reinhardtii*. *FEBS Lett.* **2013**, *587*(4), 370–377.

28. Makepeace, K. A. T. T.; Mohammed, Y.; Rudashevskaya, E. L.; Petrotchenko, E. V.; Vögtle, F.-N. N.; Meisinger, C.; Sickmann, A.; Borchers, C. H. Improving Identification of In-organello Protein-Protein Interactions Using an Affinity-enrichable, Isotopically Coded, and Mass Spectrometry-cleavable Chemical Crosslinker. *Mol. Cell. Proteomics* **2020**, *19*(4), 624–639.

29. Szklarczyk, D.; Gable, A. L.; Lyon, D.; Junge, A.; Wyder, S.; Huerta-Cepas, J.; Simonovic, M.; Doncheva, N. T.; Morris, J. H.; Bork, P.; Jensen, L. J.; Mering, C. von STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **2019**, *47*(D1), D607–D613.

30. Kanehisa, M.; Sato, Y. KEGG Mapper for inferring cellular functions from protein sequences. *Protein Sci.* **2020**, *29*(1), 28–35.

31. Kanehisa, M.; Goto, S.; Sato, Y.; Furumichi, M.; Tanabe, M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* **2012**, *40*(D1), D109–D114.

32. Kanehisa, M.; Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **2000**, *28*(1), 27.

33. Consortium, T. U. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* **2019**, *47*(D1), D506–D515.

34. Mi, H.; Muruganujan, A.; Huang, X.; Ebert, D.; Mills, C.; Guo, X.; Thomas, P. D. Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). *Nat. Protoc.* **2019**, *14*(3), 703–721.

35. Yang, X.; Yang, S.; Qi, H.; Wang, T.; Li, H.; Zhang, Z. PlaPPISite: a comprehensive resource for plant protein-protein interaction sites. *BMC Plant Biol.* **2020**, *20*(1), 1–11.

36. Orbán-Németh, Z.; Beveridge, R.; Hollenstein, D. M.; Rampler, E.; Stranzl, T.; Hudecz, O.; Doblmann, J.; Schlögelhofer, P.; Mechtler, K. Structural prediction of protein models using distance restraints derived from cross-linking mass spectrometry data. *Nat. Protoc.* **2018**, *13*(3), 478.

37. Yang, J.; Zhang, Y. I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Res.* **2015**, *43*(W1), W174-81.

38. Van Zundert, G. C. P.; Rodrigues, J. P. G. L. M.; Trellet, M.; Schmitz, C.; Kastritis, P. L.; Karaca, E.; Melquiond, A. S. J.; Van Dijk, M.; De Vries, S. J.; Bonvin, A. M. J. J. The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J. Mol. Biol.* **2016**, *428*(4), 720–725.

39. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Meng, E. C.; Couch, G. S.; Croll, T. I.; Morris, J. H.; Ferrin, T. E. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* **2021**, *30*(1), 70–82.

40. Goddard, T. D.; Huang, C. C.; Meng, E. C.; Pettersen, E. F.; Couch, G. S.; Morris, J. H.; Ferrin, T. E. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci.* **2018**, *27*(1), 14–25.

41. Schneidman-Duhovny, D.; Inbar, Y.; Nussinov, R.; Wolfson, H. J. PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res.* **2005**, *33*, W363–W367.

42. Wang, J.; Youkharibache, P.; Zhang, D.; Lanczycki, C. J.; Geer, R. C.; Madej, T.; Phan, L.; Ward, M.; Lu, S.; Marchler, G. H.; Wang, Y.; Bryant, S. H.; Geer, L. Y.; Marchler-Bauer, A. ICn3D, a web-based 3D viewer for sharing 1D/2D/3D representations of biomolecular structures. *Bioinformatics* **2020**, *36*(1), 131–135.

43. Combe, C. W.; Fischer, L.; Rappsilber, J. XiNET: Cross-link Network Maps With Residue Resolution. *Mol. Cell. Proteomics* **2015**, *14*(4), 1137.

44. Schweppe, D. K.; Zheng, C.; Chavez, J. D.; Navare, A. T.; Wu, X.; Eng, J. K.; Bruce, J. E. XLinkDB 2.0: integrated, large-scale structural analysis of protein crosslinking data. *Bioinformatics* **2016**, *32*(17), 2716–2718.

45. Zheng, C.; Weisbrod, C. R.; Chavez, J. D.; Eng, J. K.; Sharma, V.; Wu, X.; Bruce, J. E. XLink-DB: database and software tools for storing and visualizing protein interaction topology data. *J. Proteome Res.* **2013**, *12*(4), 1989.

46. Russel, D.; Lasker, K.; Webb, B.; Velázquez-Muriel, J.; Tjioe, E.; Schneidman-Duhovny, D.; Peterson, B.; Sali, A. Putting the Pieces Together: Integrative Modeling Platform Software for Structure Determination of Macromolecular Assemblies. *PLOS Biol.* **2012**, *10*(1), e1001244.

47. Rose, A. S.; Hildebrand, P. W. NGL Viewer: a web application for molecular visualization. *Nucleic Acids Res.* **2015**, *43*, W576.

48. Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; de Beer, T. A. P.; Rempfer, C.; Bordoli, L.; Lepore, R.; Schwede, T. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **2018**, *46*(W1), W296–W303.

49. Perez-Riverol, Y.; Csordas, A.; Bai, J.; Bernal-Llinares, M.; Hewapathirana, S.; Kundu, D. J.; Inuganti, A.; Griss, J.; Mayer, G.; Eisenacher, M.; Erez, E. P.; Uszkoreit, J.; Pfeuffer, J.; Sachsenberg, T.; Yılmaz, S.; Tiwary, S.; Cox, J.; Audain, E.; Walzer, M.; Jarnuczak, A. F.; Ternent, T.; Brazma, A.; Vizcaíno, J. A. The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* **2019**, *47*(D1), D442–D450.

50. Banks, C. A. S.; Zhang, Y.; Miah, S.; Hao, Y.; Adams, M. K.; Wen, Z.; Thornton, J. L.; Florens, L.; Washburn, M. P. Integrative Modeling of a Sin3/HDAC Complex Sub-structure. *Cell Rep.* **2020**, *31*(2), 107516.

51. Bartolec, T. K.; Smith, D.-L.; Pang, C. N. I.; Xu, Y. D.; Hamey, J. J.; Wilkins, M. R. Cross-linking Mass Spectrometry Analysis of the Yeast Nucleus Reveals Extensive Protein-Protein Interactions Not Detected by Systematic Two-Hybrid or Affinity Purification-Mass Spectrometry. *Anal. Chem.* **2020**, *92*(2), 1874–1882.

52. Klykov, O.; Steigenberger, B.; Pektaş, S.; Fasci, D.; Heck, A. J. R.; Scheltema, R. A. Efficient and robust proteome-wide approaches for cross-linking mass spectrometry. *Nat. Protoc.* **2018**, *13*(12), 2964–2990.

53. Sheng, X.; Watanabe, A.; Li, A.; Kim, E.; Song, C.; Murata, K.; Song, D.; Minagawa, J.; Liu, Z. Structural insight into light harvesting for photosystem II in green algae. *Nat. Plants* **2019**, *5*(12), 1320–1330.

54. Caldwell, P.; Luk, D. C.; Weissbach, H.; Brot, N. Oxidation of the methionine residues of *Escherichia coli* ribosomal protein L12 decreases the protein's biological activity. *Proc. Natl. Acad. Sci. U. S. A.* **1978**, *75*(11), 5349.

55. Shcherbik, N.; Pestov, D. G. The Impact of Oxidative Stress on Ribosomes: From Injury to Regulation. *Cells* **2019**, *8*(11), 1379.

56. Simsek, D.; Barna, M. An emerging role for the ribosome as a nexus for post-translational modifications. *Curr. Opin. Cell Biol.* **2017**, *45*, 92–101.

57. Venezia, N. D.; Vincent, A.; Marcel, V.; Catez, F.; Diaz, J. J. Emerging role of eukaryote ribosomes in translational control. *Int. J. Mol. Sci.* **2019**, *20*(5).

58. Deroo, S.; Hyung, S.-J.; Marcoux, J.; Gordiyenko, Y.; Koripella, R. K.; Sanyal, S.; Robinson, C. V. Mechanism and rates of exchange of L7/L12 between ribosomes and the effects of binding EF–G. *ACS Chem. Biol.* **2012**, *7*(6), 1120.

59. Gao, Y. G.; Selmer, M.; Dunham, C. M.; Weixlbaumer, A.; Kelley, A. C.; Ramakrishnan, V. The structure of the ribosome with elongation factor G trapped in the posttranslocational state. *Science* **2009**, *326*(5953), 694–699.

60. Datta, P. P.; Sharma, M. R.; Qi, L.; Frank, J.; Agrawal, R. K. Interaction of the G′ Domain of Elongation Factor G and the C-Terminal Domain of Ribosomal Protein L7/L12 during Translocation as Revealed by Cryo-EM. *Mol. Cell* **2005**, *20*(5), 723–731.

61. Wang, X.; Kirkpatrick, J. P.; Launay, H. M. M.; de Simone, A.; Häussinger, D.; Dobson, C. M.; Vendruscolo, M.; Cabrita, L. D.; Waudby, C. A.; Christodoulou, J. Probing the dynamic stalk region of the ribosome using solution NMR. *Sci. Rep.* **2019**, *9*(1), 1–9.

62. Zhang, Y.; Launay, H.; Schramm, A.; Lebrun, R.; Gontero, B. Exploring intrinsically disordered proteins in Chlamydomonas reinhardtii. *Sci. Rep.* **2018**, *8*(1), 1–11.

63. Merchant, S. S.; Prochnik, S. E.; Vallon, O.; Harris, E. H.; Karpowicz, S. J.; Witman, G. B.; Terry, A.; Salamov, A.; Fritz-Laylin, L. K.; Maréchal-Drouard, L.; Marshall, W. F.; Qu, L. H.; Nelson, D. R.; Sanderfoot, A. A.; Spalding, M. H.; Kapitonov, V. V.; Ren, Q.; Ferris, P.; Lindquist, E.; Shapiro, H.; Lucas, S. M.; Grimwood, J.; Schmutz, J.; Grigoriev, I. V.; Rokhsar, D. S.; Grossman, A. R.; Cardol, P.; Cerutti, H.; Chanfreau, G.; Chen, C. L.; Cognat, V.; Croft, M. T.; Dent, R.; Dutcher, S.; Fernández, E.; Fukuzawa, H.; González-Ballester, D.; González-Halphen, D.; Hallmann, A.; Hanikenne, M.; Hippler, M.; Inwood, W.; Jabbari, K.; Kalanon, M.; Kuras, R.; Lefebvre, P. A.; Lemaire, S. D.; Lobanov, A. V.; Lohr, M.; Manuell, A.; Meier, I.; Mets, L.; Mittag, M.; Mittelmeier, T.; Moroney, J. V.; Moseley, J.; Napoli, C.; Nedelcu, A. M.; Niyogi, K.; Novoselov, S. V.; Paulsen, I. T.; Pazour, G.; Purton, S.; Ral, J. P.; Riaño-Pachón, D. M.; Riekhof, W.; Rymarquis, L.; Schroda, M.; Stern, D.; Umen, J.; Willows, R.; Wilson, N.; Zimmer, S. L.; Allmer, J.; Balk, J.; Bisova, K.; Chen, C. J.; Elias, M.; Gendler, K.; Hauser, C.; Lamb, M. R.; Ledford, H.; Long, J. C.; Minagawa, J.; Page, M. D.; Pan, J.; Pootakham, W.; Roje, S.; Rose, A.; Stahlberg, E.; Terauchi, A. M.; Yang, P.; Ball, S.; Bowler, C.; Dieckmann, C. L.; Gladyshev, V. N.; Green, P.; Jorgensen, R.; Mayfield, S.; Mueller-Roeber, B.; Rajamani, S.; Sayre, R. T.; Brokstein, P.; Dubchak, I.; Goodstein, D.; Hornick, L.; Huang, Y. W.; Jhaveri, J.; Luo, Y.; Martínez, D.; Ngau, W. C. A.; Otillar, B.; Poliakov, A.; Porter, A.; Szajkowski, L.; Werner, G.; Zhou, K. The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science* **2007**, *318*(5848), 245–251.

64. Zaffagnini, M.; Michelet, L.; Massot, V.; Trost, P.; Lemaire, S. D. Biochemical Characterization of Glutaredoxins from *Chlamydomonas reinhardtii* Reveals the Unique Properties of a Chloroplastic CGFS-type Glutaredoxin. *J. Biol. Chem.* **2008**, *283*(14), 8868–8876.

65. Gao, X.-H.; Zaffagnini, M.; Bedhomme, M.; Michelet, L.; Cassier-Chauvat, C.; Decottignies, P.; Lemaire, S. D. Biochemical characterization of glutaredoxins from *Chlamydomonas reinhardtii*: Kinetics and specificity in deglutathionylation reactions. *FEBS Lett.* **2010**, *584*(11), 2242–2248.

66. Marchler-Bauer, A.; Zheng, C.; Chitsaz, F.; Derbyshire, M. K.; Geer, L. Y.; Geer, R. C.; Gonzales, N. R.; Gwadz, M.; Hurwitz, D. I.; Lanczycki, C. J.; Lu, F.; Lu, S.; Marchler, G. H.; Song, J. S.; Thanki, N.; Yamashita, R. A.; Zhang, D.; Bryant, S. H. CDD: conserved domains and protein three-dimensional structure. *Nucleic Acids Res.* **2013**, *41*(D1), D348–D352.

67. Couso, I.; Evans, B. S.; Li, J.; Liu, Y.; Ma, F.; Diamond, S.; Allen, D. K.; Umen, J. G. Synergism between Inositol Polyphosphates and TOR Kinase Signaling in Nutrient Sensing, Growth Control, and Lipid Metabolism in *Chlamydomonas*. *Plant Cell* **2016**,

28(9), 2026–2042.

68. Couso, I.; Pérez-Pérez, M. E.; Ford, M. M.; Martínez-Force, E.; Hicks, L. M.; Umen, J. G.; Crespo, J. L. Phosphorus Availability Regulates TORC1 Signaling via LST8 in *Chlamydomonas*. *Plant Cell* **2020**, *32*(1), 69–80.

69. Ingargiola, C.; Duarte, G. T.; Robaglia, C.; Leprince, A.-S.; Meyer, C. The Plant Target of Rapamycin: A Conduc TOR of Nutrition and Metabolism in Photosynthetic Organisms. *Genes (Basel).* **2020**, *11*(11), 1285.

70. Graciet, E.; Lebreton, S.; Gontero, B. Emergence of new regulatory mechanisms in the Benson–Calvin pathway via protein–protein interactions: a glyceraldehyde-3-phosphate dehydrogenase/CP12/phosphoribulokinase complex. *J. Exp. Bot.* **2004**, *55*(400), 1245–1254.

71. Moparthi, S. B.; Thieulin-Pardo, G.; Mansuelle, P.; Rigneault, H.; Gontero, B.; Wenger, J. Conformational modulation and hydrodynamic radii of CP12 protein and its complexes probed by fluorescence correlation spectroscopy. *FEBS J.* **2014**, *281*(14), 3206–3217.

72. Matsumura, H.; Kai, A.; Maeda, T.; Tamoi, M.; Satoh, A.; Tamura, H.; Hirose, M.; Ogawa, T.; Kizu, N.; Wadano, A.; Inoue, T.; Shigeoka, S. Structure Basis for the Regulation of Glyceraldehyde-3-Phosphate Dehydrogenase Activity via the Intrinsically Disordered Protein CP12. *Structure* **2011**, *19*(12), 1846–1854.

73. Mileo, E.; Lorenzi, M.; Erales, J.; Lignon, S.; Puppo, C.; Breton, N. Le; Etienne, E.; Marque, S. R. A.; Guigliarelli, B.; Gontero, B.; Belle, V. Dynamics of the intrinsically disordered protein CP12 in its association with GAPDH in the green alga *Chlamydomonas reinhardtii*: a fuzzy complex. *Mol. Biosyst.* **2013**, *9*(11), 2869–2876.

74. Yu, A.; Xie, Y.; Pan, X.; Zhang, H.; Cao, P.; Su, X.; Chang, W.; Li, M. Photosynthetic Phosphoribulokinase Structures: Enzymatic Mechanisms and the Redox Regulation of the Calvin-Benson-Bassham Cycle. *Plant Cell* **2020**, *32*(5), 1556–1573.

75. Park, S.; Khamai, P.; Garcia-Cerdan, J. G.; Melis, A. REP27, a Tetratricopeptide Repeat Nuclear-Encoded and Chloroplast-Localized Protein, Functions in D1/32-kD Reaction Center Protein Turnover and Photosystem II Repair from Photodamage. *Plant Physiol.* **2007**, *143*(4), 1547.

76. Dewez, D.; Park, S.; García-Cerdán, J. G.; Lindberg, P.; Melis, A. Mechanism of REP27 Protein Action in the D1 Protein Turnover and Photosystem II Repair from Photodamage. *Plant Physiol.* **2009**, *151*(1), 88–99.

77. Theis, J.; Schroda, M. Revisiting the photosystem II repair cycle. *Plant Signal. Behav.* **2016**, *11*(9), 1218587.

78. Jo, G. A.; Lee, J. M.; No, G.; Kang, D. S.; Kim, S. H.; Ahn, S. H.; Kong, I. S. Isolation and characterization of a 17-kDa FKBP-type peptidyl-prolyl cis/trans isomerase from *Vibrio anguillarum*. *Protein Expr. Purif.* **2015**, *110*, 130–137.

79. Vallon, O. *Chlamydomonas* Immunophilins and Parvulins: Survey and Critical Assessment of Gene Models. *Eukaryot. Cell* **2005**, *4*(4), 836.

80. Pérez-Pérez, M. E.; Mauriès, A.; Maes, A.; Tourasse, N. J.; Hamon, M.; Lemaire, S. D.; Marchand, C. H. The Deep Thioredoxome in *Chlamydomonas reinhardtii*: New Insights into Redox Regulation. *Mol. Plant* **2017**, *10*(8), 1107–1125.

81. Erales, J.; Lorenzi, M.; Lebrun, R.; Fournel, A.; Etienne, E.; Courcelle, C.; Guigliarelli, B.; Gontero, B.; Belle, V. A New Function of GAPDH from *Chlamydomonas reinhardtii*: A Thiol−Disulfide Exchange Reaction with CP12. *Biochemistry* **2009**, *48*(25), 6034–6040.

82. Erales, J.; Mekhalfi, M.; Woudstra, M.; Gontero, B. Molecular Mechanism of NADPH-Glyceraldehyde-3-phosphate Dehydrogenase Regulation through the C-Terminus of CP12 in *Chlamydomonas reinhardtii*. *Biochemistry* **2011**, *50*(14), 2881–2888.

83. Ford, M. M.; Smythers, A. L.; McConnell, E. W.; Lowery, S. C.; Kolling, D. R. J.; Hicks, L. M. Inhibition of TOR in *Chlamydomonas reinhardtii* Leads to Rapid Cysteine Oxidation Reflecting Sustained Physiological Changes. *Cells* **2019**, *8*(10).

84. Werth, E. G.; McConnell, E. W.; Couso Lianez, I.; Perrine, Z.; Crespo, J. L.; Umen, J. G.; Hicks, L. M. Investigating the effect of target of rapamycin kinase inhibition on the *Chlamydomonas reinhardtii* phosphoproteome: from known homologs to new targets. *New Phytol.* **2019**, *221*(1), 247–260.

CHAPTER 7: Conclusions and Future Directions

7.1 Conclusions

During stress responses, changes in protein expression and activity through PTMs and/or PPIs contribute to signaling networks aimed at preserving cellular homeostasis[1,2]. The work presented in this dissertation focused on the development and implementation of LC-MS/MS, discovery-based approaches to investigate changes in PTMs or PPIs following stress. Specifically, proteolysis in *Arabidopsis thaliana* and protein phosphorylation in *Enterococcus faecalis* were examined using label-free quantitative peptidomics and proteomics, while XL-MS was implemented for the identification of PPIs in *Chlamydomonas reinhardtii*.

Arabidopsis TOPs are critical components in plant response to oxidative stress and are required for a fully functioning immune response to pathogens[3,4]. A quantitative LC-MS/MS-based peptidomics method was developed and optimized to profile the *Arabidopsis* peptidome (**Chapter 2**). This was applied in a differential peptidome analysis between Col-0 wild type and *top1top2* knock-out mutant to quantify 1111 peptides and reveal 350 peptides as potential TOPs peptide substrates. From this, 10 direct substrates were validated using *in vitro* enzyme assays with recombinant TOPs and synthetic candidate peptides. The early stages of the *Arabidopsis* immune response were characterized by analyzing the peptidomes in locally infected tissue of Col-0 wild type and *top1top2* null mutant plants (**Chapter 3**). Across these samples, 2830 peptides were quantified and revealed 1260 peptides in Col-0 and 727 peptides in *top1top2* that were significantly changing across the infection timepoints. Then, Col-0 and *top1top2* peptidomes were compared at each inoculation timepoint to distinguish TOP-mediated

proteolysis. An additional 10 direct TOPs substrates were validated using *in vitro* enzyme assays and these peptides were derived from proteins involved with photosynthesis, ATP synthesis/binding, carbon fixation, and fatty acid synthesis/beta-oxidation, implicating TOPs-mediated proteolysis in these processes during plant immune responses.

One critical determinant of *Enterococcus faecalis* antimicrobial resistance is the transmembrane kinase IreK[5]. To better understand stress-modulated phosphorylation events contributing to antimicrobial resistance the phosphoproteomes of untreated and cell wall-active antimicrobial treated wild type *E. faecalis* were compared (**Chapter 4**). Overall, 161 unique phosphopeptides from 94 unique proteins were quantified and 87 phosphopeptides were determined to be significantly changing across the treatments. This analysis was then expanded to include phosphoproteomes of untreated and treated *E. faecalis* Δ*ireK* null mutant to determine the dependence of stress-mediated phosphorylation on IreK. This revealed 44 potential IreK substrates under basal conditions and 79 and 83 IreK-dependent phosphopeptides after treatment with chlorhexidine and ceftriaxone, respectively. Follow up biochemical studies were conducted to validate direct phosphorylation of GpsB and MltG by IreK, reinforcing the association between IreK, *E. faecalis* defense response, and cell wall homeostasis.

In addition to this quantitative work investigating PTMs and their role in stress responses, this dissertation contains detailed procedures for IP-MS and XL-MS approaches to identify novel PPIs (**Chapter 5**). This developed XL-MS platform using the crosslinker DSSO was applied to elucidate the global protein interactome of *Chlamydomonas reinhardtii*, a unicellular green alga used studied as a model organism for biochemical processes including photosynthesis, autophagy, cell cycle control, and lipid accumulation (**Chapter 6**)[6–8]. This analysis resulted in the identification of 769 unique crosslinks, 157 and 612 of which were interlinks and intralinks,

respectively. Intralinks were mapped to homology models and showed that 96% were within the distance restraint for DSSO, thus adding confidence to the crosslinking analysis. These experimentally determined intralinks were used as constraints to guide computational modeling to predict the structure of elongation Factor 3 and ABC transporter. For the detected interlinks, many confirmed existing complexes known in *C. reinhardtii*, while others suggested the existence of novel complexes, including the interaction between Mg-chelatase and an uncharacterized Mg-chelatase associated protein without a homolog in higher plants. Overall, this work enables further targeted analysis of PPIs in *C. reinhardtii* that may be essential to understanding its metabolic processes and could be applied to study changes in PPIs due to stress.

7.2 Future Directions

The research presented here describes the regulation of proteins in response to stress and identifies proteins involved in complex stress signaling networks. Further investigation is required to determine how these proteins and modifications contribute to overall organism survival. This knowledge could be leveraged to manipulate implicated pathways to produce pro-life or pro-death outcomes. In plants, this will contribute to minimizing the effect of stressors and preventing crop yield loss. For bacteria, understanding antimicrobial resistance mechanisms can identify therapeutic targets for the prevention or treatment of life-threatening infections.

7.2.1 Validation of Bioactive Peptides

While novel TOP direct substrates were discovered in **Chapter 2** and **Chapter 3**, further characterization of these peptides is required to determine if they are degradation products or bioactive defense signaling molecules. Human thimet oligopeptidase processes intracellular peptides, where it is known to produce peptides for MHC-1 antigen presentation as well as

degrade several neuropeptides[9]. For *Arabidopsis* TOPs, this could mean that its confirmed substrates and cleaved peptide products could both have defense signaling functions. To test this, Col-0 plants can be treated with synthetic peptides (both the substrates and determined peptide products), and assessed for the susceptibility of these treated plants to biotic stress compared to untreated plants. Likewise, *top1top2* null mutant plants could be treated as well and compared to untreated Col-0 and mutant plants following bacterial inoculation to determine if the peptide treatment restores a similar resistance phenotype to the Col-0 plants.

7.2.2 Biochemical Validation and Mechanistic Studies for IreK Phosphorylation

Various follow up studies that serve to confirm and characterize candidate IreK substrates identified in **Chapter 4** include: 1) *in vitro* kinase assays with recombinant IreK and candidates, 2) production of epitope-tagged candidates in *E. faecalis* for enrichment and confirmation of *in vivo* phosphorylation with and without cell wall-active antimicrobial treatment, 3) genetic knockouts and overexpression of candidates to assess their role in antimicrobial resistance, and 4) mutagenesis at candidate phosphosites to determine the role of phosphorylation towards antimicrobial resistance *in vivo*. While some of these avenues are outside the scope of the Hicks Laboratory, there are certain experiments that could benefit from the incorporation of LC-MS/MS analysis. For example, identified IreK substrates that contain multiple phosphosites, where the importance of each individual site to antimicrobial resistance is unknown. It would therefore be beneficial to analyze whether altered *E. faecalis* antimicrobial resistance following mutagenesis of a particular site is due to that specific phosphorylation event or the influence of that event on other phosphosites. In this experiment, substrate mutants that cause a change in resistance will be epitope-tagged for enrichment from protein lysates. The phosphorylation on this mutant can then be quantitatively compared to this protein enriched from the wild type *E.*

faecalis strain following LC-MS/MS analysis to determine if the presence/absence of a particular phosphosite modulates the phosphorylation at other residues.

IreK has been shown to autophosphorylate at multiple sites *in vitro*[10], but the extent of this phosphorylation *in vivo* is currently unknown. Furthermore, IreK phosphorylation increases in a stress-specific manner *in vivo*[5], but specific sites that impact IreK function in response to stress have not been determined. To identify these phosphosites, IreK will be epitope-tagged and enriched from *E. faecalis* with or without antimicrobial treatment. Purified IreK will then be proteolyzed and analyzed using LC-MS/MS analysis. Preliminary experiments have been performed in the Hicks Laboratory, but no IreK phosphopeptides have been detected in the methods that were implemented. This consisted of an in-solution digestion/in-gel digestion with trypsin (including chymotrypsin), and TiO_2 phosphopeptide enrichment, all analyzed on a Sciex TripleTOF 5600 in the Hicks Laboratory. One issue is that the proteolyzed peptide containing T163, T166, and T168 (within the IreK activation loop, phosphorylation of these sites contributes to increased resistance[10]) could be triply phosphorylated and is large which hinders ionization and analysis. Future experiments should focus on targeted analyses (PRM) on the Thermo Q Exactive HF-X in the Mass Spectrometry Core Laboratory and the use of other proteases that provide peptides of a more manageable size for analysis.

7.2.3 Quantitative XL-MS

While qualitative XL-MS approaches can be implemented to discover novel PPIs, quantitative XL-MS methods are required to detect PPIs changes in different stress or disease states. For example, quantitative XL-MS studies could provide a better understanding into how nutrient deprivation results in triacylglycerol accumulation in *C. reinhardtii*[8,11]. Since crosslinkers have maximum crosslinking distances, changes in protein conformation or PPIs in

different conditions can be detected by quantitative comparisons of crosslinked peptide intensities. It should be noted that changes in these abundances can also reflect the prevention of a crosslink due to steric hinderances or surface-inaccessibility in these new conformations and not the reduction of a specific PPI[12]. The abundance of a crosslinked residue pair is calculated using the abundances of identified peptides containing a crosslink between these residues. This is similar to a global proteome analysis using bottom-up proteomics where the abundances of multiple peptides are consolidated to provide a protein abundance. Unlike the global proteome analysis, crosslinked residue pairs are supported by fewer peptides, highlighting the necessity of analytical reproducibility for quantitative accuracy[13]. Many software packages can support crosslinked peptide label-free quantification including Skyline[14], MassChroQ[15], and xTract[16]. These tools have capabilities for chromatographic alignment and peak matching, which assists in the quantification of crosslinked peptides that were not identified in every sample due to poor fragment spectra quality.

Since linear, non-crosslinked peptide greatly outnumber and suppress the ionization of crosslinked peptides, chromatographic enrichment or affinity enrichment (with affinity group-containing crosslinkers) is necessary to improve the depth of coverage of crosslinked peptides. Maintaining reproducibility during enrichment is a challenge and the number of fractions collected for each replicate severely increases total acquisition time for label-free quantification. Labeling techniques (*e.g.*, isotope-labelled crosslinkers or isobaric labelling) have been applied to quantitative XL-MS studies and allow for the pooling of samples, which increases reproducibility and decreases acquisition time[17,18]. Due to the low abundance of crosslinked peptides, the reproducibility of XL-MS studies can be hindered by data-dependent acquisition (DDA) if certain crosslinked peptides are not selected for fragmentation. This also affects

methods using isobaric labelling because low precursor abundance leads to low reporter ion intensities, which can result in inaccurate quantification[13]. Alternatively, data-independent acquisition (DIA) can be used to analyze the fragment ions of all detected precursors[19], but these methods require the creation of reference fragment spectra for spectra library matching, are not compatible with isobaric labeling techniques, and Spectronaut[20] is currently the only DIA-based quantification software that supports an input of crosslinked peptide identifications.

REFERENCES

1. Fulda, S.; Gorman, A. M.; Hori, O.; Samali, A. Cellular stress responses: Cell survival and cell death. *Int. J. Cell Biol.* **2010**, *2010*, 214074.

2. Spriggs, K. A.; Bushell, M.; Willis, A. E. Translational Regulation of Gene Expression during Conditions of Cell Stress. *Mol. Cell* **2010**, *40*(2), 228–237.

3. Moreau, M.; Westlake, T.; Zampogna, G.; Popescu, G.; Tian, M.; Noutsos, C.; Popescu, S. The *Arabidopsis* oligopeptidases TOP1 and TOP2 are salicylic acid targets that modulate SA-mediated signaling and the immune response. *Plant J.* **2013**, *76*(4), 603–614.

4. Westlake, T. J.; Ricci, W. A.; Popescu, G. V.; Popescu, S. C. Dimerization and thiol sensitivity of the salicylic acid binding thimet oligopeptidases TOP1 and TOP2 define their functions in redox-sensitive cellular pathways. *Front. Plant Sci.* **2015**, *6*(327), 1–16.

5. Labbe, B. D.; Kristich, C. J. Growth- and stress-induced PASTA kinase phosphorylation in *Enterococcus faecalis*. *J. Bacteriol.* **2017**, *199*(21), e00363-17.

6. Pérez-Pérez, M. E.; Couso, I.; Heredia-Martínez, L. G.; Crespo, J. L. Monitoring Autophagy in the Model Green Microalga *Chlamydomonas reinhardtii*. *Cells* **2017**, *6*(4), 36.

7. Cross, F. R.; Umen, J. G. The Chlamydomonas cell cycle. *Plant J.* **2015**, *82*(3), 370–392.

8. Park, J. J.; Wang, H.; Gargouri, M.; Deshpande, R. R.; Skepper, J. N.; Holguin, F. O.; Juergens, M. T.; Shachar-Hill, Y.; Hicks, L. M.; Gang, D. R. The response of *Chlamydomonas reinhardtii* to nitrogen deprivation: a systems biology analysis. *Plant J.* **2015**, *81*(4), 611–624.

9. Gewehr, M. C. F.; Teixeira, A. A. S.; Santos, B. A. C.; Biondo, L. A.; Gozzo, F. C.; Cordibello, A. M.; Eichler, R. A. S.; Reckziegel, P.; Silva, R. N. O. Da; Santos, N. B. Dos; Camara, N. O. S.; Castoldi, A.; Barreto-Chaves, M. L. M.; Dale, C. S.; Senger, N.; Lima, J. D. C. C.; Seelaender, M. C. L.; Inada, A. C.; Akamine, E. H.; Castro, L. M.; Rodrigues, A. C.; Neto, J. C. R.; Ferro, E. S. The Relevance of Thimet Oligopeptidase in the Regulation of Energy Metabolism and Diet-Induced Obesity. *Biomolecules* **2020**, *10*(2), 321.

10. Labbe, B. D.; Hall, C. L.; Kellogg, S. L.; Chen, Y.; Koehn, O.; Pickrum, A. M.; Mirza, S. P.; Kristich, C. J. Reciprocal regulation of PASTA kinase signaling by differential modification. *J. Bacteriol.* **2019**, *209*(10), e00016-19.

11. Siaut, M.; Cuiné, S.; Cagnon, C.; Fessler, B.; Nguyen, M.; Carrier, P.; Beyly, A.; Beisson, F.; Triantaphylidès, C.; Li-Beisson, Y.; Peltier, G. Oil accumulation in the model green alga *Chlamydomonas reinhardtii*: characterization, variability between common laboratory strains and relationship with starch reserves. *BMC Biotechnol.* **2011**, *11*(1), 1–15.

12. Chen, Z. A.; Pellarin, R.; Fischer, L.; Sali, A.; Nilges, M.; Barlow, P. N.; Rappsilber, J. Structure of Complement C3(H2O) Revealed By Quantitative Cross-Linking/Mass Spectrometry And Modeling. *Mol. Cell. Proteomics* **2016**, *15*(8), 2730–2743.

13. Chen, Z. A.; Rappsilber, J. Protein Dynamics in Solution by Quantitative Crosslinking/Mass Spectrometry. *Trends Biochem. Sci.* **2018**, *43*(11), 908–920.

14. MacLean, B.; Tomazela, D. M.; Shulman, N.; Chambers, M.; Finney, G. L.; Frewen, B.; Kern, R.; Tabb, D. L.; Liebler, D. C.; MacCoss, M. J. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **2010**, *26*(7), 966–968.

15. Valot, B.; Langella, O.; Nano, E.; Zivy, M. MassChroQ: A versatile tool for mass spectrometry quantification. *Proteomics* **2011**, *11*(17), 3572–3577.

16. Walzthoeni, T.; Joachimiak, L. A.; Rosenberger, G.; Röst, H. L.; Malmström, L.; Leitner, A.; Frydman, J.; Aebersold, R. XTract: software for characterizing conformational changes of protein complexes by quantitative cross-linking mass spectrometry. *Nat. Methods* **2015**, *12*(12), 1185–1190.

17. Fischer, L.; Chen, Z. A.; Rappsilber, J. Quantitative cross-linking/mass spectrometry using isotope-labelled cross-linkers. *J. Proteomics* **2013**, *88*, 120–128.

18. Yu, C.; Huszagh, A.; Viner, R.; Novitsky, E. J.; Rychnovsky, S. D.; Huang, L. Developing a Multiplexed Quantitative Cross-Linking Mass Spectrometry Platform for Comparative Structural Analysis of Protein Complexes. *Anal. Chem.* **2016**, *88*(20), 10301–10308.

19. Müller, F.; Rappsilber, J. A protocol for studying structural dynamics of proteins by quantitative crosslinking mass spectrometry and data-independent acquisition. *J. Proteomics* **2020**, *218*, 103721.

20. Bruderer, R.; Bernhardt, O. M.; Gandhi, T.; Miladinović, S. M.; Cheng, L.-Y.; Messner, S.; Ehrenberger, T.; Zanotelli, V.; Butscheid, Y.; Escher, C.; Vitek, O.; Rinner, O.; Reiter, L. Extending the Limits of Quantitative Proteome Profiling with Data-Independent Acquisition and Application to Acetaminophen-Treated Three-Dimensional Liver Microtissues. *Mol. Cell. Proteomics* **2015**, *14*(5), 1400–1410.