# Mass Spectrometry Characterization of Post-Translational Modifications on the *Arabidopsis* Growth Regulator Protein RGA

Ellen Hutchins Speers
Charlottesville, VA

B.S. Chemistry, College of William and Mary, 2011
M.S. Chemistry, Northwestern University, 2013

A Dissertation Presented to the Graduate Faculty of the University of Virginia
in Candidacy for the Degree of Doctor of Philosophy
Department of Chemistry

University of Virginia
May, 2020

# Acknowledgements

I'd like to thank all the excellent people who've helped on this long journey. Dr. Donald Hunt, you've been a brilliant and understanding advisor, and I'm glad to have had the opportunity to work with one of the greats of mass spectrometry. Thank you for sharing your breadth of knowledge, linking the big picture and the precise details. Dr. Jeffrey Shabanowitz, thank you for sharing your instrumental know-how, and for encouraging me to think like a scientist. Dr. Dina Bai, you're not just a computer wizard, but you're the heart of the lab. Thanks for so many heart-to-heart conversations. Thanks also to Dr. Paisley Myers and Dr. Ben Barnhill, who showed me the ropes and shared many tips and tricks to get those procedures working. I'd also like to thank all the Hunt lab members I've known, past and present: Dr. Lissa Anderson, Dr. Stacy Malaker, Dr. Scott Ugrin, Dr. Stephanie Miller-Lehman, Garrett Tanner, Dr. Josh Hinkle, Dr. Elizabeth Duselis, Dr. Rob D'Ippolito, Mohammed Azizzanjani, Eric Hunt, Keira Mahoney, Negin Ghafourian, Hossain Shadman, Maria Panepinto, and Xi Peng. Dr. Mark Ross and Emily Zahn, you've been a great help in this project, and I look forward to seeing where you take it.

In addition, I'd like thank my committee members, Dr. David Cafiso, Dr. Ken Hsu, Dr. Jill Venton, and Dr. David Brautigan, for reading my thesis. Also on my Candidacy committee were Dr. Rebecca Pompano and Dr. Cameron Mura.

I'd like to thank my collaborators at Duke, Dr. Tai-Ping Sun, who has greatly advanced the studies of RGA over the years, and Dr. Rodolfo Zentella, who handled all the plant biology of these experiments. Without them, this project would never have bloomed.

To those who've inspired me on this scientific odyssey, the professors I've known before, thank you: Dr. Greg Clines, Dr. Bob Orwoll of William and Mary, Dr. Richard Silverman and Dr.

Tom O'Halloran of Northwestern. Ms. Johnson, thank you for getting me interested in Chemistry in the first place!

My friends and family have been so important. Erika, my best friend since middle school, my partner in crime, thanks for so many great memories. Also, thanks to my friends in Chemistry, Thushani, Jessica, Matt, TJ, Chris, Anna, Charles, and Jason. We've fought many battles and won many victories together. Tammy Shull and Brenda Griswold, thank you for your wisdom and encouragement. Of course, I can't forget my parents. Mom and Dad, thank you for so many years of love and support, thank you for nurturing my love for science and God, and challenging me to be the best I can be. Thank you also, Leigh, my birth mom, the one who gave me life. And ultimately, I'd like to thank my Heavenly Father. Thank you for giving us this awesome world to discover, and an amazing life both now and forever.

# Table of Contents

# List of Figures and Tables

**Chapter 1**



**Chapter 2**

# Thesis Abbreviations

| | |
|---|---|
| $^{o}C$ | degrees Celsius |
| · | radical species |
| $^{12}C$ | carbon-12 |
| $^{13}C$ | carbon-13 |
| 6His-3xFLAG | tandem purification tags (poly-histidine and FLAG) |
| a | acetylation |
| *a* | Matthieu parameter |
| A | Solvent A, 0.5% acetic acid in water |
| Å | Angstrom |
| A, Ala | alanine |
| α | alpha |
| *A. thaliana* | *Arabidopsis thaliana* |
| AC | alternating current |
| AC | analytical column |
| ACPC | pre-column attached to an analytical column |
| AGC | automatic gain control |
| aK | acetylated lysine |
| AMBIC | ammonium bicarbonate |
| Angio | Angiotensin I peptide (DRVYIHPFHL) |
| AtGAI | *Arabidopsis thaliana* GIBBERELLIC ACID INSENSITIVE, a DELLA protein |
| AtRGL1/2/3 | *Arabidopsis thaliana* RGA-LIKE1/2/3, DELLA proteins |
| b | b ion |
| B | Solvent B, 0.5% acetic acid in 66% acetonitrile, 33% water |
| β | beta |
| Bar | 750.1 Torr, 14.5 pounds per square inch |
| BLAST | Basic Local Alignment Search Tool |
| BZR1 | BRASSINAZOLE-RESISTANT 1, a transcription factor |
| c | c ion |
| C | carboxyl terminus of a peptide or protein |

| | |
|---|---|
| C, Cys | cysteine |
| C18 | octadecylsilane bound to a silica stationary phase bead |
| CAD | collision-activated dissociation |
| CHAPS | 3-[(3-cholamidopropyl)dimethylammonio]-1-propane sulfonate |
| cm | centimeter |
| COBALT | Constraint-based Multiple Alignment Tool |
| D, Asp | aspartic acid |
| Da | Dalton |
| DC | direct current, a set voltage |
| DELLA | DELLA protein, domain (RGA residues 1-111), or motif (RGA 44-48) |
| DNA | deoxyribonucleic acid |
| DTT | dithiothreitol |
| E, Glu | glutamic acid |
| ECD | electron capture dissociation |
| EL1 | EARLY-FLOWERING1, a rice kinase |
| ESI | electrospray ionization |
| ETD | electron transfer dissociation |
| ETnoD | electron transfer with no dissociation |
| eV | electron volt, $1.602*10^{-19}$ Joules |
| ExPASy | Expert Protein Analysis System |
| f, fuc | fucose, fucosylation |
| F, Phe | phenylalanine |
| FAB | fast atom bombardment |
| FDR | false discovery rate |
| FLAG | FLAG tag |
| fmol | femtomole ($10^{-15}$ mol) |
| fS, fT | fucosylated serine, threonine |
| FT | Fourier transform |
| FT-ICR | Fourier transform ion cyclotron resonance |
| FTMS | Fourier transform mass spectrometer |
| FWHM | full width at half-height |

| | |
|---|---|
| g, GlcNAc | N-acetylglucosamine, GlcNAcylation |
| G, Gly | glycine |
| GA | gibberellin |
| GA1 | ent-copalyl diphosphate synthase 1, a gibberellin biosynthesis enzyme |
| *ga1-3* | null mutant of GA1 |
| GID1 | GIBBERELLIN INSENSITIVE DWARF1 protein |
| GRAS | GRAS (GAI, RGA, and SCR) protein or domain (RGA residues 212-581) |
| gS, gT | GlcNAcylated serine, threonine |
| GVIG | GVIGTTVTTTTTTTTAAGESTR peptide, from RGA (residues 186-207) |
| h, Hex | Hexose, Hexosylation |
| H, His | histidine |
| $H_2O$ | water |
| HAT | histone acetyltransferase |
| HCD | higher-energy collisional dissociation |
| HDAC | histone deacetylase |
| HPLC | high-performance liquid chromatography |
| hS, hT | hexosylated serine, threonine |
| HvSLN1 | *Hordeum vulgare* SLENDER PROTEIN1, the barley DELLA protein |
| Hz | Hertz |
| I, Ile | isoleucine |
| i.d. | inner diameter |
| IAM | iodoacetamide |
| IRM | ion-routing multipole |
| ITMS | ion trap mass spectrometer |
| K, Lys | lysine |
| Kasil | potassium silicate solution |
| kDa | kilodalton ($10^3$ Da) |
| keV | kiloelectron volt ($10^3$ eV) |
| kV | kilovolt ($10^3$ V) |
| L, Leu | leucine |
| LC | liquid chromatography |

| | |
|---|---|
| LC-MS | liquid chromatography – mass spectrometry |
| LExLE | motif in the DELLA domain (RGA residues 66-70) |
| LHRI, LHRII | leucine heptad repeat I/II, subdomains in the GRAS domain (RGA residues 219-273 and 371-403, respectively) |
| LIT | linear ion trap |
| LMCO | low-mass cutoff |
| LSN | LSNHGTSSSSSISK(DK) peptide, from RGA |
| LSN15 | LSNHGTSSSSSISK (residues 12-26) |
| LSN17 | LSNHGTSSSSSISKDK (residues 12-28) |
| LTQ | linear ion trap quadrupole |
| M, Met | Methionine |
| $(M+2H)^{+\cdot}$ | charge-reduced molecular ion |
| $(M+nH)^{n+}$ | molecular ion with n charges |
| m/z | mass to charge ratio |
| MA | mass area |
| MALDI | matrix-assisted laser desorption ionization |
| Mb | millions of base pairs |
| min | minute |
| µL | microliter ($10^{-6}$ L) |
| µm | micrometer ($10^{-6}$ m) |
| µM | micromolar ($10^{-6}$ mol/L) |
| mL | milliliter ($10^{-3}$ L) |
| mm | millimeter ($10^{-3}$ m) |
| mM | millimolar ($10^{-3}$ mol/L) |
| mol | mole ($6.022 \times 10^{26}$) |
| MoRF | molecular recognition factor |
| mRNA | messenger ribonucleic acid |
| ms | millisecond |
| MS | mass spectrometry, mass spectrometer, mass spectrum |
| MS1 | first stage of mass spectrometry/spectrum (of a tandem MS experiment), full mass spectrum |

| | |
|---|---|
| MS2, MS/MS | second stage of mass spectrometry/spectrum (of a tandem MS experiment), tandem mass spectrometry |
| N | amino terminus of a polypeptide |
| N, Asn | asparagine |
| NaCl | sodium chloride |
| ND | not detected (below the limit of detection) |
| ng | nanogram ($10^{-9}$ g) |
| $Ni^{2+}$-NTA | nickel (II) – nitrilotriacetic acid column |
| nL | nanoliter ($10^{-9}$ L) |
| o.d. | outer diameter |
| O-fucose | O-linked fucosylation |
| OGA | O-GlcNAcase |
| O-GlcNAc | O-linked N-acetylglucosamination |
| OGT | O-GlcNAcyltransferase |
| O-hexose | O-linked hexosylation |
| OsSLR1 | *Oryza sativa* SLENDER RICE1, the rice DELLA protein |
| p, $PO_4$ | phosphate, phosphorylation |
| PC | pre-column |
| PFYRE | PFYRE subdomain of the GRAS domain (RGA residues 415-402) |
| pH | potential hydrogen |
| PIF | PHYTOCHROME INTERACTING FACTOR, transcription factors |
| pmol | picomole ($10^{-12}$ mol) |
| POFUT | protein O-fucosyltransferase |
| Poly-S/T | residues 112 – 211 of RGA, including the VIP, SCSS, and GVIG peptides |
| ppm | parts per million |
| pS, pT, pY | phosphorylated serine, threonine, tyrosine |
| psi | pounds per square inch |
| PtDELLA | DELLA protein from *Paulownia tomentosa* (princess tree) |
| PTM | post-translational modification |
| *q* | Matthieu parameter |
| Q, Gln | glutamine |
| QMF | quadrupole mass filter |

| | |
|---|---|
| r | radius |
| r | resolving power, resolution |
| R, Arg | arginine |
| RF | radiofrequency |
| RGA | REPRESSOR of GA1-3, an *Arabidopsis* DELLA protein |
| $RGA^{GKG}$ | RGA synthetic gene with an inserted lysine between G185 and G186 |
| RNA | ribonucleic acid |
| RP-HPLC | reverse phase high-performance liquid chromatography |
| rpm | rotations per minute |
| RT | retention time |
| s | seconds |
| S, Ser | serine |
| SAW | SAW subdomain of the GRAS domain (RGA residues 505-581) |
| SCF | SCF E3 ubiquitin ligase complex |
| SCSS | (LK)SCSSPDSMVTSTSTGTQIGK peptide, from RGA |
| SCSS20 | SCSSPDSMVTSTSTGTQIGK (residues 167-185 + an inserted lysine) |
| SCSS22 | LKSCSSPDSMVTSTSTGTQIGK (residues 165-185 + an inserted lysine) |
| SDS | sodium dodecyl sulfate |
| SDS-PAGE | SDS polyacrylamide gel electrophoresis |
| SEC | SECRET AGENT, an O-GlcNAcyltransferase |
| *sec-3* | SEC null mutant |
| SLY1 | SLEEPY1, an F-box protein |
| *sly1-10* | mutant of SLY1 with an altered C-terminal target-interacting domain |
| SPY | SPINDLY, a protein O-fucosyltransferase |
| *spy-12/-15/-19* | mutants of SPY with a point mutation in the catalytic domain |
| *spy-8* | mutant of SPY with a deletion in tetratricopeptide regions 2 and 3 |
| T, Thr | threonine |
| TaRHT1, TaRHT-B1/-D1 | *Triticum aestivum* REDUCED HEIGHT PROTEIN 1, wheat DELLA proteins |
| TF | transcription factor protein |
| TOF | time of flight |
| TOPP4 | SERINE/THREONINE-PROTEIN PHOSPHATASE PP1 isozyme 4, an *Arabidopsis* phosphatase |

| | |
|---|---|
| *topp4-1* | mutant of TOPP4 with a point mutation |
| Torr | $1.332*10^{-3}$ bar |
| TR | transcription regulator protein |
| Tris-HCl | Tris(hydroxymethyl)aminomethane-HCl |
| TVHYNP | motif in the DELLA domain (RGA residues 88 – 93) |
| U | DC quadrupolar potential |
| V | RF AC potential |
| V | volts, voltage |
| V, Val | valine |
| v:v | volume-to-volume ratio |
| Vaso | Vasoactive intestinal peptide fragment 1-12 (HSDAVFTDNYTR) |
| VHIID | VHIID subdomain of the GRAS domain (RGA residues 292-357) |
| VIP | VIPGNAIYQFPAIDSSSSSNNQNK(R) peptide, from RGA |
| VIP24 | VIPGNAIYQFPAIDSSSSSNNQNK (residues 140-163) |
| VIP25 | VIPGNAIYQFPAIDSSSSSNNQNKR (residues 140-164) |
| VvGAI | *Vitis vinifera* GIBBERELLIC ACID INSENSITIVE, a grapevine DELLA protein |
| W, Trp | tryptophan |
| x | any amino acid |
| y | y ion |
| Y, Tyr | tyrosine |
| z | charge |
| z· | z radical ion |
| ZmD8 | *Zea mays* DWARF8, a maize DELLA protein |
| Ω | trapping RF angular frequency |
| ω | secular frequency |

# Abstract

The world population has increased several-fold in the past century, but food production has increased at a greater rate, thanks to scientific developments like the Green Revolution. The Green Revolution included the development of sturdy semi-dwarf crops, which were able to bear heavy heads of grain without falling over in wind or rain. These plants were later found to have mutations in the gibberellin signaling pathway, including DELLA proteins. These master regulator proteins integrate multiple hormonal and environmental signaling pathways in order to repress plant growth. Gibberellin represses DELLA activity and causes its proteasomal degradation. DELLA proteins are also affected by post-translational modifications (PTMs).

PTMs like phosphorylation have varied and important effects on proteins, influencing their stability, structure, activity, and interactions with other proteins. Mass spectrometry is the premier technique for studying protein PTMs; it can determine the precise sites and levels of PTMs on a protein. This work uses tandem mass spectrometry to study REPRESSOR OF GA1-3 (RGA), a DELLA protein in the model plant *Arabidopsis thaliana*. RGA has two poly-S/T stretches that are heavily modified with different PTMs. Phosphorylation, GlcNAcylation, hexosylation, and fucosylation levels and sites on RGA were characterized in its native *Arabidopsis*. Different *Arabidopsis* mutants were examined, including the gibberellin biosynthesis enzyme *ga1*, the adaptor subunit of the SCF E3 ubiquitin ligase complex *sly1*, and *sly1* in combination with mutants of the O-GlcNAcyltransferase *sec* or the O-fucosyltransferase *spy*. A new modification of RGA was discovered: acetylation. These are some of the first acetylation sites seen on any DELLA protein in any organism. In addition, acetylation and phosphorylation occurred at nearby residues, and acetylation levels correlated with phosphorylation levels across the different mutants. Both phosphorylation and acetylation were increased in gibberellin-deficient *ga1-3* mutant vs. *sly1-10*

mutant plants; these mutants differ in the binding of GID1. These modifications are likely linked and play a role in the regulation of RGA by stabilizing it.

# Chapter 1: Introduction to the Dissertation

## 1.1 Proteins and Post-Translational Modifications

Proteins are integral components of cells. Proteins provide cell structure and many functions such as metabolism, signaling, material recycling, growth and senescence, and more. Proteins are coded for by genes, which are encoded in stretches of DNA. The gene is transcribed from DNA into mRNA in the nucleus. The mRNA is then translated by a ribosome into a protein. A protein is a linear chain of L-alpha-amino acids, also called residues. There are twenty common amino acids that comprise a protein, ranging in mass from glycine (57 Da) to tryptophan (186 Da). Each amino acid has the same backbone, connected to one another by amide bonds, but differing side chains. The side chains each have different properties. For example, lysine (Lys, K) and arginine (Arg, R) are basic residues that are positively charged at cellular pH. Serine (Ser, S) and threonine (Thr, T) possess a hydroxyl group (-OH), and tyrosine (Tyr, Y) has a phenolic group. Proteins have several levels of structure. The primary structure is the amino acid sequence. The secondary structure consists of local folds such as beta-sheets and alpha-helices formed as the chain folds up. These sheets and helices then fold up further in a form dependent on the sequence into the tertiary structure of the protein. If the protein exists in a complex with multiple subunits, the assembly is referred to as quaternary structure (1).

Post-translational modifications (PTMs) are chemical groups added to a protein after it has been synthesized from an mRNA template. PTMs allow for a protein molecule to be reversibly and repeatedly modified and controlled over its lifetime, allowing for more rapid and precise modulation than simply upregulating or downregulating transcription and

translation of its gene (in minutes, rather than hours) (1,2). PTMs may be rapidly added to or removed from a protein, and a given protein may be modified at many different sites with a number of different PTMs, with different combinations across time. PTMs are usually added to a protein by a specific enzyme, which recognizes a motif (often an array of nearby amino acids). There are hundreds of identified PTMs. PTMs are crucial in many cellular processes such as signaling, growth, and metabolism (1). The structures of several PTMs studied in this work are given in Figure 1.1.

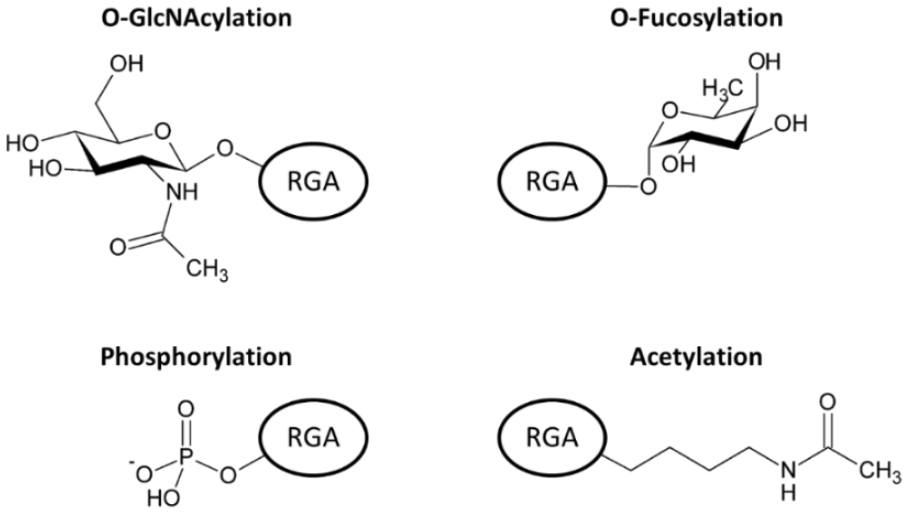


**Figure 1.1 Examples of post-translational modifications found on the protein RGA.** O-GlcNAcylation and O-fucosylation are found on S or T; phosphorylation is found on S, T, or Y; acetylation is found on K.

Phosphorylation, one of the most common PTMs, is a transient modification. A phosphate group (+$HPO_3$, +79.9663 Da) is transferred to a protein from adenosine triphosphate by a kinase; it is removed by a phosphatase. Phosphorylation usually occurs on threonines, serines, and tyrosines, with percent abundances of 80-85%, 10-15%, and 0-5% in plants, respectively. Phosphorylation is frequently used to turn proteins “on” or “off.” For example, phosphorylation cascades are involved in and regulate cell signaling, gene transcription and translation, protein-protein interactions, cell growth and senescence (1,3,4). An estimated one third of plant proteins are phosphorylated at any given time (2).

O-GlcNAcylation was first discovered by Torres and Hart in 1984 (5). β-N-acetylglucosamine (GlcNAc) is a glucose with an appended N-acyl group (+$C_8H_{13}NO_5$, +203.0794 Da). GlcNAcylation can occur on serine, threonine, or asparagine residues. GlcNAc can be found on proteins as part of glycan chains on a residue (usually N-linked) or singly (usually O-linked). O-GlcNAcylation is not found on secreted or cell surface proteins. O-GlcNAc is transferred to a protein from uridine-5'-diphosphate-GlcNAc (UDP-GlcNAc) by an O-GlcNAcyltransferase (OGT) enzyme. O-GlcNAc is removed by an O-GlcNAcase (OGA). UDP-GlcNAc biosynthesis is affected by almost all metabolic pathways, and its level is indicative of the metabolic status of the cell. Like phosphorylation, O-GlcNAc is a transient modification that is added to and removed from proteins quite rapidly. This affects cell stress response, transcriptional regulation, protein synthesis and turn over, metabolic monitoring and modulation, and cell cycle regulation. Unlike phosphorylation, in which there are many kinases in a given species that phosphorylate proteins, there are only one to a few OGTs and OGAs per species (1,6–9).

O-fucosylation was first discovered by Kentzer et al. in 1990 in human urokinase (10). Fucose (+$C_6H_{10}O_4$, +146.0579 Da) is a mono-dehydrated glucose. Fucosylation can occur on serine or threonine residues. Fucose can be found on proteins as part of glycan chains on a residue (usually N-linked) or singly (usually O-linked) (1). Fucose is transferred by a protein O-fucosyltransferase (POFUT) from guanosine-5'-diphosphate-fucose (GDP-fucose) to the target protein. O-fucosylation can have several different effects. For example, it is necessary for the function of the developmental protein Notch, and plays a role in proper folding and secretion of proteins with thrombospondin type I repeats in the endoplasmic reticulum (11). It also can affect protein-protein interactions to control gene expression (12).

Acetylation is a modification found on lysine residues and the N-terminus of proteins (1). N-terminal lysine acetylation of histones was described as a regulator of RNA transcription by Allfrey et al. in 1964 (13). The acetyl group (+$CH_2CO$, + 42.0105 Da) is a very common modification. Over 80% of human cytoplasmic proteins are acetylated at their N-terminus. The acetyl group is transferred from acetyl-CoA by an acetyltransferase enzyme (HAT), and removed by a deacetylase (HDAC). Acetylation affects the activity of cytoskeletal proteins, transcription factors, molecular chaperones, and histones (1). Acetylating a lysine neutralizes the positive charge on the lysine. Histones, which serve as scaffolding for DNA to wrap around, have a number of lysines in their N-terminal tail region; acetylating these lysines weakens the interaction between the histones and negatively-charged DNA, allowing easier access to the DNA for transcription and transcription-regulating proteins (14).

**1.2 Plant Biology**

In 1798, Thomas Malthus described the famine and disaster that would result should population overtake food supply in his *Essay on the Principle of Population*: "Famine seems to be the last, the most dreadful resource of nature. The power of population is so superior to the power in the earth to produce subsistence for man, that premature death must in some shape or other visit the human race" (15). Indeed, the world population has increased several-fold in the last century, but the Green Revolution in agriculture has staved off a Malthusian catastrophe thanks to numerous scientific developments and discoveries. For example, in the years between 1961 and 1993, increasing yields of cereal crops outstripped population growth, food prices halved, and the percentage of malnourished people fell from 35% to 21%. In India between 1961 and 1998, the population doubled, but the per capita crop production

increased 16% with a mere 6% increase in cropland acreage (16). Much of this increase is due to the development of semi-dwarf mutants of wheat, maize, and rice by plant geneticists such as Norman Borlaug. Borlaug was awarded the Nobel Peace Prize in 1970 for his efforts. Semi-dwarf crops have shorter (50-100% normal), more robust stems that resist being broken by wind, rain, or the heavy weight of high-yielding heads of grain. Additionally, the semi-dwarf crops have superior distribution of synthesized complex organic compounds between the harvested grain and the rest of the plant. Notably, these semi-dwarf plants were found to be insensitive to gibberellin signaling. Gibberellins are diterpenoid carboxylic acids, and are a class of plant growth hormone responsible for seed germination, stem elongation, leaf expansion, the start of flowering, and more (17,18). Bioactive gibberellins are present at a concentration of several nanograms per gram of plant. They are generally synthesized in growing tissues, and can be transported between different parts of the plant (19,20).

Gibberellin is not a growth activator, but a repressor of growth repressors. As seen in Figure 1.2, gibberellin binds to the nuclear protein GID1 (GIBBERELLIN INSENSITIVE DWARF 1), changing its conformation so that GID1 binds to DELLA proteins. DELLA proteins repress plant growth by sequestering growth-promoting transcription factors (TFs) from binding to DNA, sequestering transcription regulators (TRs) that prevent transcription of growth-restricting genes, and binding to growth-restricting TFs or TRs to enhance their activity (transactivation). When GID1 (with gibberellin) binds to a DELLA protein, it promotes the assembly of the SCF E3 ubiquitin ligase complex which polyubiquitylates the DELLA protein, marking it for degradation via the 26S proteasome. With lower levels of the DELLA protein, the TFs and TRs it would have sequestered are free to promote the expression of genes necessary for plant growth (18,20–23).

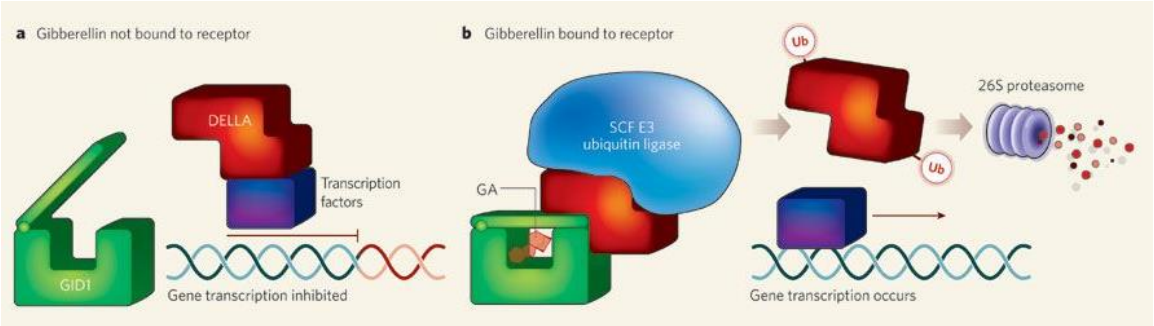


**Figure 1.2 Gibberellin regulation of DELLA proteins.** (a) in the absence of gibberellin (GA), GID1 does not readily bind DELLAs, so they are free to sequester target transcription factors. (b) in the presence of GA, GID1 binds to the DELLA, changing its conformation and allowing the SCF E3 ubiquitin ligase complex (including SLY1) to bind to the DELLA and cause its polyubiquitylation (Ub) and degradation. Figure from Hedden 2008 (23).

There are a number of mutant phenotypes in plants that interrupt gibberellin signaling. Some mutants can be restored to a normal phenotype by treating the plant with gibberellin, while some are insensitive to gibberellin treatment. GID1 is an example of the latter: in a nonfunctional GID1 mutant, the presence or absence of gibberellin does not matter, as the mutant GID1 is unable to interact with the DELLA protein to promote its ubiquitylation by the SCF E3 ubiquitin ligase complex. SLY1 (SLEEPY1) mutants are also gibberellin-insensitive, and in fact have greatly increased levels of gibberellin relative to wild-type plants. SLY1 is an F-box protein that is part of the SCF E3 ubiquitin ligase complex; non-functional SLY1 mutants prevent the complex from binding to and polyubiquitylating DELLA proteins for degradation, even in the presence of gibberellin. On the other hand, mutations in proteins in the gibberellin biosynthesis pathway, such as GA1 (ent-copalyl diphosphate synthase 1), are responsive to gibberellin treatment (18,20,21,23–25). Green Revolution mutants include *sd1*, a nonfunctional mutant of a rice gibberellin biosynthesis gene, and *TaRht-B1b* and *TaRht-D1b* in wheat, which are DELLA protein mutants lacking the GID1-binding site, so they are not proteasomally degraded and are constitutively present (17).

REPRESSOR OF GA1-3 (RGA) is a DELLA protein which was first discovered by Silverstone et al. in the lab of Tai-Ping Sun in 1997 (26). *Ga1-3* is a nonfunctional mutant of GA1; these plants produce very little gibberellin. They exhibit extreme dwarfism and do not germinate (24,26,27). Double mutants of *ga1-3* and *rga* possess a milder phenotype than *ga1-3*, wild-type-RGA plants. RGA is 587 amino acids long (64 kDa) (27). It has a C-terminal GRAS effector domain for transcription regulation and a N-terminal DELLA domain where GID1 binds. There are two N-terminal stretches rich in serines and threonines, one at the beginning of the DELLA domain and one between the DELLA and GRAS domains (18,28). These studies use a modified form of RGA (28), the sequence of which is shown in Figure 1.3. There are two N-terminal tags for purification, a poly-histidine tract (29) and a FLAG tag tract (30). In addition, there is a lysine inserted between G185 and G186 to add an additional trypsin cleavage site in the middle of the poly-S/T region to generate the new peptides SCSSPDSMVTSTSTGTQIGK and GVIGTTVTTTTTTTTAAGESTR. The modified protein has a mass of 68.4 kDa (28).

RGA is one of the DELLA proteins of *Arabidopsis thaliana*, a useful model organism for studying plant biology. It is a small plant that can be grown in petri dishes under fluorescent light, so it can be grown in large quantities in a laboratory setting. It has a short generation time, reaching maturity in 5-6 weeks (vs. 17-18 weeks for maize). It has a well-characterized and small genome, 125 million base pairs (Mb) over 5 chromosomes. Humans have 3,000 Mb over 23 chromosomes, fruit flies have 130 Mb. Maize has a larger genome than humans, at 5,000 Mb. *Arabidopsis* has been used as a model organism extensively with many resources and mutants available, making it an attractive subject for further study (31,32).

MRGSHHHHHHDYKDHDGDYKDHDIDYKDDDDKTDPMKRDHHQFQGRLSNH
GTSSSSSSISKDKMMMVKKEEDGGGNMDDELLAVLGYKVRSSEMAEVALK
LEQLETMMSNVQEDGLSHLATDTVHYNPSELYSWLDNMLSELNPPPLPAS
SNGLDPVLPSPEICGFPASDYDLKVIPGNAIYQFPAIDSSSSSNNQNKRL
KSCSSPDSMVTSTSTGTQIG(K)GVIGTTVTTTTTTTTAAGESTRSVILV
DSQENGVRLVHALMACAEAIQQNNLTLAEALVKQIGCLAVSQAGAMRKVA
TYFAEALARRIYRLSPPQNQIDHCLSDTLQMHFYETCPYLKFAHFTANQA
ILEAFEGKKRVHVIDFSMNQGLQWPALMQALALREGGPPTFRLTGIGPPA
PDNSDHLHEVGCKLAQLAEAIHVEFEYRGFVANSLADLDASMLELRPSDT
EAVAVNSVFELHKLLGRPGGIEKVLGVVKQIKPVIFTVVEQESNHNGPVF
LDRFTESLHYYSTLFDSLEGVPNSQDKVMSEVYLGKQICNLVACEGPDRV
ERHETLSQWGNRFGSSGLAPAHLGSNAFKQASMLLSVFNSGQGYRVEESN
GCLMLGWHTRPLITTSAWKLSTAAY

**Figure 1.3 Sequence of tagged RGA$^{GKG}$.** The His and FLAG tags are indicated in red; the poly-S/T regions N-terminal to the DELLA domain and between the DELLA and GRAS domains are indicated in green. The lysine inserted between G185 and G186 is parenthesized.

Two *Arabidopsis* proteins that post-translationally modify RGA are SPINDLY (SPY) and SECRET AGENT (SEC). SPY was discovered by Jacobsen and Olszewski in 1993, where mutations in SPY caused a similar phenotype to excessive gibberellin exposure. SPY was found to influence gibberellin signal transduction (33). SEC was discovered as an OGT by Hartweck et al. in 2002 (34). SPY and SEC were found to both affect *Arabidopsis* gamete development and flowering in somewhat overlapping fashions (34,35).

Previous work in the Hunt lab by Ben Barnhill, in collaboration with the Sun lab (Duke University) determined that SPY is a POFUT in *Arabidopsis,* the first nuclear POFUT discovered in any organism (12). Based on comparative genomic analysis SPY was predicted to be an OGT, but this was determined not to be the case when SPY failed to O-GlcNAcylate RGA (28). O-fucosylation of RGA promotes its binding to target transcription factors, so SPY upregulates

RGA activity. *Spy* mutants grow taller than wild-type SPY plants, as the mutation represses the repression of plant growth by RGA (12). SEC is an OGT that acts upon RGA. O-GlcNAcylation of RGA inhibits RGA binding to its target transcription factors, thus failing to sequester them and freeing them to upregulate plant growth. SEC and SPY thus have different effects on RGA activity. Zentella et al. of the Sun lab proposed that O-fucosylation may promote RGA adopting an active conformation, whereas O-GlcNAcylation may promote an inactive conformation. The active configuration promotes the ability of the GRAS domain to interact with the transcription factors it sequesters, like BZR1 and PIFs (12,28).

A study by Xu et al. in 2017 (36) identified 971 O-GlcNAcylated peptides across 262 *Arabidopsis* proteins. Many of these proteins were involved in the regulation of protein expression (transcription, translation, and chromatin remodeling), hormone signal transduction (gibberellin, auxin, etc.), stress response, and flower development. SPY was one of the O-GlcNAcylated proteins. However, they did not achieve full coverage of the O-GlcNAcylated proteome; they did not detect any O-GlcNAcylated DELLA proteins, including RGA (36).

Studies have shown that O-GlcNAc and phosphorylation are linked modifications. In some cases, a decrease in one correlates with an increase in another. Wells et al. demonstrated that a quaternary complex of a phosphatase and an OGT can remove a phosphate and then add a GlcNAc in concert (37). Certain residues on some proteins can be modified with either PTM. Kinases and phosphatases can be O-GlcNAcylated, and OGTs and OGAs can be phosphorylated (7–9). There have been fewer studies of O-GlcNAcylation than phosphorylation in plants. Martínez-Turiño et al. examined phosphorylation and O-GlcNAcylation in the Plum Pox Virus, which infects plants. The viral proteins are transcribed,

translated, and post-translationally modified in plant cells – including O-GlcNAcylation by SEC. The viral capsid protein can be simultaneously phosphorylated and GlcNAcylated, including at adjacent amino acids (gT24 and pS25). The researchers evaluated O-GlcNAc and phosphorylation levels in tobacco plants with normal and deficient levels of SEC, and found that levels of pS25 doubled in SEC-deficient plants compared to wild-type (38). It has been found that in rice, the casein kinase I EARLIER FLOWERING1 (EL1) promotes lower gibberellin response; EL1 inhibits gibberellin signaling by increasing the stability and activity of the DELLA protein OsSLR1 (39).

Acetylation has been studied in *Arabidopsis* previously. *Arabidopsis* has 16 lysine acetyltransferases and 18 lysine deacetylases (40–44). Hartl et al. in 2017 looked at substrates of deacetylases and found 91 acetylated non-histone proteins. The deacetylases of *Arabidopsis* affect the timing of flowering, seed germination, hormone-related stress responses, and ribulose-1,5-bisphosphate-carboxylase/oxygenase activity, which is important in fixing $CO_2$ in photosynthesis. Many of the identified proteins were located in the nucleus and plastids (such as chloroplasts), and had functions including transcriptional control, gluconeogenesis, photosynthesis, the TCA cycle, and tetrapyrrole synthesis (40). In other organisms acetylation and phosphorylation have been found to occur simultaneously on a protein and affect each other (45,46), though very little research on such interactions in plants have been performed (44). RGA acetylation has not been observed before.

### 1.3 Proteomics

Genomic research determines the DNA sequence of an organism, including all of its genes. The *Arabidopsis thaliana* genome was completed in 2000 (47), and the human genome

in 2004 (48). Genomics lays the groundwork for proteomics, which is the study of the large, complicated array of proteins expressed in a given cell or organism simultaneously or over time (49). There are around 27,000 *Arabidopsis* genes (32), but between alternate splicing of genes and a vast variety of PTMs (oftentimes occurring at different sites in different combinations at different times on a given protein), there are more than 600,000 proteoforms (1,50).

### *1.3.1 Protein Isolation*

Given the huge complexity of a proteome, separation of the component proteins is usually necessary prior to analysis. Electrophoretic separations are often used to separate proteins in a complex mixture according to charge in an electric field. Such techniques include polyacrylamide gel electrophoresis (PAGE) and isoelectric focusing; these can be combined in orthogonal directions as 2D electrophoresis to allow for separation along two different axes. Bands containing a protein of interest may be excised. However, the separation and excision may not isolate a single protein (1).

Proteins may also be separated chromatographically according to their characteristics. High-pressure liquid chromatography (HPLC) uses solvent (the mobile phase) flowing over a column packed with beads (the stationary phase) under pressure. Reverse-phase HPLC uses a hydrophobic stationary phase such as beads coated with aliphatic C18 chains, and a hydrophilic mobile phase such as 0.5% acetic acid in water. The composition of the mobile phase increases in hydrophobicity as the HPLC gradient proceeds. Protein hydrophobicity varies according to the amino acid composition; proteins with lower hydrophobicity (such as ones with many polar or charged residues) will have low affinity for

the stationary phase and will thus elute from the column earlier than hydrophobic proteins. Liquid chromatography can be used for both preparative purification of proteins and analysis of proteins when linked to an instrument such as a mass spectrometer (1).

Affinity chromatography is used to enrich for certain proteins based on a specific marker, the ligand. Ligand-bearing protein will bind to the affinity column preferentially while other proteins will not bind and will be washed off the column. The isolated target protein is then eluted. Antibodies are often used, as they recognize and bind to a specific antigen (the ligand), such as a specific region on a specific folded protein (1). Anti-PTM antibodies are either specific for a modification at a particular site (recognizing the amino acids surrounding the PTM as well as the PTM itself), or with broader specificity (recognizing the PTM, with less amino acid sequence specificity). However, the specific antibodies only recognize a known sequence and site, and the broader antibodies have lower binding affinity and specificity (6). Other types of affinity chromatography include nickel(II)-bearing columns with an affinity for histidines (29) and immobilized-metal affinity chromatography with iron(III), which has a high affinity for the phosphate group found on phosphoproteins (51).

### *1.3.2 Proteases*

Proteins are often digested by proteases into peptides prior to analysis. This may be performed before, after, or between separation(s), but prior to HPLC into a mass spectrometer. A number of different proteases are commonly used, with trypsin being the one most widely employed. Trypsin cleaves C-terminal to arginine and lysine residues (but not if the arginine or lysine is followed by a proline, or if the lysine is acetylated (52)). Trypsin produces peptides that are amenable to mass spectrometric analysis (described below)

because the C-terminal residue is basic and thus frequently protonated under common electrospray ionization (ESI) conditions. When the peptide is fragmented by collision-activated dissociation (CAD), both the N-terminal (from the amine group) and C-terminal (from the R/K) fragments will bear at least one positive charge, and ions from both ends will thus be detectable. The same is true for electron-transfer dissociation (ETD) fragments from peptides with at least 3 charges (3,53). CAD and ETD are described below in section 1.4.4. Contrariwise, other methods analyze intact proteins. In between these two are limited digestion techniques, which use more selective proteases to generate larger peptide fragments (54).

Once peptides have been separated, they are identified by determining the amino acid sequence. Multiple methods exist for sequencing proteins. In 1950 Pehr Edman described an eponymous method involving sequential phenylthiocarbamylation and acid cleavage of amino acids from the N-terminus of a peptide, followed by chromatographic comparison of each cleaved amino acid derivative to standards (55). This process was laborious and restricted to short peptides (1).

Mass spectrometry is a superior method for proteomics studies, including the profiling of PTMs and their site-localization. Mass spectrometry has a high level of sensitivity (down to the level of attomoles) and accurate mass determination, especially with higher resolution instruments. Tandem mass spectrometry (MS/MS) is used to sequence proteins or peptides. For PTM analysis, it is best to analyze purified proteins (whether pre-separated or with HPLC-MS; shotgun proteomics for PTMs requires more material and instrument runs). Protein sequence coverage is variable, with more abundant proteins usually having more identified peptides (56). Genome sequencing made large scale proteomic mass spectrometry

possible. For example, in 1994 Eng, McCormack, and Yates used MS/MS to identify peptides from whole cell lysates of *E. coli* or yeast based on protein sequence databases of each species (57).

## 1.4 Mass Spectrometry

Mass spectrometry is an analytical method which measures the mass to charge (m/z) ratio of ions in a sample. There are several steps to mass spectrometric analysis: the sample must first be ionized and introduced into the instrument, the ions must be routed to the mass analyzer (which measures the m/z of the ions), analyzed, and detected. In a tandem mass spectrometry experiment, ions of a certain m/z are selected and isolated from the others, and then subjected to fragmentation. The fragments are then analyzed in a similar fashion to the original ion. The original ion is called the precursor, and the fragment ions are called product ions (58).

### *1.4.1 Ionization Methods*

Electron impact ionization, developed in the 1940s by Alfred Nier, is a "hard" ionization technique that bombards the analyte with a stream of electrons (10s of electron volts (eV)); this energy is high enough to fragment ions in addition to ionizing them prior to entering the mass spectrometer. Chemical ionization was developed by Frank Field and Burnaby Munson in 1966 as a technique which uses electrons to bombard and ionize methane instead of the analyte itself; the methane ions then ionize the analyte. This causes less fragmentation than electron impact ionization, such that the intact precursor ion is observable. However, these techniques are only suitable for analyzing volatile, low molecular

weight species and thus not usable for analyzing peptides. Fast atom bombardment (FAB) is a soft ionization technique developed in 1981 by Michael Barber et al. In FAB, a stream of argon atoms (keV) bombards a solid or liquid sample on a copper substrate, sputtering the analyte into the gas phase while ionizing it. FAB is able to ionize larger molecules, including small proteins (59–61).

In the current era, the techniques of matrix-assisted laser desorption ionization (MALDI) and electrospray ionization (ESI) have become the predominant techniques for analyzing proteins. Both ESI and MALDI are capable of ionizing proteins in the tens to hundreds of kilodalton mass range. MALDI was developed by Michael Karas and Franz Hillenkamp, and a similar technique by Koichi Tanaka (62,63). ESI was first used in mass spectrometry by John Fenn and Masamichi Yamashita in 1985 (64). In 2002 John Fenn and Koichi Tanaka were awarded half of the Nobel Prize in Chemistry. In MALDI, the sample is added to a solution with a matrix (usually a low-mass aromatic acid) that can absorb ultraviolet light. The solution is deposited on a metal plate where it dries and co-crystallizes into a spot. The spot is then irradiated with pulses of an ultraviolet laser (typically 337 nm or 355 nm in wavelength), which energizes the matrix so it sublimes (along with the analyte molecules) into the gas phase. The matrix transfers a proton to the analyte molecule and an electric field passes the analyte ion (typically singularly-charged) into the mass spectrometer. MALDI is usually performed with a time-of-flight mass analyzer, and is not suitable for direct interfacing with a constant flow of sample from an LC (53,62,63,65).

ESI is the technique used in the studies presented here, and will be explained in more detail (see Figure 1.4). ESI functions at atmospheric pressure and can be linked to direct flow from a chromatographic column, as it generates ions from the liquid phase. The dilute

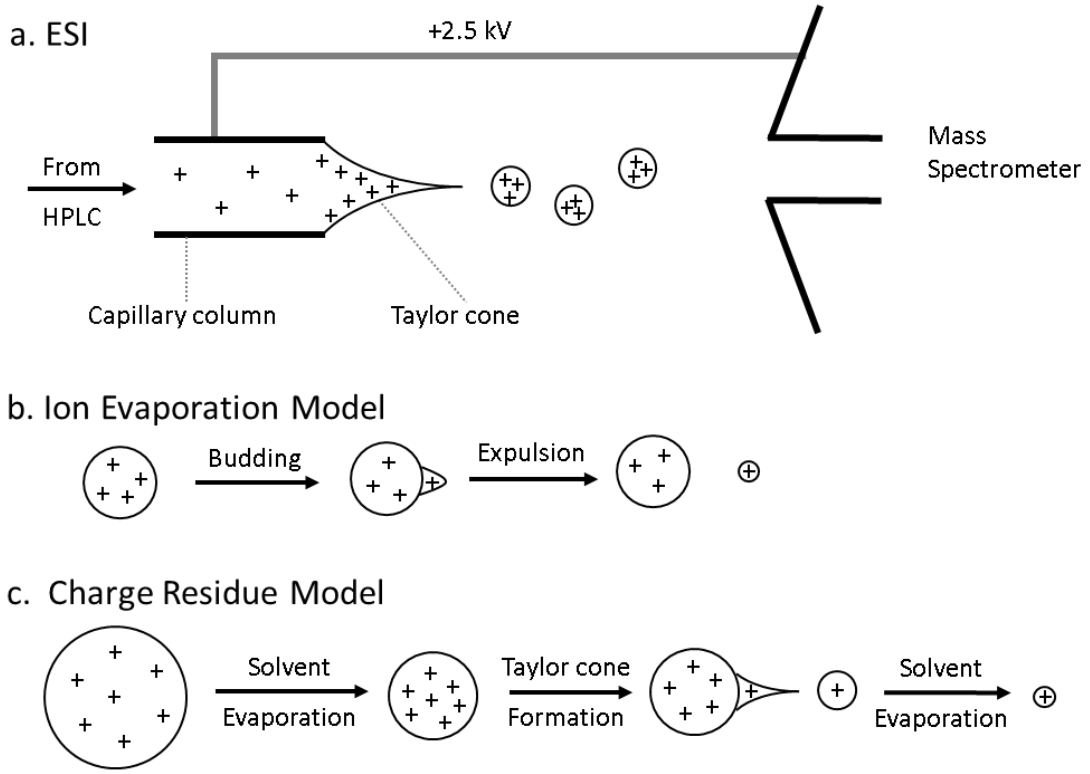


**Figure 1.4 Electrospray ionization.** (a) Sample analyte ions at the end of the capillary column are attracted by a voltage difference towards the mass spectrometer inlet, and form a Taylor cone that droplets bud off from. (b) Ion evaporation model. (c) Charge residue model.

solution of sample flows from the column at a low flow rate under pressure from an HPLC. In ESI, a differential voltage (+2-5 kV) is applied between the instrument inlet and the column (in the present experiments, the waste line from the Y splitter of the HPLC flow). Cations are attracted towards the instrument inlet and accumulate at the end of the column; analyte proteins in an acidic solution are easily protonated at basic residues such as lysine and arginine, and a higher abundance of protons (such as at the column tip) enhances this. When enough charges accumulate at the end of the column, the Coulombic repulsion between the charges matches the surface tension of the solution (the Rayleigh limit), a Taylor cone forms, and droplets bud off the tip of the cone. As the droplets travel through the air and then the heated capillary at the front of the instrument, the solvent evaporates, approaching the

Rayleigh limit again and causing a new Taylor cone to form on the droplets, leading to generation of ever smaller droplets. There are two theories of how the analyte ions become isolated in the gas phase; the ion evaporation model and the charge residue model (Figure 1.4b and 1.4c, respectively). The ion evaporation model states that at very small droplet sizes (~20 nm), solvated ions are expelled from the droplet; this model describes the behavior of smaller analytes well (less than 1000 Da). The charge residue model states that each sufficiently small droplet (~200 nm) sprayed from a Taylor cone contains only one analyte ion which is desolvated by evaporation; this model describes the behavior of larger analyte ions well. At larger ion sizes hydrophobic peptides ionize more efficiently than hydrophilic ones, and at high concentrations the former may suppress the signal of the latter. Thus, phosphorylated and glycosylated peptides tend to be suppressed by their unmodified counterparts, though this is less pronounced with nano-ESI (as used in this work), which uses very small capillary diameters (a few µm at the tip) and solvent flow rates (~100 nL/min). ESI produces multiply charged analyte cations (66–69).

### *1.4.2 Mass Analyzers*

The quadrupole ion trap (QIT, or 3D trap) was developed by Wolfgang Paul and Helmut Steinwedel in 1960. Paul and Hans Dehmelt won half of the 1989 Nobel Prize in Physics (alongside Norman Ramsey) for the development of the ion trap technique. The 3D trap uses a quadrupolar field to trap ions. It has three electrodes, all hyperbolic: two endcaps and one central ring. $r_o$ is the inner radius of the ring electrode and $2z_o$ is the distance between the apexes of the endcaps. In theory $r_o^2 = 2z_o^2$, though 3D traps are usually manufactured with an increased $2z_o$. An RF potential is applied to the ring electrode and the endcaps are

grounded to give a quadrupolar field that generates a parabolic potential well (described by Dehmelt) in which to confine the ions. Axial resonant excitation is produced by applying a supplemental AC potential (100s of mV) across the endcaps. Ions are injected into the trap and cooled via collisions. Axial excitation of an ion at its secular frequency imbues it with kinetic energy, so it oscillates in the chamber at greater velocity and moves further along the z axis (60).

Linear ion traps (LITs) were developed by John Syka and colleagues and described in their 2002 article (70). The linear ion trap is analogous to a 3D trap that has been stretched lengthwise. As seen in Figure 1.5, The LIT consists of four hyperbolic rods arranged parallel to each other (4mm apart in the y direction, 5.5 mm apart in the x direction). The four rods are divided into three sections: two shorter end sections (12 mm long) and a longer center section (37 mm; where the ions are trapped). The DC voltages on the three sections are changed at different times to admit, trap, and mix ions along the z axis as desired. The presence of the end sections greatly reduces fringe field effects, keeping the applied fields axially consistent for the full length of the center section in order to minimize undesired axial ejection. The trap contains a helium bath gas (approximately $10^{-5}$ Torr) which helps trap injected ions through kinetic cooling via collisions. A quadrupolar field is applied to the rods to trap ions. Opposite rods form pairs: the two rods in the x-plane form a pair, and the two rods in the y-plane form another pair. The x rods have narrow slits (0.25 mm, 30 mm long) in the center through which ions are ejected for detection. When a quadrupolar RF field is applied to the rods in order to trap ions radially, each pair is 180 degrees out of phase with the other (70).

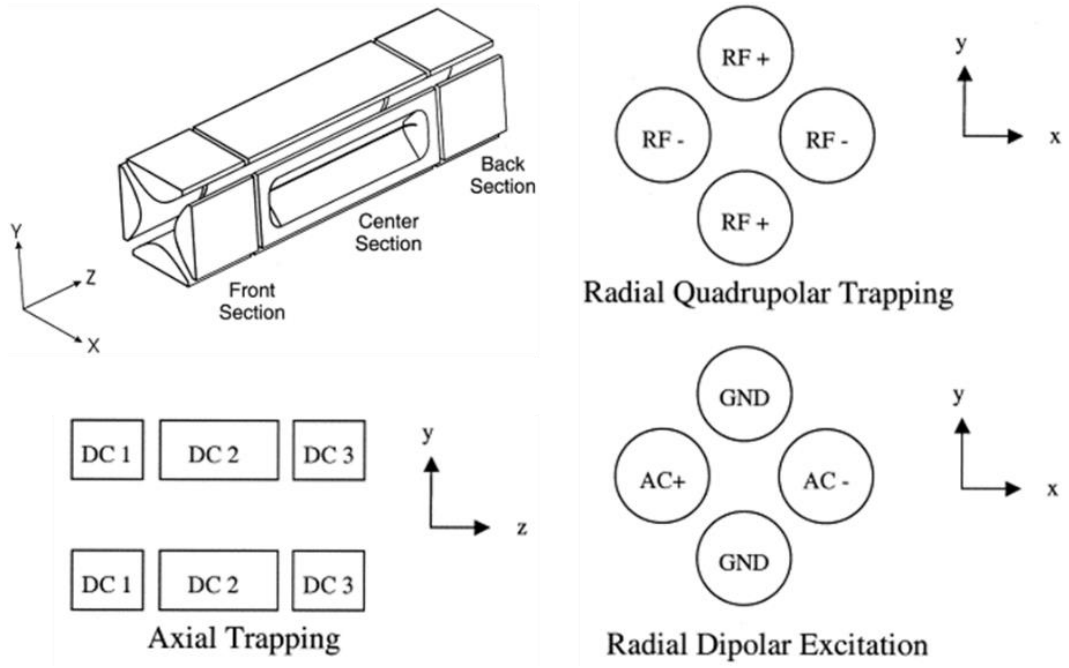


**Figure 1.5 Linear ion trap diagram and potentials in operation.** Ions are trapped axially with DC potentials on the front, center, and back sections of the four rods. Ions are trapped radially with a quadrupolar field generated by RF potentials on each pair of rods (the x-rods 180 degrees out of phase from the y-rods, represented by + and -). Ions are excited with a supplemental AC applied across the x-rods for isolation, activation, or ejection. Figure from Schwartz 2002 (70).

The Mathieu equation (Equation 1.1) describes the motion of an ion in a quadrupolar field, such as those generated in a 3D trap or LIT. As the RF potential on the LIT rod pairs varies sinusoidally in time, the ions are alternately attracted and repelled from the rods. Positive ions are attracted to the negative pair and repulsed from the positive pair, but the amplitude and polarity of the charges on each rod pair changes sinusoidally over time, so ions orbit in circles in the LIT with a certain frequency, called the secular frequency (ω; see Equation 1.2). In the Mathieu equation, ξ is proportional to the product of Ω (the angular frequency of the trapping RF potential) and time. u is the displacement of the ion from the center of the field, and $a_u$ and $q_u$ are the Mathieu parameters for an ion with displacement u (in the x, y, z, or r direction). For a LIT where the x-rods spacing and the y-rods spacing are identical, u is defined

radially (r) from the center of the trap, and the axial potential component is negligible due to the three axial sections of the trap. Thus, the values of *a* and *q* are given in Equation 1.3 below (70,71):

Equation 1.1: $\frac{\mathrm{d}^2\mathrm{u}}{\mathrm{d}\xi^2} + (a_u - 2q_u\cos(2\xi)) * \mathrm{u} = 0$ , ξ= Ωt/2

Equation 1.2: $\omega \approx \frac{\Omega}{2}\sqrt{a_u + \frac{q_u^2}{2}}$

Equation 1.3: $a_x = -a_y = \frac{8\mathrm{zU}}{\mathrm{m}\Omega^2\mathrm{r_o^2}}$ , $q_x = -q_y = \frac{-4\mathrm{zV}}{\mathrm{m}\Omega^2\mathrm{r_o^2}}$

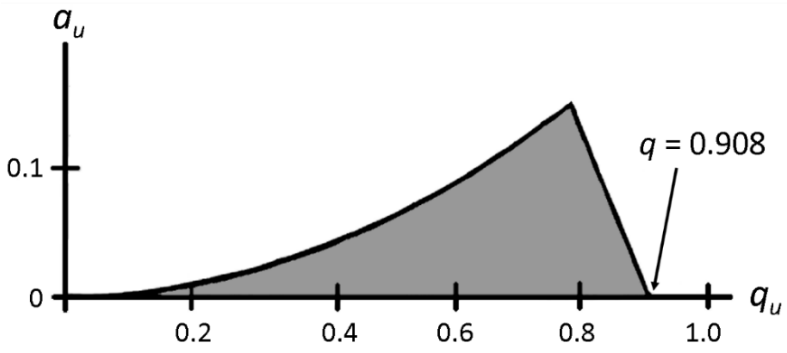


**Figure 1.6 Mathieu stability diagram.** *a* is proportional to the DC voltage, *q* to the RF AC voltage. Ions are stable when within the shaded area. When the DC voltage is 0, the ions fall along the x-axis. At *q* = 0.908, the ions become unstable and are no longer trappable. Figure adapted from March 1997 (60).

The $a_u$ parameter is proportional to U (the DC potential applied to the rods in a quadrupolar manner), and the $q_u$ parameter is proportional to V (the RF AC 0-to-peak potential). z is the charge of the ion, and m is the mass of the ion. The regions of ion stability are given in a Mathieu stability diagram, with *a* and *q* as the axes (Figure 1.6 shows the stability region near the origin, which is the one used in mass spectrometry). Ions within the stable region have stable trajectories and can be trapped; ions outside the stable region have unstable trajectories and will be ejected or expire on the rods. As seen in Equation 1.3, the *a* and *q* parameters are inversely proportional to the m/z of an ion. In the linear ion trap, the DC potential is set to 0, so *a* = 0. Which ions are stable is thus only determined by the RF

potential. The stable *q* values range from 0 to 0.908, though practically the lower limit is about 0.1. Since *q* is inversely proportional to the m/z of an ion (Equation 1.4), an ion of a higher m/z will have a lower *q* in a field of a given RF potential V. The secular frequency of the ion is approximately inversely proportional to its m/z (Equation 1.5) (60,71).

Equation 1.4: With $a$ =0, $\frac{m}{z} = \frac{4V}{q\Omega^2 r_0^2} = \frac{V\sqrt{2}}{\omega\Omega r_0^2}$

Equation 1.5: $\omega \approx \frac{\Omega}{2}\sqrt{0 + \frac{q^2}{2}} = \frac{\Omega q}{2\sqrt{2}} = (\frac{z}{m})(\frac{V\sqrt{2}}{\Omega r_0^2})$

By applying a supplemental AC at a given frequency in the x-axis, ions of the corresponding m/z (secular frequency = supplemental AC frequency) are resonantly excited and gain kinetic energy. The excited ions may be subjected to collision-activated dissociation (CAD) fragmentation or ejected through the slits in the x-rods to be detected. Ejection requires more energy and thus a larger supplemental AC voltage. Alternatively, a given species may be isolated by ejecting the other species via applying resonance excitation over a wide band of frequencies, with the exception of a narrow band around the secular frequency of the desired ion. Schwartz 2002 describes the use of a 5-500 kHz band with a gap for the desired ion at *q* = 0.83. The isolated ions are then activated with a single frequency, ~1V resonance excitation at an activation *q* of 0.25-0.35. A *q* of 0.25, as used for the experiments in this work, gives sufficient time (due to trapping) and energy for collisions while allowing a reasonable low mass cut off (LMCO) of ions detectable, which is determined by *q*. At lower *q* the LMCO for a given precursor m/z decreases, but trapping and fragmentation efficiency also decrease. Then, to eject ions to the detectors for mass analysis, the RF voltage is increased to bring ions of increasing m/z to *q* = 0.88, where they are resonantly excited by a supplemental AC applied to the x-rods and ejected to the detector. The supplemental AC

essentially creates a “hole” in the Mathieu stability diagram at $q$ = 0.88 where ions “fall through” and are ejected. As the RF voltage V increases linearly, the m/z of ions reaching $q$ =0.88 also increases linearly in a purely quadrupolar field, according to Equation 1.3. Thus, the ion intensity is measured over time, and a specific time is linearly correlated with a specific RF voltage and thus a specific m/z. In order to improve resolution, the ion cloud distribution in space is reduced via collisional cooling with the bath gas (70–72).

Linear ion traps have several advantages over older, 3D traps. The larger volume allows more ions to be trapped before severe space charging occurs. Storage space charging (how many ions can be stored at once) is greater than activation space charging (how many ions can be stored while activating select ions). These are both greater than the isolation limit (how many charges can be isolated), which is greater than the spectral limit (how many charges can be isolated and still produce a good spectrum). These values can vary over several orders of magnitude. A linear ion trap may hold 10x or more charges than a comparable 3D trap. Thus, one scan can offer good ion intensities and resolution, and the sensitivity and limit of detection is superior. LITs have more stringent manufacturing precision requirements than 3D traps. Linear ion traps have higher trapping efficiencies than 3D traps (~5% vs. up to ~100%), since ions entering 3D traps must face an RF field along the axis of injection, whereas in LITs they only effectively face the DC differential of the center section and endcap. Also, it is easier to trap a wide m/z range with a single RF amplitude with a LIT vs. a 3D ion trap, so the instrument may inject ions into the trap with a single injection RF (corresponding to a 100 m/z LMCO) and simultaneously trap ions of 150-2000 Da. Finally, the LIT is more readily connected in line with another mass analyzer, such as an Orbitrap (70,73).

The Orbitrap mass analyzer was developed by Makarov in 1999 (74). Orbitraps are high-resolution mass analyzers/detectors. As seen in Figure 1.7, the Orbitrap consists of two outer electrodes with a cup-shaped inner surface and a coaxial, spindle-shaped inner electrode. It is radially symmetric about the spindle axis. Ion frequencies are measured in a non-destructive manner, via the image current they induce on the outer electrodes. The Orbitrap has a quadro-logarithmic field due to the DC potential between the outer electrodes and inner spindle electrode. The Orbitrap has high resolution (up to 1 million), good dynamic range (1:5000), and high ion storage capacity. It is derived from the Kingdon trap, the Paul (3D) trap, and the FTICR. The Kingdon trap uses a purely DC field (between a central wire and surrounding cylindrical electrode; there are also endcap electrodes) to trap ions in a radial logarithmic field. In 1981 Knight introduced a split into the cylindrical outer electrode through which to introduce ions. This altered trap could use axial resonant excitation to eject trapped ions to an outside detector, or monitor image current between the split electrodes. For the Orbitrap, the central wire was replaced with a spindle electrode, and the outer cylinder was replaced with two cup-shaped outer electrodes (75–77).

The outer electrodes of the Orbitrap are used to create the trapping field (with a potential difference relative to the spindle electrode) and then detect the image current of oscillating ions. Ions in an Orbitrap undergo simple harmonic oscillation along the z axis, driven only by a DC voltage difference between the spindle and outer electrodes and the Orbitrap geometry. The ions also orbit the spindle radially (r), but the potential and thus movement of ions in the axial (z) and radial directions are independent of each other. If injected perpendicular to the spindle axis, the ions start with little to no axial velocity, which allows for more accurate and precise determination of the oscillation frequency (which is

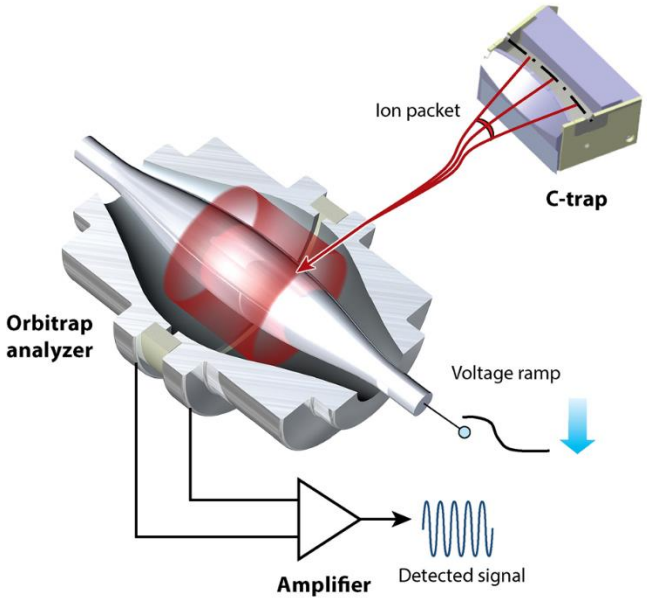


**Figure 1.7 Orbitrap mass analyzer and C-trap.** The C-trap focuses an ion packet and injects it into the Orbitrap perpendicular to the z-axis (the spindle axis) at an offset from the center. The ions orbit around the spindle and oscillate back and forth along the z-axis over time (the ion path is represented in red). The axial oscillations cause an image current detected by the two outer electrodes over time to produce the detected signal. Figure from Eliuk 2015 (76).

inversely related to the square root of the m/z, see Equation 1.6). The ions are injected at a z-axis offset so that they start oscillating axially about the spindle, without the need for additional excitation. The axial oscillations are detected via the image current on the two outer electrodes (which are split at z = 0). The signal is converted from analog to digital, and deconvoluted from the time domain via Fourier transformation to give the ion frequencies and hence m/z values (75–77).

Equation 1.6: $\omega = \sqrt{(\frac{z}{m}) * k}$

Ions are accumulated for injection into the Orbitrap in the C-trap. The C-trap is a bent RF-only quadrupole containing 1 mTorr of nitrogen as a bath gas which collisionally cools the ions into a tighter packet. The ions are then focused and transferred into the Orbitrap, where the DC voltage across the spindle and outer electrodes is ramped to electrodynamically squeeze the radius of the ion packet and its radial distance from the spindle before the voltage is set and image current detection commences. This squeezing, in combination with

the later arrival of larger m/z ions, causes larger m/z ions to orbit at a larger radial distance. Ions need to be injected into the trap with the right amount of kinetic energy to move in a nearly circular orbit in about the spindle, rather than an eccentric orbit. Ions of the same m/z ratio will dephase radially but much more slowly in the axial direction, forming a thin ring around the spindle that moves back and forth axially – it is this axial movement that is detected by the outer electrodes and which corresponds to the m/z. After the initial injection occurs and the orbits stabilize, the image current detection begins (75–77).

### *1.4.3 Orbitrap Fusion*

The mass spectrometer used in this research is an Orbitrap Fusion™ Tribrid™ Mass Spectrometer from Thermo Fisher Scientific (78). This instrument accepts ions via ESI. It possesses a quadrupole mass filter, an Orbitrap mass analyzer, and two sequential linear ion trap mass analyzers (a high-pressure cell and a low-pressure cell). As shown in Figure 1.8, the linear ion traps are located beyond the C-trap which directs ions into the Orbitrap (78).

Electrosprayed positive ions are directed through the instrument with increasingly negative voltages (0 to -30 V). Ions enter the instrument through a heated capillary to the S-lens (a stacked-ring ion guide), where they are focused into a tight beam. At the end of the S-lens is a discharge ion source for calibration or fragmentation (this is the source of ETD reagent ions). Ions then proceed through the active-beam guide (a curved quadrupole with axial DC field), which directs ions around a 90-degree bend. Neutral molecules do not curve in the electric field and so are pumped away by the vacuum system rather than proceeding deeper into the instrument. The ions are fed into the quadrupole mass filter, which like the LIT has four parallel rods. The QMF can be operated in either AC-only mode (which allows for

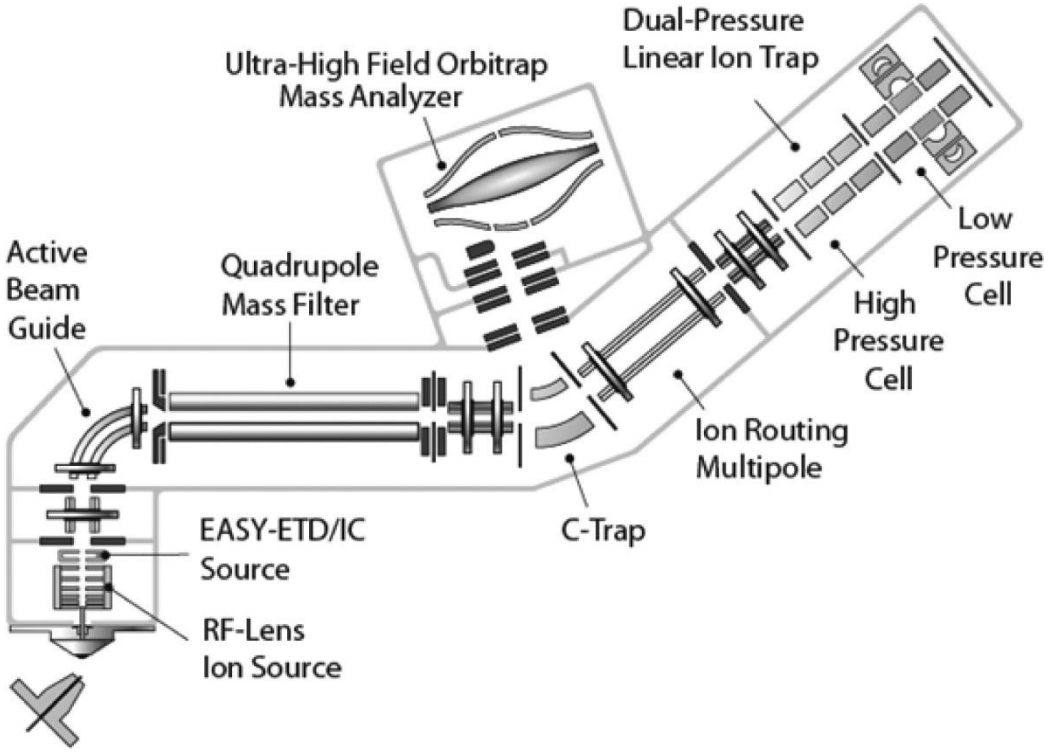


**Figure 1.8 Layout of the Orbitrap Fusion™ Tribrid™ mass spectrometer (Thermo Fisher Scientific).** Ions enter the instrument inlet at the bottom left, and are routed through the instrument with voltage differentials. The instrument has a "T" architecture, with the Orbitrap branched off at the C-trap prior to the IRM and dual LITs. Figure from Senko 2013 (78).

transmission of a large window of ions, such as 150-2000 m/z), or in AC and DC mode to allow a specific range of ions through (as wide as 1200 m/z or as narrow as 0.7 m/z). Ions are trapped radially with RF AC voltages on the multipole rods. The AC defines the center of the window, and the DC/AC ratio determines the width of the window (78).

The QMF leads to the C-trap. From there the ions enter the ion-routing multipole (IRM), where they are accumulated by trapping them with 5-10 mTorr of nitrogen gas between positively-charged endcaps. Collisions with the nitrogen slow the ions down and aid the electric fields in trapping them. The analyte species may be fragmented in the IRM during accumulation. The precursor or fragment ions may be routed to the LITs in the rear of the

instrument or back through the C-trap to the Orbitrap (routing via an axial DC field). The Orbitrap has a -5 kV differential applied to its central electrode to inject ions into the Orbitrap. The ions follow a Z-shaped path to minimize nitrogen transfer into the low-pressure ($10^{-10}$ Torr) Orbitrap. Senko reported a resolution of 120,000 at m/z 200 with a 256 ms transient in the Orbitrap. The Fusion has two LITs in a dual cell arrangement (at $1e^{-5}$ and $7e^{-5}$ Torr). MS/MS product ions are scanned at 66.6 m/z units per ms. CAD and ETD fragmentation are performed in the LITs. The back, low-pressure LIT detector has two high-energy dynodes that when struck by ions produce electrons which are directed to one electron multiplier. The LIT traps ions regardless of charge polarity, allowing for ETD with reagent radical anions and peptide cations. (78). The front, high pressure LIT is used for isolating and fragmenting ions, while the rear, low-pressure cell is used for mass analysis of ions (76).

The Orbitrap Fusion possesses several time-saving advantages. QMF isolation removes the need for waveform-based isolation in the LIT. The Fusion has a high-field, compact Orbitrap and is operated with eFT (enhanced Fourier Transform), which takes into account the off-axis injection of ions into the Orbitrap (which causes a type of phase shift) and causes a virtual doubling of the effective transient time. The compact Orbitrap and eFT allow for shorter transients to achieve the same resolution. Most notably, the T-shaped configuration of the instrument allows parallelization of analysis. The instrument is able to perform a high-resolution Orbitrap full scan while simultaneously isolating and accumulating ions in the IRM, while the LIT acquires a tandem MS scan. This reduces the instrument cycle time and allows for more rapid analysis and deeper analysis over a given time period. The parallelization requires more sophisticated software, which also allows for more complicated experiments than previous mass spectrometers, such as a top-speed data-dependent analysis

with multiple selection parameters vs. a set number of MS2s per cycle (such as the top 10 most abundant ions) (76,78).

Automatic gain control (AGC) is used to prevent space-charging and its deleterious effects. A packet of ions is isolated very briefly (<1 ms) in the IRM and then injected into the low-pressure LIT to determine this number of charges. This occurs before every scan, and is used to calculate how long ions should be injected into the trap to reach a certain number of charges. The instrument acquires a high-resolution MS1 scan in the Orbitrap, producing a spectrum of the detected ions and the intensity of each. The instrument selects one of the species for fragmentation (in a 2 Da window centered around the monoisotopic peak, as determined by monoisotopic peak selection [MIPS]). That species (which is still entering the instrument via ESI through its inlet) is isolated by the QMF and accumulated in the IRM. The ions may be fragmented by CAD or ETD in the low-pressure LIT or HCD in the IRM; simultaneously, the Orbitrap is acquiring the next MS1 spectrum. One instrument cycle encompasses one high-resolution MS1 in the Orbitrap followed by a number of MS2 fragmentations and low-resolution scans. The cycle can be designed to acquire a set number of MS2s per cycle (e.g. the top 10 most abundant ions) or as many scans as possible in a set amount of time (e.g. the top most abundant ions in 3 seconds) (76,78). Dynamic exclusion is used to prevent repeated selection of high-intensity precursors in order to reach lower-intensity precursors in the sample. In dynamic exclusion, a precursor ion that has been selected for fragmentation x times within a set time period is added to an exclusion list, so that it is not selected and fragmented for another set period of time (54).

### *1.4.4 Fragmentation Methods*

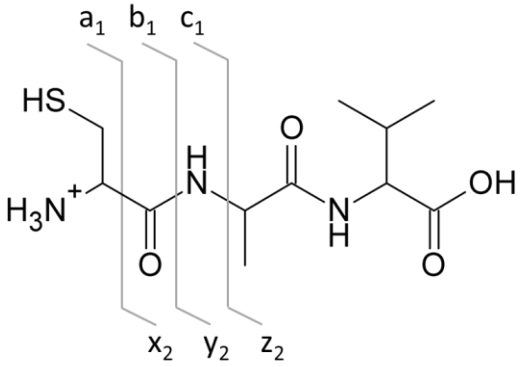


**Figure 1.9 Fragmentation modes of a peptide.** CAD produces b and y ions, ETD c and z· ions. a, b, and c ions are numbered from the N-terminus; x, y, and z ions are numbered from the C-terminus.

Mass spectrometry was first used to sequence peptides in 1959 by Klaus Biemann by reducing the carbonyl and cleaving the bond between the $C_\alpha$ and $CH_2$ (the reduced C=O) (79). Donald Hunt and his lab were the first to sequence peptides using tandem mass spectrometry in 1981, with collision-activated dissociation in a triple quadrupole (80). Peptides can be fragmented at different points along their amide backbone, producing complementary pairs of ions (a and x; b and y; c and z). a, b, and c ions originate from the N-terminus of the peptide; x, y, and z ions originate from the C-terminus. An ion is numbered according to how many residues it encompasses from its respective terminus. These fragmentation sites are shown in Figure 1.9. Which types of ions are produced depends on the fragmentation method (53).

Collision-activated dissociation (Figure 1.10) was developed by Keith Jennings in 1968 (81), building on previous work by Futrell, Miller, Shannon, and McLafferty (59). In CAD, the isolated precursor peptide is trapped in an ion trap with a bath gas, usually He or $N_2$. The precursor ion is resonantly excited by a supplemental AC at its secular frequency, increasing its kinetic energy and velocity so it collides repeatedly with the bath gas (71). The many collisions gradually build up vibrational internal energy. This vibrational energy is distributed throughout the bonds of the peptide, and after hundreds of collisions enough energy has accumulated that one of the weakest bonds breaks. In the peptide backbone, this is the peptide bond (the amide bond, carbonyl carbon to amide nitrogen) with an additional mobile

proton. This fragmentation produces b and y ions. The fragmentation of thousands of peptide molecules, at random peptide bonds along the backbone, creates b and y series of ions. Two sequential ions in a series are separated by one amino acid. Each amino acid has a characteristic mass. Modified amino acids also have an indicative mass (167 for phosphoserine, for example). Thus, by subtracting the masses of two sequential ions, one determines the amino acid represented; doing this for a full series reveals the sequence of the peptide. The b and y series are mirror images and confirm each other. (53,82).

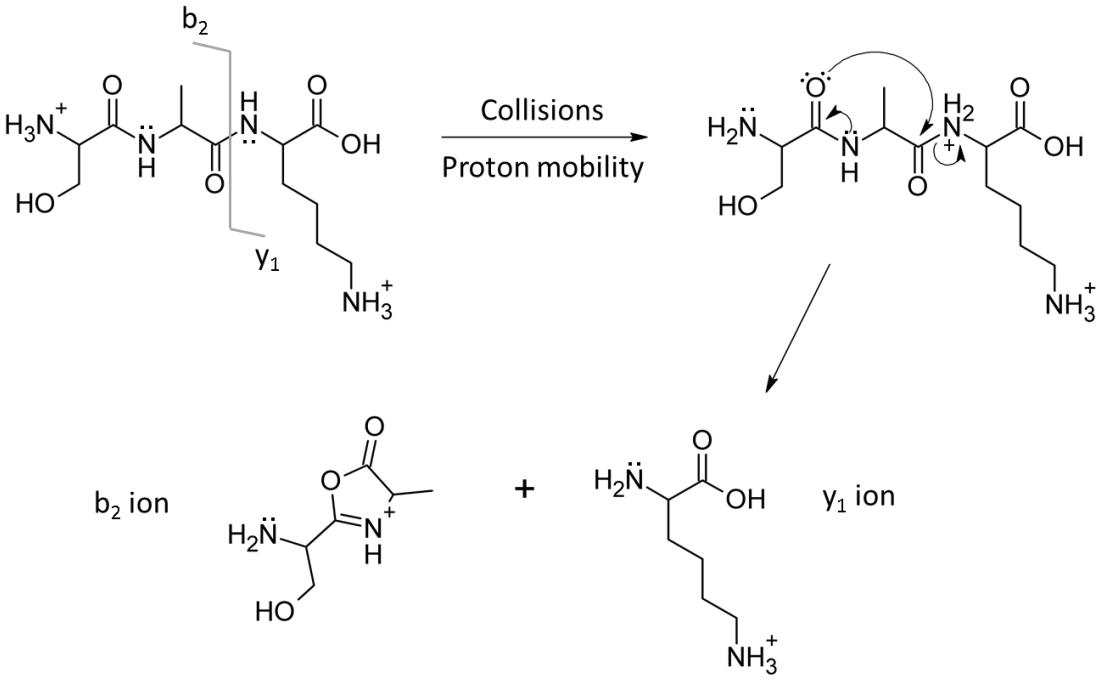


**Figure 1.10 Collision-activated dissociation (CAD) mechanism.**

CAD backbone cleavage occurs according to the mobile proton model. In this model, fragmentation is charge-directed, as fragmentation requires a proton at the cleavage site. Protons that are normally localized at basic sites (such as the N-terminus, Arg, or Lys) are freed and are able to move along the peptide backbone as the precursor ion gains energy from the repeated collisions. Good sequence coverage from complete b and y ion series

requires randomized proton location to cover all backbone amide bonds. There are several models of how CAD proceeds. In one (illustrated in Figure 1.10), the mobile proton lands on an amide nitrogen. The oxygen of the N-terminal carbonyl attacks the carbon of this amide carbonyl, displacing the pair of electrons forming the C-N bond onto the protonated nitrogen and severing the bond to produce an oxazolone b ion and a complementary y ion (83,84).

Unfortunately for PTM analysis, the P-O bond of phosphorylated residues and the O-C glycan bond are fairly labile, so the modification is often lost in CAD. This makes site-localizing the PTM from only CAD spectra very difficult. The loss of the modification from the precursor is usually one of the most abundant peaks in the spectrum. The loss of 98 Da from a phosphorylated fragment is indistinguishable from the loss of 18 Da (water, another common loss) from a non-phosphorylated version of the same fragment. For O-GlcNAc, CAD gives an ion at m/z 204, the oxonium ion of O-GlcNAc, and the charge reduced $(M+nH-203)^{n+}$, which is the predominant species observed. Another fragmentation method is necessary in order to site-localize PTMs (53,85).

Electron-transfer dissociation (ETD) was developed in the Hunt lab in 2004 (86). It was an improvement on the older technique of electron-capture dissociation (ECD), developed in the lab of Fred McLafferty in 1998 (87). ECD was performed in a Fourier transform ion cyclotron resonance MS (FT-ICR-MS), where peptide cations were reacted with low-energy electrons (a few eV) to produce radical cations that then underwent cleavage at the backbone N-$C_\alpha$ bond. This preserved labile PTMs; however, ECD was only readily useable with an FT-ICR instrument (84,85,87).

As shown in Figure 1.11, ETD uses a polycyclic aromatic radical anion reagent (such as fluoranthene) rather than a beam of electrons to produce a radical peptide cation. The radical

anion reagent is trapped in the ion trap with the precursor. The reagent transfers its radical electron to the precursor peptide, reducing the charge on the precursor by one (and making the fluoranthene neutral); this reaction is exothermic by around 4-5 eV. The electron is captured by a peptide amide carbonyl, leading to cleavage of the pi bond and localization of the radical electron to the carbonyl carbon. The resulting carbonyl oxygen anion abstracts a proton from a neighboring basic amino acid. The N-$C_\alpha$ bond then breaks, leaving an N-terminal c ion and a C-terminal radical z· ion (with the radical on the $C_\alpha$) When the bond N-terminal to a proline is broken, no dissociation occurs because the fragments remain covalently linked. The electron transfer and subsequent fragmentation occurs very fast, too fast for the energy inputted to rearrange via vibrational relaxation, so labile bonds such as phosphorylation or glycans are preserved. This allows the PTM to be site-localized (85,88).

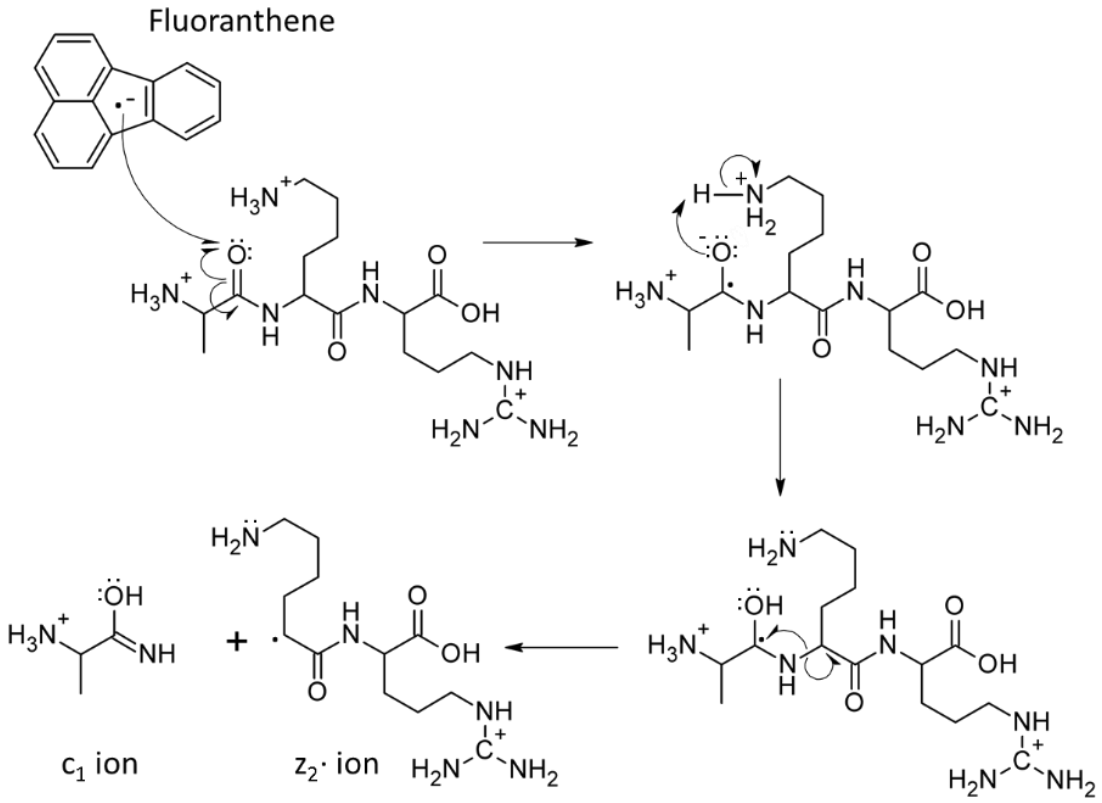


**Figure 1.11 Electron-transfer dissociation (ETD) fragmentation mechanism.**

ETD works better for peptides of higher charge density. Since ETD reduces the charge of the precursor ion by one, when a 2+ precursor fragments into two products, only one of them would bear a charge. Thus, only the c or z· product ion (but not both) would be detectable. Furthermore, sometimes the two fragments will fail to separate, in a case of ETnoD (electron transfer, no dissociation). In ETnoD the fragments are held together by noncovalent interactions, such as hydrogen-bonding. ETnoD will be more pronounced for a +2 precursor, since only one fragment bears a charge, so there is no other charge to repel the first. Indeed, ETD spectra of +2 precursors are dominated by the $(M+2H)^{+\cdot}$, with only a few of the highest mass c and z· ions seen (89).

CAD and ETD are complementary techniques. CAD functions better with lower charge density (thus a lower charge state precursor): more charges on a peptide means more restriction on where a mobile proton goes, so a few sites will be preferentially cleaved and complete sequence coverage will probably not occur. Using both fragmentation methods to analyze a sample allows for better sequence coverage on a protein than either alone, especially when there are labile PTMs. (54,85).

Another commonly used fragmentation method is higher-energy C-trap dissociation (or higher-energy collisional dissociation, HCD), developed by Olsen et al. in 2007 (90). They sought to improve upon CAD by conducting fragmentation in the C-trap of an LTQ Orbitrap, using an increased trapping RF voltage to trap more fragment ions. However, this high voltage made it harder to trap low m/z product ions, so they moved to fragmenting in a dedicated octupole collision cell located beyond the C-trap. The octupole had an RF frequency voltage and a variable DC offset. In this setup, low mass ions were seen (90). In an ion trap, the CAD RF voltage for trapping product ions imposes a LMCO according to the m/z of the precursor

ion (usually one third of the precursor m/z). HCD allows for ion activation (accelerating precursor ions into the collision cell with a DC potential difference) independent of product ion trapping, so the LMCO of IT CAD no longer exists. HCD can produce side-chain fragmentation, as HCD tends to input more energy into precursor ions vs. CAD, and all ions are activated (vs. the selected precursor in CAD). HCD spectra tend to have high y-ion coverage (54,84).

**1.5 Data Analysis**

Once a sample has been analyzed by the mass spectrometer, the data exist in a raw state as lists of scans, each with a list of ion m/z values observed and their abundances. This data may be visualized with a program such as Xcalibur Qual Browser (Thermo Fisher). The program can display a mass spectrum, with all the ions (and their abundances) occurring at a particular time; or a chromatogram, with the abundance of a particular ion over time. To turn these diagrams into proteomic information, however, interpretation is necessary. De novo sequencing divines the identity of a peptide directly from the mass spectrum. For large data sets, however, search engine programs are used to generate proposed peptides. Database search engines compare spectra to a list of peptides generated from in silico digestion of one or more proteins. More recently, hybrid programs have been developed which combine limited de novo interpretation to identify ions from a spectrum with more targeted database searching to identify peptides. The hybrid search engine Byonic (Protein Metrics, Inc) was used in this work. Manual validation of peptide hits is advisable to confirm the results, and was also used here (53,91,92).

Relative quantification of post-translational modifications is important for determining the effects of different *Arabidopsis* mutants on plant growth regulation and growth protein interactions with one another. Each sample in this work is from a specific mutant. Each harvest and immunopurification of RGA from a batch of *Arabidopsis* plants tends to produce 5-10 pmol (0.34 to 0.68 ug) of protein, which is enough for several mass spectrometric analyses. The levels of modified vs. unmodified forms of a given peptide are calculated by comparing the chromatographic peak areas for the corresponding MS1 ions. The most abundant modifications of the RGA peptides of interest may be of higher abundance than the unmodified form (for example, the GlcNAcylated form of GVIGTTVTTTTTTTTAAGESTR is frequently more abundant that the unmodified form). However, the least abundant modifications are often present at less than 1% of the abundance of the unmodified peptide, though they are still important. Phosphorylation, O-glycosylation, and lysine acetylation are reversible modifications that may only be found on RGA at specific spatiotemporal distributions (1,50). Dynamic exclusion of precursors from fragmentation assists with detecting the low-level modifications, but there are still many species eluting from the column simultaneously. The modified forms of a peptide usually elute within a few minutes of the unmodified form. Site-localizing PTMs is more difficult still, and will be discussed in Chapter 2.

Andrew Dawdy and Benjamin Barnhill (12,28) have conducted much research into the post-translational modifications and of RGA and their functions. They have studied RGA modification by *sec* and *spy* mutants in tobacco and *Arabidopsis*, but not with the $RGA^{GKG}$ form in *Arabidopsis*. This work examines the interplay of phosphorylation and other PTMs on RGA through the analysis of different *Arabidopsis* mutants. Additionally, this work identified

two novel acetylation sites on RGA, some of the first acetylation identified on a DELLA protein. By better understanding the regulation of RGA and plant growth networks in *Arabidopsis*, we can better understand how commercially important plants with RGA orthologs such as maize or rice grow, and how we may improve that growth.

## 1.6 References


1. Walsh G, Proteins: Biochemistry and Biotechnology, 2nd ed. Chichester, UK: John Wiley & Sons, Ltd; 2014.
2. Černý M, Novák J, Habánová H, Cerna H, Brzobohatý B, Role of the Proteome in Phytohormonal Signaling, Biochim Biophys Acta, **2016**;1864(8):1003–1015.
3. Mann M, Jensen ON, Proteomic Analysis of Post-Translational Modifications, Nat Biotechnol, **2003**;21(3):255–261.
4. Van Wijk KJ, Friso G, Walther D, Schulze WX, Meta-Analysis of *Arabidopsis thaliana* Phospho-Proteomics Data Reveals Compartmentalization of Phosphorylation Motifs, Plant Cell, **2014**;26(6):2367–2389.
5. Torres C-R, Hart GW, Topography and Polypeptide Distribution of Terminal N-Acetylglucosamine Residues on the Surfaces of Intact Lymphocytes, J Biol Chem, **1984**;259(5):3308–3317.
6. Wang Z, Hart GW, Glycomic Approaches to Study GlcNAcylation: Protein Identification, Site-Mapping, and Site-Specific O-GlcNAc Quantitation, Clin Proteomics, **2008**;4:5–13.
7. Wells L, Vosseller K, Hart GW, Glycosylation of Nucleocytoplasmic Proteins: Signal Transduction and O-GlcNAc, Science, **2001**;291(5512):2376–2378.
8. Zeidan Q, Hart GW, The Intersections between O-GlcNAcylation and Phosphorylation: Implications for Multiple Signaling Pathways, J Cell Sci, **2010**;123(1):13–22.
9. Hart GW, Slawson C, Ramirez-Correa G, Lagerlof O, Cross Talk between O-GlcNAcylation and Phosphorylation: Roles in Signaling, Transcription, and Chronic Disease, Annu Rev Biochem, **2011**;80:825–858.
10. Kentzer EJ, Buko AM, Menon G, Sarin VK, Carbohydrate Composition and Presence of a Fucose-Protein Linkage in Recombinant Human Pro-Urokinase, Biochem Biophys Res Commun, **1990**;171(1):401–406.
11. Luther KB, Haltiwanger RS, Role of Unusual O-Glycans in Intercellular Signaling, Int J Biochem Cell Biol, **2009**;41(5):1011–1024.
12. Zentella R, Sui N, Barnhill B, Hsieh W-P, Hu J, Shabanowitz J, Boyce M, Olszewski NE, Zhou

P, Hunt DF, Sun T, The *Arabidopsis* O-Fucosyltransferase SPINDLY Activates Nuclear Growth Repressor DELLA, Nat Chem Biol, **2017**;13(5):479–485.

13. Allfrey VG, Faulkner R, Mirsky AE, Acetylation and Methylation of Histones and Their Possible Role in the Regulation of RNA Synthesis, Proc Natl Acad Sci U S A, **1964**;51(5):786–794.

14. Li C, Choi H-P, Wang X, Wu F, Chen X, Lü X, Jing R, Ryu H, Wang X, Azadzoi K, Yang J-H, Post-Translational Modification of Human Histone by Wide Tolerance of Acetylation, Cells, **2017**;6(4):34.

15. Malthus TR, An Essay on the Principle of Population, London: Joseph Johnson; 1798.

16. Trewavas AJ, The Population/Biodiversity Paradox. Agricultural Efficiency to Save Wilderness, Plant Physiol, **2001**;125(1):174–179.

17. Hedden P, The Genes of the Green Revolution, Trends Genet, **2003**;19(1):5–9.

18. Davière J-M, Achard P, Gibberellin Signaling in Plants, Development, **2013**;140(6):1147–1151.

19. Yamaguchi S, Gibberellin Metabolism and its Regulation, Annu Rev Plant Biol, **2008**;59(1):225–251.

20. Griffiths J, Murase K, Rieu I, Zentella R, Zhang Z-L, Powers SJ, Gong F, Phillips AL, Hedden P, Sun T, Thomas SG, Genetic Characterization and Functional Analysis of the GID1 Gibberellin Receptors in *Arabidopsis*, Plant Cell, **2006**;18(12):3399–3414.

21. Ariizumi T, Murase K, Sun T, Steber CM, Proteolysis-Independent Downregulation of DELLA Repression in *Arabidopsis* by the Gibberellin Receptor GIBBERELLIN INSENSITIVE DWARF1, Plant Cell, **2008**;20(9):2447–2459.

22. Van De Velde K, Ruelens P, Geuten K, Rohde A, Van Der Straeten D, Exploiting DELLA Signaling in Cereals, Trends Plant Sci, **2017**;22(10):880–893.

23. Hedden P, Gibberellins Close the Lid, Nature, **2008**;456:455–456.

24. Sun T, Kamiya Y, The *Arabidopsis* GA1 Locus Encodes the Cyclase ent-Kaurene Synthetase A of Gibberellin Biosynthesis, Plant Cell, **1994**;6(10):1509–1518.

25. Sasaki A, Itoh H, Gomi K, Ueguchi-Tanaka M, Ishiyama K, Kobayashi M, Jeong D-H, An G, Kitano H, Ashikari M, Matsuoka M, Accumulation of Phosphorylated Repressor for Gibberellin Signaling in an F-Box Mutant, Science, **2003**;299(5614):1896–1898.

26. Silverstone AL, Mak PYA, Martínez EC, Sun T, The New RGA Locus Encodes a Negative Regulator of Gibberellin Response in *Arabidopsis thaliana*, Genetics, **1997**;146(3):1087–1099.

27. Silverstone AL, Ciampaglio CN, Sun T, The *Arabidopsis* RGA Gene Encodes a Transcriptional Regulator Repressing the Gibberellin Signal Transduction Pathway, Plant Cell, **1998**;10(2):155–169.

28. Zentella R, Hu J, Hsieh W-P, Matsumoto PA, Dawdy AW, Barnhill B, Oldenhof H, Hartweck

LM, Maitra S, Thomas SG, Cockrell S, Boyce M, Shabanowitz J, Hunt DF, Olszewski NE, Sun T, O-GlcNAcylation of Master Growth Repressor DELLA by SECRET AGENT Modulates Multiple Signaling Pathways in *Arabidopsis*, Genes Dev, **2016**;30(2):164–176.

29. Hochuli E, Döbeli H, Schacher A, New Metal Chelate Adsorbent Selective for Proteins and Peptides Containing Neighbouring Histidine Residues, J Chromatogr A, **1987**;411:177–184.

30. Einhauer A, Jungbauer A, The FLAG™ Peptide, a Versatile Fusion Tag for the Purification of Recombinant Proteins, J Biochem Biophys Methods, **2001**;49:455–465.

31. Meinke DW, Cherry JM, Dean C, Rounsley SD, Koornneef M, *Arabidopsis thaliana*: A Model Plant for Genome Analysis, Science, **1998**;282(5389):662–682.

32. Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, Muller R, Dreher K, Alexander DL, Garcia-Hernandez M, Karthikeyan AS, Lee CH, Nelson WD, Ploetz L, Singh S, Wensel A, Huala E, The *Arabidopsis* Information Resource (TAIR): Improved Gene Annotation and New Tools, Nucleic Acids Res, **2012**;40(Database):D1202–D1210.

33. Jacobsen SE, Olszewski NE, Mutations at the SPINDLY Locus of *Arabidopsis* Alter Gibberellin Signal Transduction, Plant Cell, **1993**;5(8):887–896.

34. Hartweck LM, Scott CL, Olszewski NE, Two O-linked N-Acetylglucosamine Transferase Genes of *Arabidopsis thaliana* L. Heynh. Have Overlapping Functions Necessary for Gamete and Seed Development, Genetics, **2002**;161(3):1279–1291.

35. Hartweck LM, Genger RK, Grey WM, Olszewski NE, SECRET AGENT and SPINDLY Have Overlapping Roles in the Development of *Arabidopsis thaliana* L. Heyn., J Exp Bot, **2006**;57(4):865–875.

36. Xu S-L, Chalkley RJ, Maynard JC, Wang W, Ni W, Jiang X, Shin K, Cheng L, Savage D, Hühmer AFR, Burlingame AL, Wang Z-Y, Proteomic Analysis Reveals O-GlcNAc Modification on Proteins with Key Regulatory Functions in *Arabidopsis*, Proc Natl Acad Sci U S A, **2017**;114(8):E1536-1543.

37. Wells L, Kreppel LK, Comer FI, Wadzinski BE, Hart GW, O-GlcNAc Transferase is in a Functional Complex with Protein Phosphatase 1 Catalytic Subunits, J Biol Chem, **2004**;279(37):38466–38470.

38. Martínez-Turiño S, Pérez JDJ, Hervás M, Navajas R, Ciordia S, Udeshi ND, Shabanowitz J, Hunt DF, García JA, Phosphorylation Coexists with O-GlcNAcylation in a Plant Virus Protein and Influences Viral Infection, Mol Plant Pathol, **2018**;19(6):1427–1443.

39. Dai C, Xue H-W, Rice early flowering1, a CKI, Phosphorylates DELLA Protein SLR1 to Negatively Regulate Gibberellin Signalling, EMBO J, **2010**;29(11):1916–1927.

40. Hartl M, Füßl M, Boersema PJ, Jost J-O, Kramer K, Bakirbas A, Sindlinger J, Plöchinger M, Leister D, Uhrig RG, Moorhead GBG, Cox J, Salvucci ME, Schwarzer D, Mann M, Finkemeier I, Lysine Acetylome Profiling Uncovers Novel Histone Deacetylase Substrate Proteins in *Arabidopsis*, Mol Syst Biol, **2017**;13(10):949.

41. Finkemeier I, Laxa M, Miguet L, Howden AJM, Sweetlove LJ, Proteins of Diverse Function

and Subcellular Location Are Lysine Acetylated in *Arabidopsis*, Plant Physiol, **2011**;155(4):1779–1790.

42. Wu X, Oh M-H, Schwarz EM, Larue CT, Sivaguru M, Imai BS, Yau PM, Ort DR, Huber SC, Lysine Acetylation is a Widespread Protein Modification for Diverse Proteins in *Arabidopsis*, Plant Physiol, **2011**;155(4):1769–1778.

43. Rao RSP, Thelen JJ, Miernyk JA, Is Lys-$N^{\varepsilon}$-Acetylation the Next Big Thing in Post-Translational Modifications?, Trends Plant Sci, **2014**;19(9):550–553.

44. Uhrig RG, Schläpfer P, Roschitzki B, Hirsch-Hoffmann M, Gruissem W, Diurnal Changes in Concerted Plant Protein Phosphorylation and Acetylation in *Arabidopsis* Organs and Seedlings, Plant J, **2019**;99(3):176–194.

45. Cheung P, Tanner KG, Cheung WL, Sassone-Corsi P, Denu JM, Allis CD, Synergistic Coupling of Histone H3 Phosphorylation and Acetylation in Response to Epidermal Growth Factor Stimulation, Mol Cell, **2000**;5(6):905–915.

46. Ahn S-H, Diaz RL, Grunstein M, Allis CD, Histone H2B Deacetylation at Lysine 11 is Required for Yeast Apoptosis Induced by Phosphorylation of H2B at Serine 10, Mol Cell, **2006**;24(2):211–220.

47. The *Arabidopsis* Genome Initiative, Analysis of the Genome Sequence of *Arabidopsis thaliana*, Nature, **2000**;408(6814):796–815.

48. International Human Genome Sequencing Consortium, Finishing the Euchromatic Sequence of the Human Genome, Nature, **2004**;431(7011):931–945.

49. Anderson NL, Anderson NG, Proteome and Proteomics: New Technologies, New Concepts, and New Words, Electrophoresis, **1998**;19(11):1853–1861.

50. Millar AH, Heazlewood JL, Giglione C, Holdsworth MJ, Bachmair A, Schulze WX, The Scope, Functions, and Dynamics of Posttranslational Protein Modifications, Annu Rev Plant Biol, **2019**;70(1):119–151.

51. Abelin JG, Trantham PD, Penny S, Patterson AM, Ward ST, Hildebrand WH, Cobbold M, Bai DL, Shabanowitz J, Hunt DF, Complementary IMAC Enrichment Methods for HLA-Associated Phosphopeptide Identification by Mass Spectrometry, Nat Protoc, **2015**;10(9):1308–1318.

52. Zee BM, Garcia BA, Discovery of Lysine Post-Translational Modifications through Mass Spectrometric Detection, Essays Biochem, **2012**;52:147–163.

53. Steen H, Mann M, The ABC's (and XYZ's) of Peptide Sequencing, Nat Rev Mol Cell Biol, **2004**;5(9):699–711.

54. Brodbelt JS, Ion Activation Methods for Peptides and Proteins, Anal Chem, **2016**;88(1):30–51.

55. Edman P, Method for Determination of the Amino Acid Sequence in Peptides, Acta Chem Scand, **1950**;4:283–293.

56. Mallick P, Kuster B, Proteomics: a Pragmatic Perspective, Nat Biotechnol,

**2010**;28(7):695–709.

57. Eng JK, McCormack AL, Yates III JR, An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database, J Am Soc Mass Spectrom, **1994**;5(11):976–989.

58. Domon B, Aebersold R, Mass Spectrometry and Protein Analysis, Science, **2006**;312(5771):212–217.

59. Jarchum I, Finklestein J, Newton A, De Ranieri E, Georgescu I, Fleet L, Bergin E, Faust T, Strack R, Mueck L, Barnes K, Goodman C, Johnson R, Legate K, Dellisanti CD, Fehervari Z, Dekker B, Hansell C, Kirk R, et al., Milestones in Mass Spectrometry, Nat Methods, **2015**;12:S1-21.

60. March RE, An Introduction to Quadrupole Ion Trap Mass Spectrometry, J Mass Spectrom, **1997**;32(4):351–369.

61. Barber M, Bordoli RS, Sedgwick RD, Tyler AN, Fast Atom Bombardment of Solids as an Ion Source in Mass Spectrometry, Nature, **1981**;293:270–275.

62. Karas M, Bachmann D, Bahr U, Hillenkamp F, Matrix-Assisted Ultraviolet Laser Desorption of Non-Volatile Compounds, Int J Mass Spectrom Ion Process, **1987**;78:53–68.

63. Tanaka K, Waki H, Ido Y, Akita S, Yoshida Y, Yoshida T, Matsuo T, Protein and Polymer Analysis up to m/z 100 000 by Laser Ionization Time-of-Flight Mass Spectrometry, Rapid Commun Mass Spectrom, **1988**;2(8):151–153.

64. Yamashita M, Fenn JB, Electrospray Ion Source. Another Variation on the Free-Jet Theme, J Phys Chem, **1984**;88(20):4451–4459.

65. Clark AE, Kaleta EJ, Arora A, Wolk DM, Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry: a Fundamental Shift in the Routine Practice of Clinical Microbiology, Clin Microbiol Rev, **2013**;26(3):547–603.

66. Cech NB, Enke CG, Relating Electrospray Ionization Response to Nonpolar Character of Small Peptides, Anal Chem, **2000**;72(13):2717–2723.

67. Wilm M, Principles of Electrospray Ionization, Mol Cell Proteomics, **2011**;10(7):M111.009407.

68. Cech NB, Enke CG, Practical Implications of Some Recent Studies in Electrospray Ionization Fundamentals, Mass Spectrom Rev, **2001**;20(6):362–387.

69. Martin SE, Shabanowitz J, Hunt DF, Marto JA, Subfemtomole MS and MS/MS Peptide Sequence Analysis Using Nano-HPLC Micro-ESI Fourier Transform Ion Cyclotron Resonance Mass Spectrometry, Anal Chem, **2000**;72(18):4266–4274.

70. Schwartz JC, Senko MW, Syka JEP, A Two-Dimensional Quadrupole Ion Trap Mass Spectrometer, J Am Soc Mass Spectrom, **2002**;13(6):659–669.

71. Snyder DT, Peng W-P, Cooks RG, Resonance Methods in Quadrupole Ion Traps, Chem Phys Lett, **2017**;668:69–89.

72. Savaryn JP, Toby TK, Kelleher NL, A Researcher's Guide to Mass Spectrometry-Based Proteomics, Proteomics, **2016**;16(18):2435–2443.

73. March RE, Todd JFJ, Radio Frequency Quadrupole Technology: Evolution and Contributions to Mass Spectrometry, Int J Mass Spectrom, **2015**;377:316–328.

74. Makarov A, Electrostatic Axially Harmonic Orbital Trapping: a High-Performance Technique of Mass Analysis, Anal Chem, **2000**;72(6):1156–1162.

75. Hu Q, Noll RJ, Li H, Makarov A, Hardman M, Cooks RG, The Orbitrap: a New Mass Spectrometer, J Mass Spectrom, **2005**;40(4):430–443.

76. Eliuk SM, Makarov A, Evolution of Orbitrap Mass Spectrometry Instrumentation, Annu Rev Anal Chem, **2015**;8(1):61–80.

77. Perry RH, Cooks RG, Noll RJ, Orbitrap Mass Spectrometry: Instrumentation, Ion Motion and Applications, Mass Spectrom Rev, **2008**;27(6):661–699.

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

79. Biemann K, Gapp F, Seibl J, Application of Mass Spectrometry to Structure Problems. I. Amino Acid Sequence in Peptides, J Am Chem Soc, **1959**;81(9):2274–2275.

80. Hunt DF, Buko AM, Ballard JM, Shabanowitz J, Giordani AB, Sequence Analysis of Polypeptides by Collision Activated Dissociation on a Triple Quadrupole Mass Spectrometer, Biomed Mass Spectrom, **1981**;8(9):397–408.

81. Jennings KR, Collision-Induced Decompositions of Aromatic Molecular Ions, Int J Mass Spectrom Ion Phys, **1968**;1(3):227–235.

82. Hunt DF, Yates III JR, Shabanowitz J, Winston S, Hauer CR, Protein Sequencing by Tandem Mass Spectrometry, Proc Natl Acad Sci U S A, **1986**;83(17):6233–6237.

83. Wysocki VH, Tsaprailis G, Smith LL, Breci LA, Mobile and Localized Protons: a Framework for Understanding Peptide Dissociation, J Mass Spectrom, **2000**;35(12):1399–1406.

84. Rathore D, Aboufazeli F, Huang Y, Kolli V, Fernando GS, Dodds ED, Ion Dissociation Methods in Proteomics, Encycl Anal Chem, **2015**;1–26.

85. Mikesh LM, Ueberheide B, Chi A, Coon JJ, Syka JEP, Shabanowitz J, Hunt DF, The Utility of ETD Mass Spectrometry in Proteomic Analysis, Biochim Biophys Acta, **2006**;1764(12):1811–1822.

86. Syka JEP, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF, Peptide and Protein Sequence Analysis by Electron Transfer Dissociation Mass Spectrometry, Proc Natl Acad Sci U S A, **2004**;101(26):9528–9533.

87. Zubarev RA, Kelleher NL, McLafferty FW, Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process, J Am Chem Soc, **1998**;120(13):3265–

3266.

88. Jennings KR, Analytical Applications of Ion Activation Techniques, Int J Mass Spectrom, **2015**;375:610–616.

89. Udeshi ND, Compton PD, Shabanowitz J, Hunt DF, Rose KL, Methods for Analyzing Peptides and Proteins on a Chromatographic Timescale by Electron-Transfer Dissociation Mass Spectrometry, Nat Protoc, **2008**;3(11):1709–1717.

90. Olsen J V, Macek B, Lange O, Makarov A, Horning S, Mann M, Higher-Energy C-trap Dissociation for Peptide Modification Analysis, Nat Methods, **2007**;4(9):709–712.

91. Bern M, Kil YJ, Becker C, Byonic: Advanced Peptide and Protein Identification Software, Curr Protoc Bioinforma, **2012**;40(1):13.20.1-14.

92. Bern M, Cai Y, Goldberg D, Lookup Peaks: A Hybrid of de Novo Sequencing and Database Search for Protein Identification by Tandem Mass Spectrometry, Anal Chem, **2007**;79(4):1393–1400.

# Chapter 2: Post-Translational Modifications of RGA

## 2.1 RGA and Its Protein Interactors

### *2.1.1 RGA Structure and Function*

The subject of this research is the DELLA protein RGA. RGA is in the DELLA subfamily of the GRAS family of proteins. As shown in Figure 2.1, RGA is composed of two domains, the N-terminal DELLA domain and the larger C-terminal effector GRAS domain. These domains are separated by a poly-S/T region. The GFRAS domain is conserved across GRAS family proteins, but different subfamilies (like DELLA) have different N-terminal domains. The GRAS domain has two leucine heptad repeat subdomains (LHRI and II) and the subdomains VHIID, PFYRE, and SAW (1–3). These GRAS subdomains form an α-helical cap composed of the LHRI and part of the PFYRE subdomains, and an α/β core composed of the rest of the subdomains (4).

The N-terminal region of DELLA proteins is intrinsically disordered in the free protein, except for the DELLA, LExLE, TVHYNP, and L(K/R)XI motifs. This region includes the stretch of residues before the DELLA motif (with the LSNHGTSSSSSSISK(DK) peptide, abbreviated as LSN) and the poly-S/T stretch between the DELLA and GRAS domains (with the VIPGNAIYQFPAIDSSSSSNNQNK(R), (LK)SCSSPDSMVTSTSTGTQIGK, and GVIGTTVTTTTTTTTAAGESTR peptides, abbreviated VIP, SCSS, and GVIG, respectively) (3). These four peptides are indicated in the Figure 2.1 sequence, in green. The DELLA, LExLE, and TVHYNP motifs are involved in binding to the gibberellin receptor GID1. The TVHYNP loop changes conformation when RGA binds to GID1. In the crystal structure of another *Arabidopsis* DELLA protein by Murase et al. (5), the region including the DELLA, LExLE, and

TVHYNP motifs becomes a bundle of α-helices when bound to GID1 (5,6). These motifs are molecular recognition features (MoRFs), which are often found in intrinsically disordered regions and instigate protein-target recognition. Phosphorylation often occurs in these regions (7).

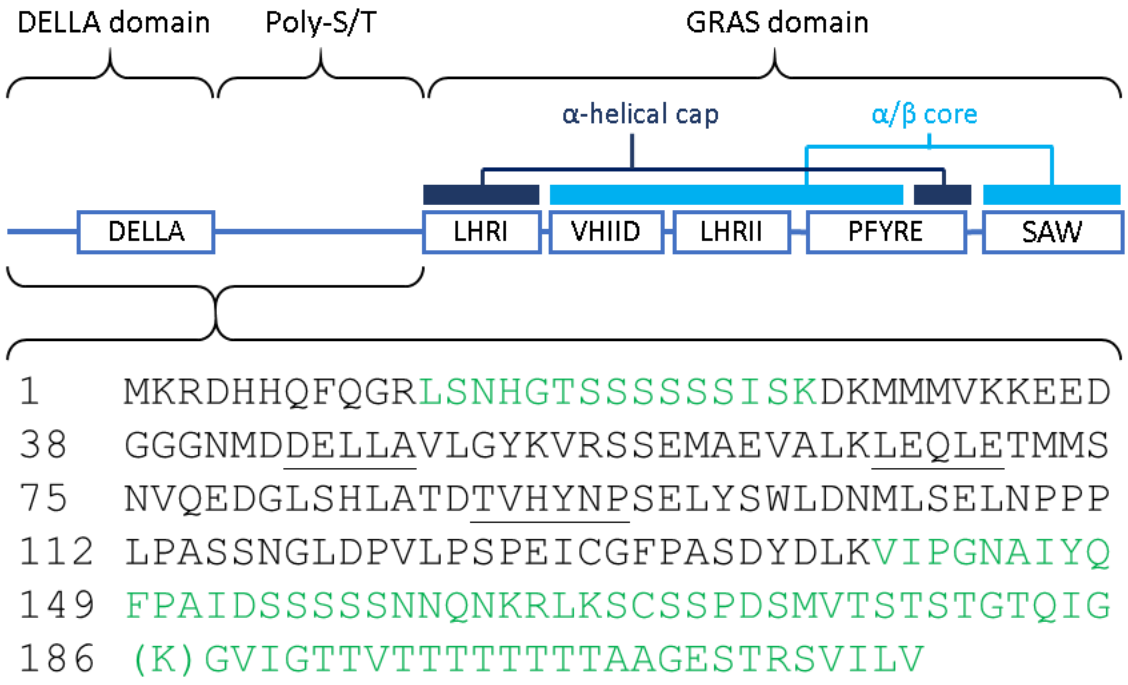


**Figure 2.1 Structure of RGA.** RGA consists of an N-terminal disordered region (including the DELLA domain and poly-S/T region) and a conserved C-terminal GRAS domain. The GRAS domain contains five subdomains that form an α -helical cap and an α/β core in the structure. The first 212 residues of RGA are shown, with the peptides studied in this work colored green. The DELLA, LExLE, and TVHYNP motifs are underlined. The parenthesized K is the inserted lysine of $RGA^{GKG}$. This figure was adapted from figures in Zentella 2017 (3) and Hirano 2017 (4).

RGA works by sequestering transcription factors (TFs) or regulators (TRs), preventing TF binding to DNA (e.g. PIFs) or TR binding to TFs (e.g. JAZs). RGA can also be a transactivation factor, turning on transcription of some genes through direct interaction with transcription factors (e.g. IDDs). RGA sequesters targets to repress growth through its LHRI and (sometimes) SAW subdomains, and transactivates targets through its LHRI subdomain (Figure 2.2). RGA also homodimerizes at its LHRI subdomain. RGA regulated proteins include

PHYTOCHROME INTERACTING FACTOR (PIF3/4), BRASSINAZOLE RESISTANT (BZR1), JASMONATE ZIM-DOMAIN (JAZ1/3), SCARECROW-LIKE (SCL3), ALCATRAZ (ALC), ABSCISIC ACID INSENSITIVE (ABI3), AUXIN RESPONSE FACTOR (ARF6/7/8), and INDETERMINATE DOMAIN (IDD1/4/5/9, JKD, MGP). PIFs regulate the response to light and the synthesis of chlorophyll. BZRs interface gibberellin- and brassinazole-signaling pathways. JAZs interface gibberellin- and jasmonate-signaling pathways, root growth, and defense against pathogens. SCL3 is a GRAS (but not DELLA) protein that influences root elongation and seed germination. All four of these contribute to hypocotyl elongation. ALC and ABIs influence fruit patterning and seed germination, respectively (1,6,8–10). DELLA proteins are thus integrators for multiple hormone signaling pathways including ethylene, auxin, abscisic acid, jasmonate, and brassinosteroid signaling, as well as environmental signals like light, abiotic stresses, and temperature (9,11,12).

DELLA mutations played important roles in the Green Revolution. For example, one of the important varieties of wheat used to produce semi-dwarf, high-yield crops was the Norin 10 variety; the Norin 10 dwarfing gene was found in over 70% of global wheat cultivars in 2002. Wheat is hexaploid; Norin 10 has mutations in two of its three homoeologous TaRHT1 genes, *TaRht-B1b* and *TaRht-D1b*. These mutants have similar effects on plant height individually and additively. DELLA proteins are highly conserved among plants. Besides TaRHT1 and RGA, DELLA proteins include grapevine VvGAI, tomato PROCERA, *Arabidopsis* AtGAI and AtRGL1-3, maize ZmD8 and ZmD9, rice OsSLR1, and barley HvSLN1. *TaRht-B1b* and *TaRht-D1b* have a truncation removing the DELLA motif, so they are constitutively active for repression of gibberellin signaling. Dwarfing mutants of TaRHT1 orthologs, including RGA, often have similar deletions in the DELLA and/or TVHYNP regions (1,8,9,13).

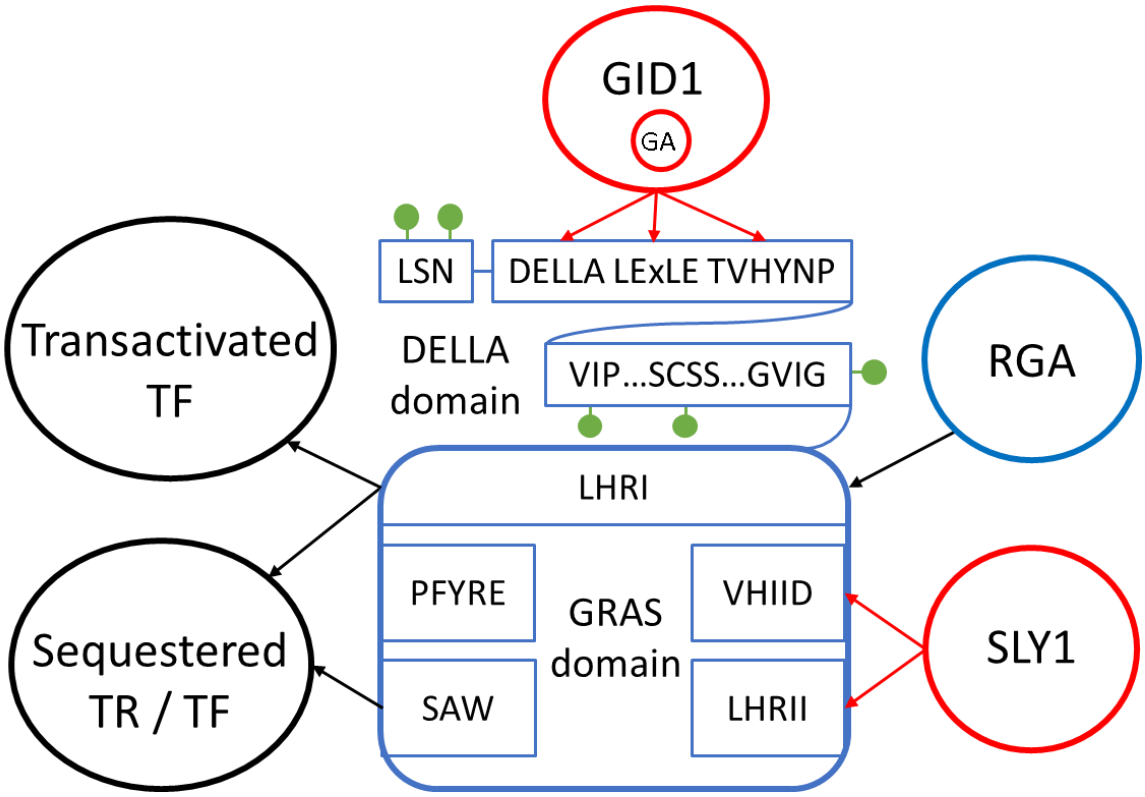


**Figure 2.2 RGA protein-protein interactions**. The unstructured DELLA domain contains the LSN and VIP…SCSS…GVIG poly-S/T regions with the PTMs (green dots) observed in this work. When bound to gibberellin (GA), GID1 binds to the DELLA, LExLE, and TVHYNP motifs in the DELLA domain; this also recruits SLY1 to bind to the VHIID and LHRII subdomains (SLY1 will then recruit the rest of the SCF E3 ubiquitin complex to polyubiquitylate RGA, not shown here). RGA binds and sequesters transcription factors (TFs) and transcription regulators (TRs) via its LHRI (and sometimes SAW) subdomains; RGA also binds to other TFs via its LHRI subdomain to enhance their transcription of target genes. RGA can homodimerize via the LHRI domain.

### *2.1.2 GA1, GID1, and SLY1: RGA Degradation*

SLEEPY1 (SLY1) is an F-box protein and part of the SCF E3 ubiquitin-ligase complex that polyubiquitylates RGA in the presence of gibberellin, causing proteasomal degradation of RGA. SLY1 acts as an adaptor for the SCF complex to target RGA; the SCF complex does not specifically target RGA without SLY1. (14). Dill et al. found that SLY1 interacts directly with the C-terminal GRAS domain of RGA but not its N-terminal DELLA domain; the DELLA domain has been shown necessary for gibberellin-induced degradation of RGA (15). The *sly1-10* mutation

consists of the deletion of the last 8 amino acids of the 151 amino-acid protein and the addition of 46 amino acids to the C-terminus. This preserves the N-terminal F-box motif which interacts with the rest of the SCF E3 complex, but drastically alters the C-terminal target-interacting domain, so it does not bind to RGA. In the *sly1-10* mutant, RGA is not ubiquitylated and degraded, so more RGA is present in a given cell (14). In addition to its biological relevance, this allows more RGA to be purified from the same amount of plant material, which enables better detection of low-level modifications and their sites.

GIBBERELLIN INSENSITIVE DWARF1 (GID1) mediates gibberellin-induced degradation of RGA by recruiting SLY1. Gibberellin and a functional DELLA domain of RGA are necessary for GID1 binding. As shown in the crystal structure by Murase et al. (5), GID1 has a gibberellin-binding pocket and a flexible N-terminal tail; when gibberellin binds, its C3 hydroxyl hydrogen bonds to Y31 on GID1, changing the conformation of GID1. The N-terminal tail of GID1 swings down to cover the gibberellin-binding pocket, and the lid can now bind the DELLA, LExLE, and TVHYNP portions of the DELLA domain of RGA (which are necessary for GID1-gibberelllin binding to RGA). Once GID1 binds to RGA, RGA changes conformation, and SLY1 can recognize and bind to its GRAS domain through the VHIID and LHRII subdomains. (1,5,8,16).

GA1 (ent-kaurene synthetase A) is the first enzyme in the gibberellin (GA) biosynthesis pathway. The *ga1-3* mutant has a large deletion in the gene that renders it a null mutant; these mutants produce almost no GA (17). In the *ga1-3* mutant, GA is scarcely present to bind to the protein GID1, so GID1 does not bind to RGA and repress it or induce its ubiquitylation and degradation. Both *sly1-10* and *ga1-3* prevent proteasomal degradation of RGA, but *ga1-3* interrupts all GA-based signaling through the absence of GA, whereas *sly1-10* does not. *Ga1-3* is a more severe phenotype than *sly1* mutants; *ga1-3* seeds fail to germinate at all, whereas

*sly1* mutants are sometimes able to germinate after a long dormancy period. Male *ga1-3* mutants are infertile, whereas *sly1* mutants are only partially infertile. However, *sly1* mutant plants have elevated levels of RGA compared to *ga1-3* plants – is the plant compensating for the elevated levels of RGA by rendering it less active (14)? In the presence of gibberellin and a functional RGA DELLA domain, overexpressing GID1 rescues *sly1* mutant phenotypes (dwarf and infertility) without decreasing the elevated levels of RGA. This indicates that GID1 can attenuate RGA activity to some extent simply by binding to its DELLA domain; the repression of gibberellin signaling by RGA depends on more than just the level of RGA (16).

### *2.1.3 SPY and SEC: RGA Post-Translational Modification*

RGA activity is also modulated via post-translational modification (PTM). The O-GlcNAcyltransferase SEC upregulates gibberellin signaling – the *sec-3* mutant has shorter hypocotyl growth, shorter final height, and reduced fertility. *sec-3* is a null mutant, with no SEC mRNA or protein present. SEC binds to RGA residues 107-285, which includes VIP…SCSS…GVIG and LHRI (18). The protein O-fucosyltransferase (POFUT) SPY also affects gibberellin signaling, as *spy* mutation partially rescues the dwarfism of *ga1-3* mutants. The mutants *spy-8*, *spy-12*, and *spy-15* were reported in Silverstone 2007 (19). *Spy-19* was reported in Zentella 2017 (3). *Spy-8* has a 23 amino acid deletion (M354-Q376, in TPR3-TPR2), *spy-12* has a G570D substitution, *spy-15* has an E567K substitution, and *spy-19* has a K665M substitution. The last three have a mutation in the catalytic domain and more severe phenotypes than *spy-8* (earlier flowering and worse fertility). *Spy-8* has 7% POFUT activity relative to wild type, while *spy-15* and *spy-19* have no detectable POFUT activity (3,19).

Neither SEC nor SPY affect RGA degradation, nor its nuclear localization. Instead, O-GlcNAcylation of RGA by SEC interferes with RGA interaction with BZR1 and PIFs, but O-fucosylation of RGA by SPY enhances RGA interactions with them. Mutant *sec* leads to lower expression of proteins repressed by DELLA (reducing BZR- and PIF-dependent responses), but mutant *spy* does the opposite. This leads to the idea, proposed by Zentella et al., of GlcNAcylation and fucosylation of RGA altering its conformation in different ways to either hinder or help the binding of target TFs and TRs (3,18).

These recent studies of SEC (18) and SPY (3) effects on RGA were a collaboration between the biology of the Sun lab of Duke University and the mass spectrometry of the Hunt lab of UVA. Andrew Dawdy of the Hunt lab studied post-translational modifications of RGA expressed in a tobacco species (*Nicotiana benthamiana*). He looked at plants which co-overexpressed *Arabidopsis* $RGA^{GKG}$ and SEC compared to ones without SEC, and plants which co-overexpressed the *Arabidopsis* casein kinase CK1. When difficulties obtaining sufficient mass spectral coverage of the SCSSPDSMVTSTSTGTQIGGVIGTTVTTTTTTTTAAGESTR poly-S/T region arose (due to poor chromatography and the lack of charge for quality ETD), a lysine insertion between G185 and G186 ($RGA^{GKG}$) was introduced. This bisected the region to give the peptides SCSSPDSMVTSTSTGTQIGK and GVIGTTVTTTTTTTTAAGESTR after trypsin digestion. He characterized a number of sites of O-GlcNAcylation, phosphorylation, and O-hexosylation. He also showed that RGA phosphorylation increased in the presence of *Arabidopsis* CK1 (18,20).

Benjamin Barnhill continued the work on RGA in the Hunt lab. He analyzed samples of $RGA^{GKG}$ from tobacco that contained wild-type or mutated *Arabidopsis* SPY. With Amanda Wriston (also of the Hunt lab) he discovered O-fucosylation on the LSN peptide, the first O-

fucosylation site in *Arabidopsis*. Additionally, he proved with the Sun lab that SPY is a POFUT, not an OGT; O-fucosylation levels were drastically reduced in *spy* mutants. He also analyzed samples of RGA from its native *Arabidopsis* in the presence of wild-type, *sec-3* or *spy-8* mutants (in the *ga1-3* background). However, the RGA in these samples did not have the lysine insertion, so he was unable to find post-translational modifications on SCSSPDSMVTSTSTGTQIGGVIGTTVTTTTTTTTAAGESTR (3,18).

Dawdy and Barnhill largely worked with RGA purified from tobacco, because tobacco has large leaves that produce a large amount of the engineered protein (tandem purification gave 10 to 50+ pmol of RGA per sample) (20). Proteins may be easily expressed in *N. benthamiana* via agroinfiltration due to its weak immune system; however, the *N. benthamiana* genome is not well characterized (21). More importantly, tobacco has its own OGT, which O-GlcNAcylated some sites of RGA even in the absence of *Arabidopsis* SEC (20).

In this present work, the post-translational modifications of $RGA^{GKG}$ in *Arabidopsis* in *sly1-10, ga1-3*, and *sly1-10* double mutants were examined. The double mutants were *sly1-10* + (*spy-8, -12, -15, -19; sec-3*). Dawdy and Barnhill analyzed RGA PTMs in *spy* or *sec* mutant samples, but this is the first work to examine RGA (with the lysine insertion) in these mutants in combination with *sly1-10* using the considerable analytical capabilities of mass spectrometry.

**2.2 Materials and Instrumentation**

Agilent Technologies (Palo Alto, CA)

1100 series high performance liquid chromatograph

1100 series vacuum degasser

Eppendorf (Hauppauge, NY)

5414R benchtop centrifuge

Labconco Corporation (Kansas City, MO)

Centrivap centrifugal vacuum concentrator

Molex (Lisle, IL)

Polymicro Technologies™ polyimide-coated fused silica capillary, 360 µm outer diameter, (50 µm or 75 µm or 150 µm) inner diameter

PQ Corporation (Valley Forge, PA)

Kasil ® 1624 (saturated potassium silicate solution)

Promega Corporation (Madison, WI)

Sequencing grade modified trypsin (#V5111)

Protein Metrics, Inc (Cupertino, CA)

Byonic search engine (v2.16.16)

Sigma Aldrich (Saint Louis, MO)

1,4-dithiothreitol, ≥97% purity

Acetic acid (glacial), ≥99.99% purity

Ammonium bicarbonate (AMBIC)

Angiotensin I acetate salt hydrate (human sequence, ≥99% purity) (Angio)

Anti-FLAG-M2 agarose beads

Formamide

Iodoacetamide (Bioultra), ≥99% purity

Tris(hydroxymethyl)aminomethane-HCl (Tris-HCl)

Vasoactive intestinal peptide fragment 1-12 (human sequence, ≥97% purity) (Vaso)

Sutter Instrument Co. (Navato, CA)

P-2000 microcapillary laser puller with fused silica adapter

Thermo Fisher Scientific (San Jose, CA/Bremen, Germany)

Orbitrap Fusion™ Tribrid™ Mass Spectrometer

Thermo Xcalibur Qual Browser (v.4.0.27.10)

VWR (Radnor, PA)

Mini vortexer

Water, LC-MS grade

Acetonitrile, LC-MS grade

YMC Company, LTD (Kyoto, Japan)

ODS-AQ, C18 5 µm spherical silica particles, 120 Å pore size

ODS-AQ, C18 15 µm spherical silica particles, 120 Å pore size

ODS-AQ, C18 5-20 µm irregular silica particles, 120 Å pore size

Zeus Industrial Products (Orangeburg, SC)

Teflon tubing, 0.012” i.d. x 0.060” o.d.

**2.3 Methods**

*These studies were performed in collaboration with Dr. Tai-Ping Sun and her lab at Duke University, who performed the Arabidopsis mutant generation, plant growth, and immunopurification. Trypsin digest of samples was performed by Dr. Rodolfo Zentella (Duke) or Ellen Speers. Mass spectrometry and data analysis were performed by Ellen Speers, and some by Emily Zahn, Dr. Mark Ross, and Dr. Jeffrey Shabanowitz of the Hunt lab.*

### *2.3.1 Plant Growth and Immunopurification of RGA$^{GKG}$*

The plant growth of transformed *Arabidopsis thaliana* was performed according to the procedure described in (18), which was modified from (22). *Arabidopsis* plants (with the *rga-24* null mutant background) were transformed with a modified form of RGA (tagged at the N-terminus with 6His-3xFLAG and with an inserted lysine between G185 and G186, designated RGA$^{GKG}$). The gene for RGA$^{GKG}$ was inserted in a plasmid, which was used to generate a transgenic, homozygous *Arabidopsis* line. This was accomplished using *Agrobacterium tumefasciens* strain GV3101 pMP90. A procedure as in (18) was used to generate plants with various mutants and combinations of mutants: *sly1-10*; *ga1-3*; and *sly1-10* + (*spy-8*, *spy-12*, *spy-15*, *spy-19,* or *sec-3*) double mutants. For affinity purification of RGA$^{GKG}$ protein (for mass spectrometry analysis), seeds were germinated in 250 mL DeLong flasks (~3000 seeds per flask) with 30 mL of liquid media (half-strength Murashige and Skoog medium, 1% sucrose) on a shaker at 100 rpm under constant light at 22$^{o}$C (18).

The immunopurification of RGA$^{GKG}$ from *Arabidopsis* was performed according to the procedure described in (18): first via the poly-histidine tag (23), then via the FLAG tag (24). A 10 g sample of 12-day-old liquid cultures of *Arabidopsis* was homogenized in extraction buffer (150 mM NaCl, 50 mM tris(hydroxymethyl)aminomethane-HCl (Tris-HCl), 0.5% 3-[(3-cholamindopropyl) dimethylammonio]-1-propane sulfonate (CHAPS), 2.5 mM 2-mercaptoethanol, 2 µg/mL aprotinin, 1 µg/mL pepstatin, 1 µg/mL leupeptin, 10 µM O-(2-Acetamido-2-deoxy-D-glucopyranosylidenamino)-N-phenylcarbamate (PuGNAc), and 20 µM carbobenzoxy-leu-leu-leucinal (MG-132)). The filtrate was mixed with 0.5 mL of pre-charged His-bind ($Ni^{2+}$-NTA) resin that had been equilibrated in extraction buffer. The mixture was incubated for 90 min with rotation. The mixture was loaded onto a disposable plastic column,

with the flow-through being discarded. The resin was washed with 5 mL of extraction buffer, and then the $RGA^{GKG}$ was eluted with extraction buffer with added 500 mM imidazole (18).

The $RGA^{GKG}$ was diluted with two volumes of extraction buffer, then mixed with 10 µL of anti-FLAG-M2-agarose beads. The sample/bead mixture was incubated at 4°C overnight with rotation. The beads were washed with 1 mL each of modified Low Salt, High Salt, and LiCl wash buffers as described in (25) (all detergents substituted with 0.5% CHAPS). The beads were washed five times with 1 mL of 100 mM ammonium bicarbonate (AMBIC), pH 7.9. The sample-bearing beads were divided between two tubes (90% for mass spectrometry analysis, 10% for SDS-PAGE analysis). For SDS-PAGE, the $RGA^{GKG}$ was eluted off the beads with 50 mM Tris-HCl, pH 8 and 1% SDS and loaded onto an 8% SDS-PAGE gel. Serial dilutions of protein standards (Bovine Serum Albumin, β-galactosidase from *E. coli*, and phosphorylase B from rabbit) were added to other lanes for quantification, and the gel was stained with Oriole fluorescent gel stain. An image was taken with a Gel-Doc, and the amount of RGA protein was determined with ImageJ software (18).

### *2.3.2 On-Beads Trypsin Digest*

Immunopurified RGA was digested on the anti-FLAG-M2-agarose beads according to the procedure in (18), with some modifications. 5 to 10 µL of protein-loaded anti-FLAG-M2-agarose beads were transferred to a low-protein binding tube. The beads were centrifuged at 1000 rpm for 30 s and the supernatant removed and replaced with 100 µL of 100 mM AMBIC pH 8.0. A fresh aliquot of 5 mM 1,4-dithiothreitol (DTT) was prepared and added to the beads in solution to a final concentration of 50 µM DTT. The sample was incubated on a shaker for 1 hour. A stock of 15 mM iodoacetamide (IAM) was freshly prepared, then added

to the beads for a final concentration of 150 µM IAM. The sample was incubated on a shaker for 1 hour in the dark. Trypsin was diluted to 100 ng/µL in 100 µM AMBIC, and 100 ng trypsin was added to the beads. The sample was incubated on a shaker in the dark overnight. The beads were then pelleted via centrifugation, and 90 µL of supernatant was removed into a low-protein binding tube. 100 µL of 100 mM AMBIC (pH 8.0) was added to the beads in the original tube, which was lightly vortexed and then centrifuged. 80-90 µL of the supernatant was combined with the first supernatant. The sample was dried in a Centrivap and then stored at -35$^{o}$C.

### *2.3.3 Column Fabrication*

*Precolumn (PC)* - A length of 360 µm o.d. x 75 µm i.d. fused silica capillary was cut. A frit was made at one end by dipping the tip of the column in a solution of 3:1 (v:v) potassium silicate (Kasil):formamide. The column was then heated in an oven either overnight or for 3-4 hours at 60$^{o}$C. The Kasil frit was trimmed to a length of 2 mm, and the column was packed on a pressure vessel with ODS-AQ C18 reverse-phase spherical silica particles, 15 µm (or 5-20 µm on older columns) diameter to a length of 8-9 cm. The PC was conditioned with Angiotensin I (DRVYIHPFHL, Angio) and Vasoactive intestinal peptide fragment 1-12 (HSDAVFTDNYTR, Vaso) repeatedly.

*Analytical column (AC)* - A length of 360 µm o.d. x 50 µm i.d. fused silica capillary was cut. A frit was made by dipping the tip of the column in a solution of 3:1 (v:v) Kasil:formamide, trimming the Kasil to 2 mm, then pushing the Kasil ~7 cm into the column with HPLC-grade water via capillary action. The column was then heated in an oven overnight or for 3-4 hours at 60$^{o}$C. The column was packed on a pressure vessel (400-700 psi) with ODS-AQ C18 reverse-

phase spherical silica particles, 5 µm diameter to a length of 6-6.5 cm. A nano-electrospray emitter tip (2-5 µm i.d.) was produced 2-3 mm above the frit with a P-2000 microcapillary laser puller. The AC was conditioned with tens of pmol of Angio/Vaso peptide mix repeatedly. A PC was connected to the back end of the AC using a 2 cm Teflon sleeve (0.060" o.d. x 0.12" i.d.). The combined ACPC was tested with 100 fmol of Angio/Vaso peptide mix on a mass spectrometer to ensure a good peptide signal intensity and chromatography (including Gaussian peak shape). This test was repeated on the day of each analysis.

### *2.3.4 Mass Spectrometry Analysis*

The dried $RGA^{GKG}$ sample was reconstituted in 0.1% acetic acid to a concentration of 500 fmol/µL RGA (according to the SDS-PAGE gel described previously), and vortexed. 1.0 pmol of RGA sample (2 µL) was pressure-loaded (~500 psi) onto the PC over 5-10 minutes, and then 100 fmol of Angio/Vaso peptide mix was also pressure-loaded to serve as quality indicators for chromatography, mass accuracy, and fragmentation. The PC was washed on the HPLC with 100% Solvent A (0.5% acetic acid in HPLC-grade $H_2O$) for 20-25 minutes (~35 bar) to desalt the column, then reattached to the AC (sometimes drying the PC first), and then 100% Solvent A was run through the ACPC to rehydrate the columns. The sample was then analyzed on the Orbitrap Fusion™ Tribrid™ Mass Spectrometer (Fusion) with a 0-60% Solvent B (66% acetonitrile, 33% $H_2O$, 0.5% acetic acid) in one-hour HPLC gradient. The gradient was as follows: start at 0% B, ramp 0-60% B over 1 hour, ramp 60-100% B over 2 min, hold at 100% B for 2 min, then ramp down 100-0% B over 2 min, and hold at 0% B for 20 min. The sample was electrosprayed from the column emitter tip at ~60-100 nL/min (achieved by splitting the solvent flow from the HPLC).

A tandem ESI-MS/MS experiment was performed. The instrument was operated in standard pressure mode, positive ion mode. High-resolution full mass (MS1) scans were performed in the Orbitrap (m/z range of 300-1500, quadrupole isolation, resolution 120,000 at m/z 200, precursor automatic gain control (AGC) target of 2E5 charges, max inject time 100 ms). Ions appearing in the MS1 scan were selected for low-resolution MS2 analysis in a data-dependent manner in order of decreasing intensity, and isolated by the quadrupole mass filter with a 2 Da window (monoisotopic precursor selection on). The experiment was performed with dynamic exclusion of repeat ions (repeat count of 1, repeat duration of 30 s, exclusion duration of 10 s). Precursors of charge 2-6 were subjected to CAD fragmentation (30% normalized collision energy) and MS2 analysis in the linear ion trap (normal scan speed, first mass 120 m/z, max inject time of 250 ms, AGC target of 1E4 charges). Precursors of charge state 3-6 were subjected to ETD fragmentation using calibrated charge-dependent reaction times and linear ion trap MS2 analysis (max inject time of 250 ms, AGC target of 1E4 charges). The cycle time was 2 s.

### *2.3.5 Data Analysis*

The data were visualized with Thermo Xcalibur Qual Browser software. Raw data files were searched using the Byonic search engine (26,27) against a database consisting of forward (and decoy reverse) sequences of $RGA^{GKG}$. The cleavage parameters were set for trypsin (C-terminal to R or K) with up to 3 missed cleavages. The precursor peptide parameters were as follows: mass less than 7,500 Da, mass error tolerance of $\pm$10 ppm, and isotope envelope adjustment of +1 or +2 allowed. The fragment mass error tolerance was $\pm$0.35 Da for both CAD and ETD. No FDR cutoff was applied. The data were searched with a fixed

modification of +57.0215 Da (carbamidomethyl, from IAM alkylation) on C. The variable modifications were +146.0579 (fucose) on S or T, +203.0794 (GlcNAc) on S or T, +162.0528 (hexose) on S or T, and + 79.9663 (phosphorylation) on S or T. Up to 12 variable modifications were allowed per precursor.

The Byonic peptide hits were manually validated by examination of the high-resolution MS1 and low-resolution fragment MS2 spectra. The amount of each peptide was quantified as the integrated peak area of the peptide. The ions selected for the peptide chromatogram consisted of all charge states and the isotope peaks of each charge state. The isotope peaks included were of intensity 10% or greater than the monoisotopic peak in the tallest charge state. The PTMs were site-localized manually as well. The percent abundance of each modified peptide relative to its unmodified form was calculated with the integrated peak areas.

## 2.4 Results

### *2.4.1 Spectral Interpretation*

In mass spectrometry, the exact isotope masses rather than the average mass of a compound are considered when individual isotopes are resolvable. The monoisotopic mass of a compound is composed of only the most common isotopes of each element (e.g. $^{12}C$, $^{14}N$). Other isotopes of the compound have one or more atoms of different isotopes (e.g. $^{13}C$, $^{15}N$). Carbon-12 ($^{12}C$, 98.9% of all carbon) is defined as exactly 12 Da. $^{13}C$ (1.07% of all carbon) has an exact mass of 13.0034 Da (28). Thus, a compound exists as an isotopic distribution, with successive isotopes differing by the mass of one neutron. For proteins, $^{13}C$ is the most important contributor to the relative abundance of each protein isotope, due to the high

number of carbon atoms and the relatively high abundance of $^{13}C$. The larger the protein, the more likely any given protein molecule will have one $^{13}C$ atom, two $^{13}C$ atoms, etc. (29,30).

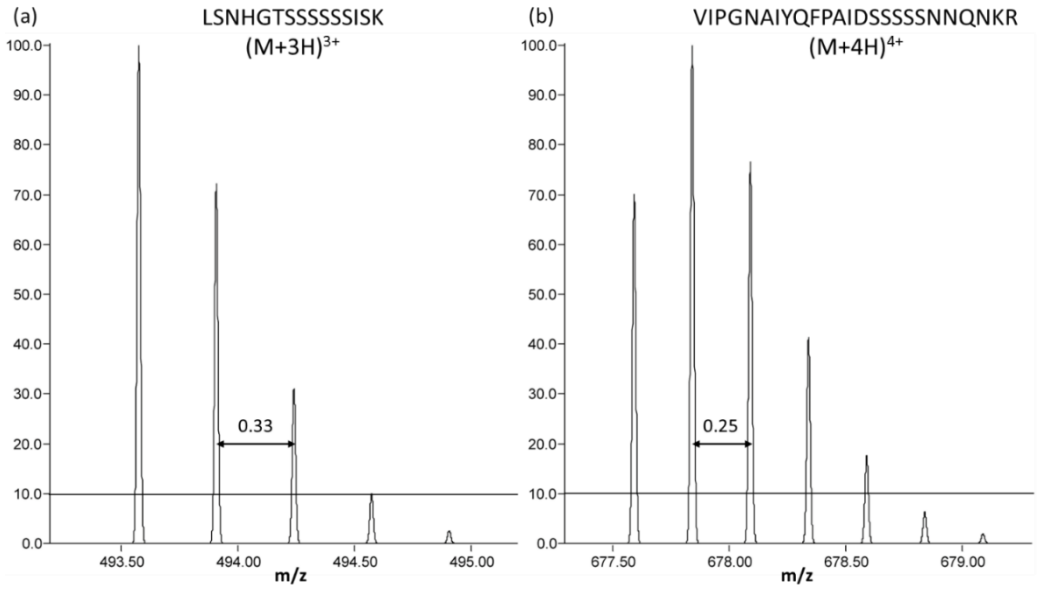


**Figure 2.3 Isotopic distributions.** The tallest isotope peak = 100%. The line at 10% indicates the lower bound of isotopes included in mass area calculations of peptide abundance. The spacing between adjacent isotope peaks is inversely proportional to the charge state. (a) Isotopic distribution of LSNHGTSSSSSSISK $(M+3H)^{3+}$. (b) isotopic distribution of VIPGNAIYQFPAIDSSSSSNNQNKR $(M+4H)^{4+}$. Figure made with the PNNL Molecular Weight Calculator program (30).

For example, as seen in Figure 2.3, the peptide LSNHGTSSSSSSISK ($C_{58}H_{100}N_{19}O_{26}$, monoisotopic mass 1478.7087) has relative isotopic abundances of 100%, 72%, 31%, 10%, etc. The monoisotopic peak is the tallest. VIPGNAIYQFPAIDSSSSSNNQNKR ($C_{116}H_{184}N_{35}O_{40}$, monoisotopic mass 2707.3439 Da), which has twice as many carbons, has relative isotopic abundances of 70%, 100%, 77%, 41%, 18%, 6.3%, etc. The one $^{13}C$ peak is the tallest. These values are predicted from the PNNL Molecular Weight Calculator program (30).

In ESI mass spectrometry, the peptide ions only rarely exist in the +1 charge state; they usually have multiple charges due to varying numbers of $H^+$ adducts. The +1 charge state is $(M+H)^+$, the +2 charge state is $(M+2H)^{2+}$, and the +n charge state is $(M+nH)^{n+}$. The isotopic distribution abundances are essentially unchanged by the addition of one or more $H^+$, but the

spacing between adjacent peaks decreases. The spacing is inversely proportional to the charge. Figure 2.3a has the 0.33 m/z spacing of an $(M+3H)^{3+}$ ion; Figure 2.3b has the 0.25 m/z spacing of an $(M+4H)^{4+}$ ion. Thus, the spacing of the isotope peaks of an ion in a spectrum can be used to calculate the mass of the uncharged species. Whether the individual isotope peaks are discernable in a mass spectrum depends on the resolution of the mass analyzer. Resolution (r) is defined in Equation 2.1 below, where FWHM is the full width of a peak at half of its maximum intensity (29).

Equation 2.1: $$r = \frac{(m/z)}{FWHM}$$

Equation 2.2: $$FWHM = \frac{(m/z)}{r}$$

For a mass analyzer, the resolution is given at a particular m/z value. High resolution mass analyzers such as an Orbitrap can achieve resolutions of over 480,000 at 200 m/z. For the Orbitrap, the resolution depends on the transient time, as more axial oscillations increase resolution. For example, the Orbitrap in Fusion has a resolution of 120,000 at 200 m/z with a 256 ms transient time, giving a mass accuracy of <5 ppm. Resolution in the Orbitrap increases linearly with transient time, but decreases in proportion to the square root of the ion m/z (31,32). A resolution of 120,000 at 200 m/z corresponds to r = 60,000 at 800 m/z, and r = 43,800 at 1500 m/z. For the 800 m/z ion with r = 60,000, the FWHM is 0.013 m/z (Equation 2.2), which is more than sufficient to resolve the isotope peaks of even a +10 species (0.10 m/z spacing). In contrast, LITs are low-resolution mass analyzers. The 800 m/z ion with a resolution of 2000 has a FWHM of 0.4 m/z, at which the isotopes of +1 ions are resolved, but the isotopes of +2 ions meld into one taller peak. The melded peak has an apparent m/z value in between the constituent isotopes (29).

The mass accuracy of an observed ion is the difference between the observed m/z value and the theoretical, calculated m/z value. For high resolution instruments the mass accuracy is expressed in parts per million (ppm; Equation 2.3). The Orbitrap Fusion has a mass accuracy of <5 ppm, as seen in these experiments. This high resolution and mass accuracy allows for data searches with narrow mass error tolerances for precursor ions, which significantly decreases the number of potential peptides considered (29,32).

Equation 2.3: $$ppm = \frac{(m/z_{observed} - m/z_{theoretical})}{m/z_{theoretical}} * 10^6$$

Figure 2.4 and Table 2.1 display examples of the chromatograms and mass spectra used in identifying a peptide, its abundance and site(s) of post-translational modification. The chromatograms display the abundance of a certain ion or set of ions over time, while the high-resolution MS1 displays all the ions detected at a given time, and the low-resolution MS2 displays all the fragment ions produced from a precursor ion selected based on the MS1. The peptide examined is O-GlcNAcylated LSNHGTSSSSSSISK (residues 12-26 of RGA), hereafter LSN15. From the top, the first two chromatograms in Figure 2.4a show the abundance of unmodified and O-GlcNAcylated peptide (respectively) over time. The LSN15 peptide elutes from the HPLC-connected C18 analytical column and is electrosprayed into the mass spectrometer and subjected to mass analysis. The O-GlcNAcylated form of LSN15 starts eluting about one minute after the unmodified form starts, but their elution profiles overlap significantly. However, mass analysis provides an orthogonal axis of separation, as the two forms have different m/z values (493.5742 and 561.2673 for the +3 charge state of the unmodified and O-GlcNAcylated forms, respectively).

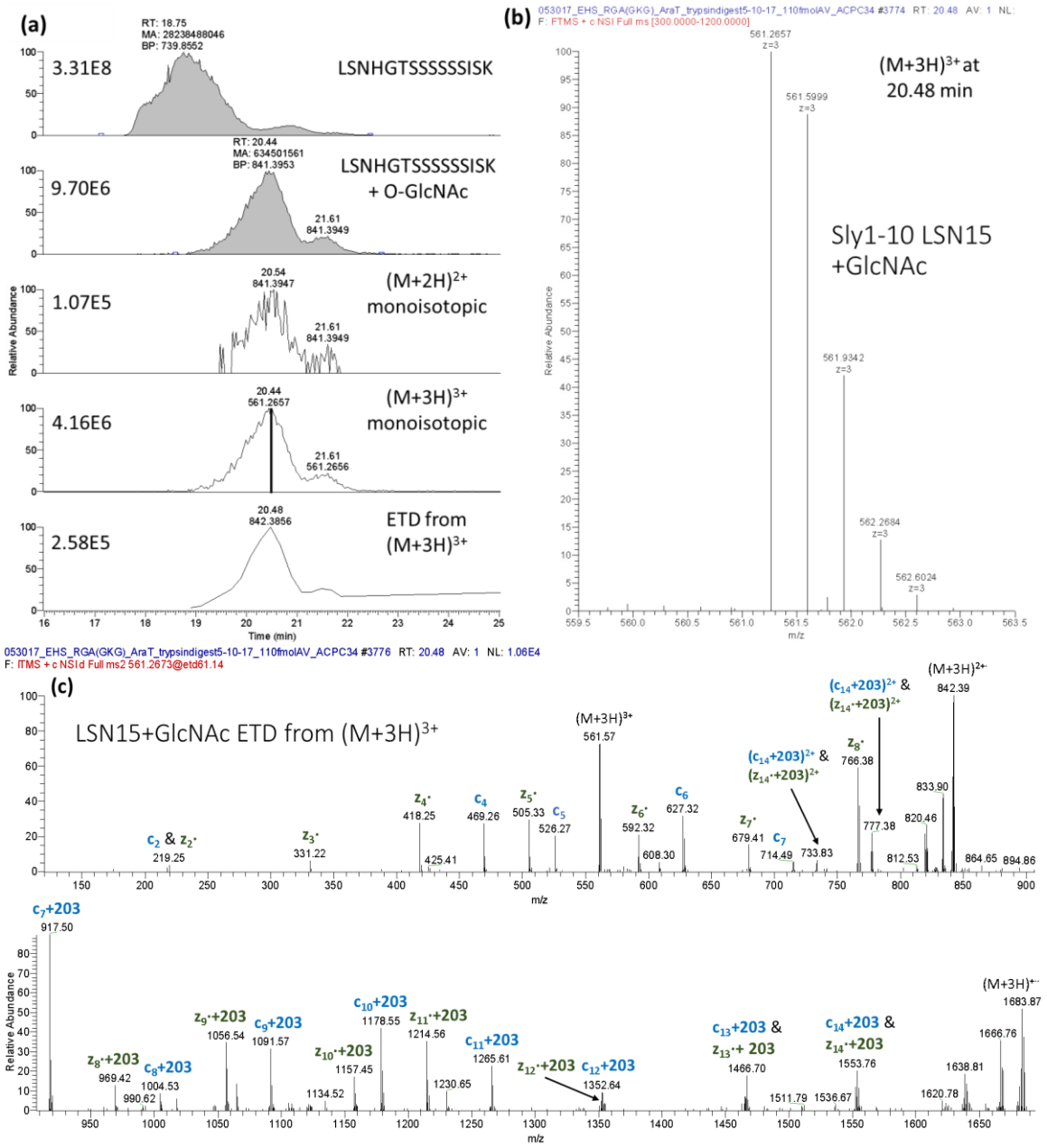


**Figure 2.4 Chromatograms and MS of O-GlcNAcylated LSNHGTSSSSSSISK from *sly1-10* (5/17).** (a) Chromatograms. The top 2 chromatograms are of LSNHGTSSSSSSISK unmodified and O-GlcNAcylated, respectively (the peaks include all major charge states and isotopes (10% cutoff)). RT – chromatographic retention time. MA – mass area of the peak (abundance of the peptide). The third and fourth chromatograms are the monoisotopic peaks of $(M+2H)^{2+}$ and $(M+3H)^{3+}$, respectively, of O-GlcNAcylated LSNHGTSSSSSSISK. The bottom chromatogram is the ETD MS2 of the $(M+3H)^{3+}$ precursor. The numbers on the left of each chromatogram are the intensity of the highest peak in each chromatogram. (b) The high-resolution MS1 of $(M+3H)^{3+}$ eluting at 20.48 min, showing the isotopic distribution. This time is indicated in the fourth chromatogram of (a) by a solid line. (c) The ETD MS2 spectrum from $(M+3H)^{3+}$ at 20.48 min. c ions are shown in blue, and z· ions in green. O-GlcNAcylated ions are indicated by "+203."

| 2+ c ions + OGlcNAc | 2+ c ions | 1+ c ions + OGlcNAc | 1+ c ions | c | Sequence | z· | 1+ z· ions | 1+ z· ions + OGlcNAc | 2+ z· ions | 2+ z· ions + OGlcNAc |
|---|---|---|---|---|---|---|---|---|---|---|
| | 66.10 | | 131.12 | 1 | L | 15 | 1478.71 | 1681.79 | 740.28 | 841.88 |
| 211.24 | 109.64 | 421.23 | 218.15 | 2 | S | 14 | 1349.61 | 1552.68 | 675.69 | 777.29 |
| 268.29 | 166.69 | 535.27 | 332.19 | 3 | N | 13 | 1262.57 | 1465.65 | 632.15 | 733.75 |
| 336.86 | 235.26 | 672.33 | 469.25 | 4 | H | 12 | 1148.53 | 1351.61 | 575.09 | 676.69 |
| 365.39 | 263.79 | 729.35 | 526.27 | 5 | G | 11 | 1011.47 | 1214.55 | 506.52 | 608.12 |
| 415.94 | 314.34 | 830.40 | 627.32 | 6 | T | 10 | 954.45 | 1157.53 | 478.00 | 579.60 |
| 459.48 | 357.88 | **917.43** | **714.35** | 7 | **S** | 9 | 853.40 | 1056.48 | 427.45 | 529.05 |
| 503.02 | 401.42 | 1004.46 | 801.39 | 8 | **S** | 8 | **766.37** | **969.45** | 383.91 | 485.51 |
| 546.56 | 444.96 | 1091.50 | 888.42 | 9 | S | 7 | 679.34 | 882.42 | 340.37 | 441.97 |
| 590.10 | 488.50 | 1178.53 | 975.45 | 10 | S | 6 | 592.31 | 795.39 | 296.83 | 398.43 |
| 633.64 | 532.04 | 1265.56 | 1062.48 | 11 | S | 5 | 505.27 | 708.35 | 253.29 | 354.89 |
| 677.18 | 575.58 | 1352.59 | 1149.51 | 12 | S | 4 | 418.24 | 621.32 | 209.75 | 311.35 |
| 733.76 | 632.16 | 1465.68 | 1262.60 | 13 | I | 3 | 331.21 | 534.29 | 166.21 | 267.81 |
| 777.30 | 675.70 | 1552.71 | 1349.63 | 14 | S | 2 | 218.13 | 421.21 | 109.63 | 211.23 |
| 841.88 | 740.28 | 1681.79 | 1478.71 | 15 | K | 1 | 131.09 | | 66.09 | |

**Table 2.1 ETD sequence coverage of O-GlcNAcylated LSNHGTSSSSSSISK at 20.48 min.** Grey cells indicate observed ions. The columns flanking the "Sequence" column give the number of residues in the c or z· ion (these numbers do not correspond to the residue number in the protein itself). The columns with "+ OGlcNAc" give the mass of the modified fragment ion; the other columns give the mass of the unmodified fragment ion. Red boxes indicate the product ions unique to gS7; blue circles indicate the product ions unique to gS8.

Each peptide is present at several different charge states; the monoisotopic peak of +2 and +3 charge states of O-GlcNAcylated LSN15 are shown as the third and fourth chromatograms in Figure 2.4a, respectively. The figure shows the maximum peak intensity for the graphed ion(s) in each chromatogram to the left of the peak. For O-GlcNAcylated LSN15, the +2 charge state is only one fortieth as tall as the +3 charge state. Figure 2.4b shows a portion of the MS1 acquired at 20.48 min, zoomed in on the +3 charge state of LSN15 + GlcNAc to show the isotopic distribution (this time is indicated in the fourth chromatogram by a black line). In order to calculate the abundance of a peptide, all peaks of 10% abundance relative to the monoisotopic peak of the most abundant charge state (as well as all detected monoisotopic peaks without that abundance) are included in one chromatogram (the second chromatogram above), from which the integrated peak area is recorded. For the +3 charge state, this includes the first four isotope peaks.

Since ETD is necessary for site localization (and effective ETD requires a higher charge density, so higher charge states are beneficial), the occurrence of ETD MS2 spectra with the +3 charge state as the precursor is plotted in the bottom chromatogram in Figure 2.4a. The precursor is isolated with a 2 m/z window centered around the monoisotopic peak. The annotated ETD MS2 spectrum taken at 20.48 min is shown in Figure 2.4c. In order to evaluate the spectrum, a list of possible c and z· ions is generated with the Fragment Calculator program developed in house by Dina Bai. Table 2.1 lists all possible fragment ions of unmodified and O-GlcNAcylated LSN15. Since the O-GlcNAc could potentially be located on any of the nine serines and threonines, all possible O-GlcNAcylated product ions are included in the table. The observed ions are colored grey to indicate coverage. Interestingly, this MS2 contains evidence of two different O-GlcNAc sites: S7 and S8 of the peptide (S18 and S19 in the full protein). The ions unique to O-GlcNAcylation on S7 (gS7) are boxed in red, while the ions unique to gS8 are circled in blue. All other observed ions would be present in either isoform. If the only isoform of LSN15 was gS7, then no $c_7$ would be observed, only $c_7$+203. Conversely, if only gS8 was present, then the opposite would be true. Since both $c_7$ and $c_7$+203 ions are present in this spectrum (and similarly $z_8$· and $z_8$·+203), both gS7 and gS8 forms of LSN15 are present in the sample and arriving in the mass spectrometer at the same time. Because both isoforms have the same mass, they are indistinguishable in MS1 spectra. In the second chromatogram in Figure 2.4a there are two overlapping peaks, at 20.4 min and 21.6 min. However, these two peaks do not separately represent gS7 vs. gS8; rather, the 21.6 min peak contains the gT6, gS8, and probably gS7 isoforms (in this case, $c_7$ and $z_8$· would be present for either gT6 or gS7).

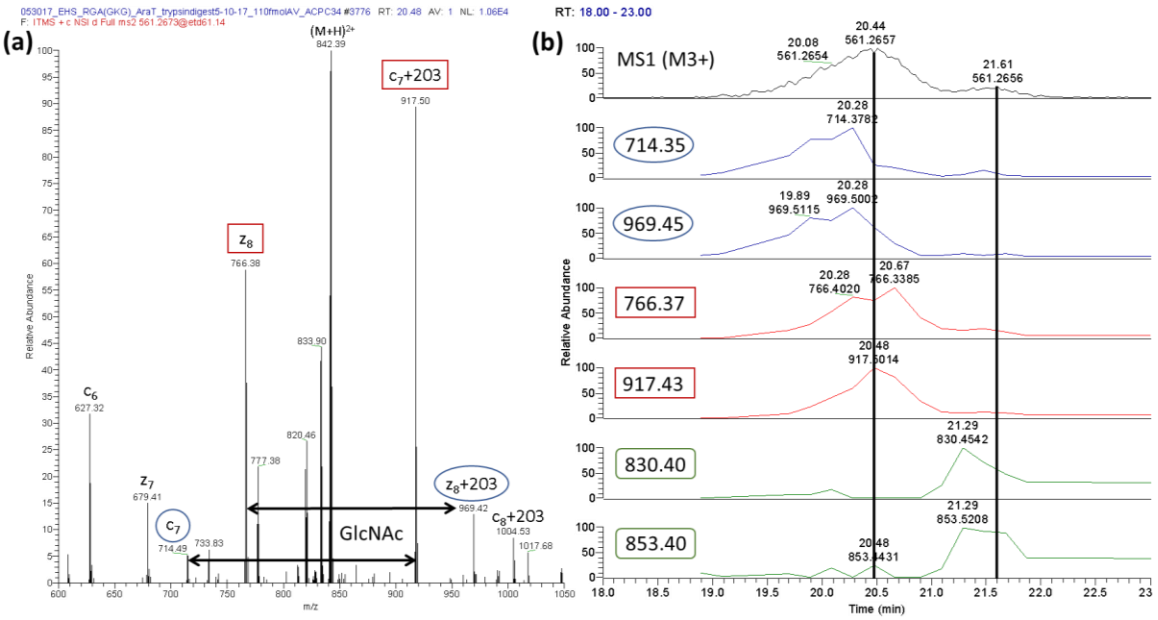


**Figure 2.5 O-GlcNAcylation of LSNHGTSSSSSSISK at different sites**. (a) ETD MS2 at 20.48 min – gS7 is present at a higher level than gS8 at this time. Ions indicative of gS7 are boxed in red; ions indicative of gS8 are circled in blue. (b) Chromatograms of characteristic MS2 ions for gS8, gS7, and gT6 showing peak retention times. Black vertical lines indicate the peaks at 20.48 min and 21.61 min retention times. Ion abundances between chromatograms are not to scale.

The rough relative amounts of isoforms at any given time can be estimated by examining the MS2s more closely. Looking at Figure 2.5a, the $z_8$· and $c_7$+203 ions (766.38 and 917.50) unique to gS7 are more abundant than $z_8$·+203 and $c_7$ ions (969.42 and 714.49) unique to gS8. Producing chromatograms of the characteristic ETD MS2 ions over time shows that the three isoforms elute at slightly different but overlapping times (Figure 2.5b). gS7 peaks at 20.5 min, gS8 slightly earlier near 20.3 min, and gT6 at 21.5 min. gS7 and gS8 are close enough in retention time that they meld into one chromatographic peak, albeit one with a long forward tail. Determining precise relative levels between the isoforms from this present data is not possible; better chromatographic separation would be required. Thus, all peptide modification levels in this work are reported without regard for different site isoforms (e.g. the overall level of di-phosphorylated LSN is reported, not broken down by site(s)).

The Byonic search engine (26,27) (Protein Metrics, Inc) was used to examine data. Byonic is a hybrid search engine that combines limited de novo sequencing with database searching. Database search algorithms such as MASCOT or SEQUEST compare a list of the ions of a spectrum to in-silico-digested peptides from a protein database (which gives a list of predicted ions). These programs are less practical for searching peptides with a wide variety of possible PTMs, as the addition of each new PTM increases the search time exponentially. De novo programs (which do not use protein databases) may be better at detecting modified peptides, but they require high-quality spectra to be effective (26,27,29).

Byonic examines a spectrum to identify likely b and y ions (for CAD) or c and z· ions (for ETD), along with associated ions (such as ones that underwent the neutral loss of water in CAD). These peaks are then used with the precursor ion mass and other parameters (such as digest parameters) to identify candidate peptides from a database of protein sequences. Byonic allows one to search for various PTMs, which can be grouped by rarity. PTMs can be fixed or variable. In digests of IAM-alkylated proteins, carbamidomethylation is set as a fixed modification expected to be present on all peptides with a cysteine; on the other hand, sub-stoichiometric PTMs like phosphorylation and O-GlcNAcylation are set as variable mods that may or may not be present on a given peptide. The search engine allows input for the digestion parameters (trypsin would be C-terminal to R or K, for instance), with an allowance for number of missed cleavages permitted (in case of incomplete digestion). Sometimes the precursor chosen for fragmentation by the instrument is not the monoisotopic peak (especially for larger peptides where the tallest peak in the isotope envelope is not the monoisotopic one), so Byonic allows for a parameter ("off by x") to adjust the isotope envelope $\pm 1$ or more Da to the actual monoisotopic (all $^{12}C$) peak. For Byonic, a higher peptide

score is a better hit (the score is based on the quality of a peptide-spectrum match, not relative to other peptides that might match that spectrum). The Byonic results file also lists the difference in score from the top-scoring peptide to the next highest for a given spectrum, either with or without regard for PTM site-localization differences (26,27).

Even with search programs like Byonic, however, site-localizing posttranslational modifications is difficult, as detailed above. The Byonic scores for each potential site-localization indicate a likely site, but the highest scoring site is often incorrect, especially for RGA. Manual verification of the site(s) is required. RGA is remarkable for the long poly-S/T stretches in the N-terminal region. Stretches of 6 to 8 serines or threonines exist in all types of organisms and multiple classes of proteins (such as structural proteins, transcription factors, other DNA-binding proteins, and kinases), but regions this extensive are rare (33). The LSNHGTSSSSSSISK peptide has S or T at 9 of its 15 residues (7 in a row); GVIGTTVTTTTTTTTAAGESTR (GVIG) has S or T at 12 of its 22 residues (8 in a row, after TTV). GVIG is the last of three consecutive poly-S/T peptides (VIPGNAIYQFPAIDSSSSSNNQNK(R), (LK)SCSSPDSMVTSTSTGTQIGK, and then GVIGTTVTTTTTTTTAAGESTR). The residue(s) in parentheses indicate longer forms of the respective peptides that occur due to incomplete trypsin cleavage; both forms are routinely seen. Any of these serines and threonines may be modified. Additionally, several sites can be modified on one protein molecule simultaneously, with either the same or different PTMs. For example, the GVIGTTVTTTTTTTTAAGESTR peptide often has one or more GlcNAcs in the poly-T region, and often with a simultaneous hexose. In the same fashion as O-GlcNAcylated LSNHGTSSSSSSISK above, all isoforms of mono-O-GlcNAcylated GVIGTTVTTTTTTTTAAGESTR have the same mass, and the chromatogram of the appropriate mass in the MS1 spectra usually has a multi-humped peak, as they all have a

similar retention time (often differing by 1 minute or less). GVIGTTVTTTTTTTTAAGESTR is often more difficult to analyze than LSNGHTSSSSSSISK because two threonines (202.0954 Da) have a similar mass to one GlcNAc (203.0794 Da) that can be difficult to distinguish with low-resolution MS2s. Site-localizing PTMs is even more difficult with multiply-modified peptides, as exponentially more isoforms are possible with each additional PTM.

### *2.4.2 Arabidopsis Mutant Samples*

Eleven *Arabidopsis* mutant samples were digested with trypsin and analyzed on the Orbitrap Fusion mass spectrometer. For each sample, the data was searched against the 6His-3xFLAG-RGA$^{GKG}$ sequence using Byonic with phosphorylation, O-GlcNAcylation, O-fucosylation, and O-hexosylation as variable modifications. Peptides, their PTM(s), and PTM site(s) were manually verified, and their abundances determined from their chromatographic peak areas as described previously. The abundances of each peptide and their PTM variants are given in Table 2.2; in a given sample, the percent abundances are calculated relative to the total abundance of all listed variants of that peptide. For example, in *sly1-10 spy-8*, mono-O-GlcNAcylated LSN comprises 2.5% of all LSN forms listed. In Table 2.2, "ND" means a peptide form was not detected, with a limit of detection of mass area ~$1*10^6$. The peptide might possibly be present, but in a very low abundance. Grey cells represent modifications that did not have any accompanying CAD or ETD MS2s, but which had a similar chromatographic retention time to the modified peptide in other samples and a ppm mass error similar to other peptides in that sample. A table with the percent abundances for each individual sample is found in the Appendix.

| LSNHGTSSSSSSISK(DK) | *ga1-3* avg | *sly1-10* avg | *sly1-10 spy-8* | *sly1-10 spy-12* | *sly1-10 spy-15* | *sly1-10 spy-19* | *sly1-10 sec-3* avg |
|---|---|---|---|---|---|---|---|
| Unmodified | 55% | 93% | 85% | 89% | 74% | 86% | 84% |
| + Fucose | 1.5% | 0.4% | 1.7% | ND | <0.1% | ND | 1.0% |
| + Hexose | ND | <0.1% | ND | <0.1% | ND | <0.1% | ND |
| + GlcNAc | 6.4% | 1.8% | 2.5% | 8.1% | 2.6% | 10.1% | ND |
| + PO4 | 17.0% | 3.1% | 5.3% | 2.2% | 19.5% | 2.9% | 8.1% |
| + 2*PO4 | 15.3% | 1.1% | 3.0% | 0.1% | 2.2% | 0.3% | 3.9% |
| + 3*PO4 | 2.8% | <0.1% | 0.2% | ND | ND | ND | 0.4% |
| + GlcNAc + PO4 | 0.5% | <0.1% | 0.5% | <0.1% | 0.3% | ND | ND |
| + acetyl | 1.5% | 0.6% | 1.5% | 0.2% | 1.1% | 0.2% | 2.9% |
| VIPGNAIYQFPAIDSSSSSNNQNK(R) | *ga1-3* avg | *sly1-10* avg | *sly1-10 spy-8* | *sly1-10 spy-12* | *sly1-10 spy-15* | *sly1-10 spy-19* | *sly1-10 sec-3* avg |
| Unmodified | 97% | 100% | 100% | 99% | 100% | 100% | 100% |
| + GlcNAc | ND | ND | ND | ND | ND | ND | ND |
| + PO4 | 2.0% | 0.2% | 0.2% | <0.1% | 0.2% | 0.1% | 0.2% |
| + 2*PO4 | 0.4% | 0.2% | ND | 1.0% | 0.1% | ND | ND |
| + Acetyl | 0.3% | <0.1% | <0.1% | ND | <0.1% | ND | <0.1% |
| (LK)SCSSPDSMVTSTSTGTQIGK | *ga1-3* avg | *sly1-10* avg | *sly1-10 spy-8* | *sly1-10 spy-12* | *sly1-10 spy-15* | *sly1-10 spy-19* | *sly1-10 sec-3* avg |
| Unmodified | 65% | 72% | 75% | 80% | 75% | 74% | 75% |
| + Fucose | ND | ND | ND | ND | ND | ND | ND |
| + GlcNAc | 4.3% | 3.7% | 5.3% | 4.6% | 5.1% | 4.4% | ND |
| + PO4 | 29.4% | 24.6% | 19.9% | 15.2% | 19.6% | 22.0% | 24.6% |
| + 2*PO4 | 1.7% | <0.1% | ND | ND | ND | ND | ND |
| GVIGTTVTTTTTTTTAAGESTR | *ga1-3* avg | *sly1-10* avg | *sly1-10 spy-8* | *sly1-10 spy-12* | *sly1-10 spy-15* | *sly1-10 spy-19* | *sly1-10 sec-3* avg |
| Unmodified | 24% | 33% | 29% | 30% | 27% | 30% | 100% |
| + Fucose | ND | <0.1% | ND | ND | ND | ND | 0.1% |
| + Hex | 6.2% | 11.8% | 9.2% | 12.0% | 9.3% | 4.1% | <0.01% |
| + 2*Hex | 0.3% | 0.3% | 0.3% | 0.5% | 0.4% | 0.6% | ND |
| + GlcNAc | 41.6% | 41.9% | 43.6% | 45.4% | 41.9% | 41.3% | 0.2% |
| + 2*GlcNAc | 16.0% | 8.0% | 10.8% | 7.2% | 13.2% | 13.1% | ND |
| + 3*GlcNAc | 3.0% | 0.6% | 1.8% | 0.8% | 2.0% | 2.2% | ND |
| + 4*GlcNAc | 0.1% | ND | ND | ND | ND | ND | ND |
| + Hex + GlcNAc | 7.1% | 3.8% | 4.4% | 4.0% | 5.4% | 6.9% | ND |
| + 2*Hex + GlcNAc | 0.2% | <0.01% | <0.1% | ND | ND | 0.1% | ND |
| + Hex + 2*GlcNAc | 1.6% | 0.4% | 1.0% | 0.5% | 1.1% | 1.8% | ND |
| + Fucose + GlcNAc | 0.2% | 0.2% | ND | ND | ND | ND | ND |
| + PO4 | ND | ND | ND | ND | ND | ND | ND |

**Table 2.2 Relative levels of post-translational modifications of RGA in *Arabidopsis* mutants.** ND: not detected (threshold ~$1*10^6$). Grey cells: no MS2s from the peptide with this modification were acquired in this sample, but it had a comparable retention time to the other samples and low ppm mass error. The *ga1-3* and *sly1-10* columns are averages of three different experiments (one sample with a 300-1200 m/z window, two with a 300-1500 m/z window (one by Speers, one by Zahn)). The *sly1-10 sec-3* column is an average of two different experiments (one by Speers, one by Zahn). The PTM levels for each individual sample are found in the Appendix.

The eleven *Arabidopsis* mutant samples consisted of seven different mutants, summarized in Table 2.3. When there were multiple samples of a given mutant, they were designated with the month and year (e.g. 7/17). There were two samples of *ga1-3* (7/17 and 8/18, which was analyzed twice), three samples of *sly1-10* (5/17, 12/17, and 3/18), one sample each of *sly1-10 (spy-8/-12/-15/-19)*, and two samples of *sly1-10 sec-3* (4/18 and 8/18). The *sly1-10* and *ga1-3* single mutants samples had wild-type SPY and SEC. The *sly1-10 spy-8/-12/-15/-19* double mutants had wild-type SEC, and *sly1-10 sec-3* had wild-type SPY.

| Mutant | First Mutation | Second Mutation | GID1 Binding | SLY1 Binding | SPY POFUT Activity | SEC OGT Activity |
|---|---|---|---|---|---|---|
| *ga1-3* | Null *ga1* | - | No | No | Yes | Yes |
| *sly1-10* | *sly1* C-term | - | Yes | No | Yes | Yes |
| *sly1-10 spy-8* | *sly1* C-term | *spy* 354-376 deletion | Yes | No | Reduced | Yes |
| *sly1-10 spy-12* | *sly1* C-term | *spy* G570D | Yes | No | No | Yes |
| *sly1-10 spy-15* | *sly1* C-term | *spy* E567K | Yes | No | No | Yes |
| *sly1-10 spy-19* | *sly1* C-term | *spy* K665M | Yes | No | No | Yes |
| *sly1-10 sec-3* | *sly1* C-term | Null *sec* | Yes | No | Yes | No |

**Table 2.3 Summary of mutants.** "Null" indicates a mutation in which the gene is not expressed at all. *sly1* C-term: deletion of the last 8 residues and addition of 46 AA to the C-terminal target interaction domain of *sly1*. POFUT: protein O-fucosyltransferase. OGT: O-GlcNAcyltransferase.

Two of the samples (the 5/17 *sly1-10* and 7/17 *ga1-3* samples) were analyzed with an MS1 window of 300 to 1200 m/z; the rest were analyzed with an MS1 window of 300 to 1500 m/z. Additionally, the *sly1-10* (12/17) and *sly1-10 spy-12* samples were analyzed with the QMF operated in mass filter mode (with both RF AC and DC voltages active). The *sly1-10* (5/17 and 3/18), *ga1-3*, and *sly1-10 (spy-8/-15/-19* or *sec-3*) samples were analyzed with the QMF in ion transmission mode (RF AC only operation).

All instrument runs had an average ppm mass error of $\pm$3.5 or smaller. This was due to the high mass accuracy of Fusion; the average ppm mass error was due to gradual drifts

over months that are periodically corrected with recalibration. The standard deviation of the average ppm mass error was less than 1.0 ppm for each run, which aids in the identification of one peptide and/or PTM variant over another for a particular mass spectrum.

The peptides discussed in this work are as follows: LSNHGTSSSSSSISK(DK) (residues 12 to 28 of wild-type RGA; abbreviated LSN) was detected in both LSNHGTSSSSSSISK (LSN15) and LSNHGTSSSSSSISKDK (LSN17) forms. VIPGNAIYQFPAIDSSSSSNNQNK(R) (residues 140 to 164 of wild-type RGA; abbreviated VIP) was detected in both VIPGNAIYQFPAIDSSSSSNNQNK (VIP24) and VIPGNAIYQFPAIDSSSSSNNQNKR (VIP25) forms. (LK)SCSSPDSMVTSTSTGTQIGK (residues 165-185 of wild-type RGA with the inserted K at the end; abbreviated SCSS) was detected in both SCSSPDSMVTSTSTGTQIGK (SCSS20) and LKSCSSPDSMVTSTSTGTQIGK (SCSS22) forms. The GVIGTTVTTTTTTTTAAGESTR peptide consisted of residues 186 to 207 of wild-type RGA, and was abbreviated GVIG. The PTMs studied here are transient, reversible modifications that may only be present on RGA in specific tissues, points in time (cell cycle, plant life cycle, under certain stimuli), etc. (34). Thus, low-abundance modifications can still be important. A two-fold or greater difference in percent abundance between two mutants was considered significant here.

### *2.4.3 GlcNAc, Fucose, and Hexose*

As expected of a mutant with a null *sec* mutation (thus lacking OGT activity) (18), GlcNAcylation was not detected in *sly1-10 sec-3* at LSNHGTSSSSSSISK(DK), (LK)SCSSPDSMVTSTSTGTQIGK, and was scarcely present in GVIGTTVTTTTTTTTAAGESTR. GlcNAcylated VIPGNAIYQFPAIDSSSSSNNQNK(R) was not observed in any sample.

At LSNHGTSSSSSSISK(DK), GlcNAcylation and fucosylation were increased over three-fold in *ga1-3* relative to *sly1-10*. GlcNAcylation was elevated in the *sly1-10 spy* double mutants relative to *sly1-10*, but only significantly in *sly1-10 spy-12/-19*. Fucosylation was significantly elevated in *ga1-3*, *sly1-10 spy-8*, and *sly1-10 sec-3* relative to *sly1-10*, but significantly decreased in the other *sly1-10 spy* mutants. Hexosylation was scarcely present in any sample.

At (LK)SCSSPDSMVTSTSTGTQIGK, there was no fucosylation detected. The cysteines of RGA (including the one in SCSS) were reduced and then alkylated to the carbamidomethylated form (+$C_2H_3NO$, +57.0214) to break any disulfide bonds and prevent their reformation. The atomic composition of a carbamidomethylated, fucosylated peptide (+$C_6H_{10}O_4$ + $C_2H_3NO$ = +$C_8H_{13}NO_5$; +146.0579 + 57.0214 = +203.0794) is identical to a GlcNAcylated (but not carbamidomethylated) peptide (+$C_8H_{13}NO_5$, +203.0794 Da). Since the two variants are identical in mass, they are nearly impossible to distinguish without MS2 fragmentation. O-GlcNAcylated peptides subjected to CAD or HCD fragmentation readily lose the GlcNAc, which appears in the MS2 spectrum as an oxonium ion at m/z 204.09, $(C_8H_{13}NO_5+H)^+$ (35). However, small m/z ions are only seen in ion trap CAD if the low-mass cutoff (LMCO) is small enough. The LMCO is determined by the precursor m/z and the *q* value at which it is activated (here *q* = 0.25). At a given activation frequency, *q* increases as the m/z of an ion decreases (they are inversely proportional), but ions are unstable near *q* = 0.908 and will not be trapped efficiently for subsequent detection. The LMCO is thus usually about one third of the precursor ion m/z, or about 0.3 for *q* = 0.25 (36,37), which for GlcNAcylated (alkylated or non) SCSS falls near or above 204 m/z.

CAD provides other ways to distinguishing the two PTM forms. The loss of the O-GlcNAc from the precursor also produces a prominent $(M+nH–203)^{n+}$ peak, which is often the

largest in the CAD spectrum. The loss of fucose from a fucosylated peptide would similarly give $(M+nH–146)^{n+}$. Since CAD results in the loss of sugar modifications, but not S-carbamidomethylation of cysteine, the alkylated fucosylated peptide form can be distinguished from the non-alkylated GlcNAcylated form via the presence or absence of carbamidomethyl on the $(M\text{-sugar})^{n+}$ peak. For example, SCSS22 + 203 (LKSCSSPDSMVTSTSTGTQIGK + 203) would have a CAD peak at 1136.54 $(M+2H–146)^{2+}$ if it was alkylated fucosylated, but it would have a 1108.03 $(M+2H–203)^{2+}$ peak if it was non-alkylated GlcNAcylated (the difference corresponds to the presence or absence of the carbamidomethyl, respectively).

Additionally, in CAD or HCD, carbamidomethylated SCSS will have b and y ions at/after the C with the alkylation (+57), while non-carbamidomethylated SCSS will have those ions without +57 (this is more readily observed with the b ion series, since for SCSS22 $b_4$ to $b_{22}$ will be affected rather than just $y_{19}$ to $y_{22}$ for the y ion series). If the GlcNAcylated amino acid is not close to the cysteine, intermediate ETD fragment ions would also show a difference (for SCSS, c ions at/after the cysteine would not show the +57, and z· ions would show +203 rather than +146 at/after the GlcNAc – see Figure 2.6). However, if the GlcNAcylated S/T and the C are close to each other, this difference is difficult to observe with incomplete sequence coverage. Overall, while the loss of the sugar in CAD curtails its usefulness in site-localization, it is useful for distinguishing alkylated fucosylated and non-alkylated GlcNAcylated peptides (35). Upon inspection, all of the SCSS + 146 in the samples was non-alkylated GlcNAcylated, rather than alkylated fucosylated, so the mass areas of the alkylated and non-alkylated GlcNAcylated peptide forms (+146 and +203, respectively) were added together in Table 2.2. There were no significant changes in the GlcNAcylation levels of SCSS.

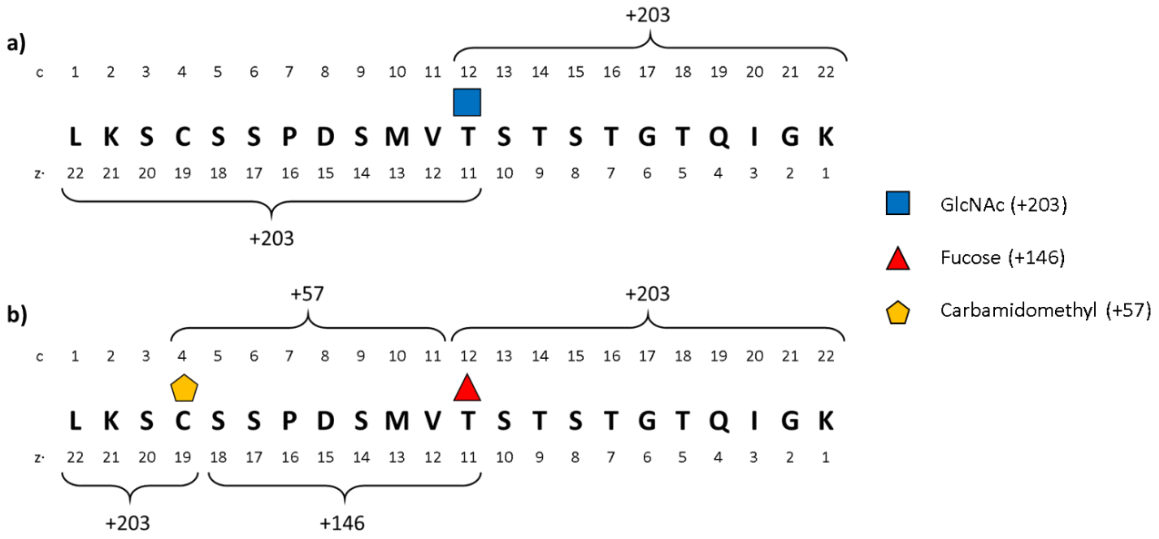


**Figure 2.6 Distinguishing LKSCSSPDSMVTSTSTGTQIGK + 203 forms.** While the MS1 masses are the same, MS2 fragmentation is able to distinguish between (a) non-alkylated GlcNAcylated and (b) alkylated fucosylated SCSS22. The fragment ions that would have masses above the unmodified version are indicated with brackets (e.g. $z_{\cdot 11} - z_{\cdot 18}$ of alkylated fucosylated SCSS22 has +146 from the fucose on T11). Blue square: GlcNAc. Red triangle: fucose. Yellow pentagon: carbamidomethyl.

GVIGTTVTTTTTTTTAAGESTR GlcNAcylation was seen across all the different mutants, with the exception of its virtual absence in *sly1-10 sec-3*. The levels of GVIG with one, two, three, or four GlcNAcs across the different samples are illustrated in Figure 2.7. GVIGTTVTTTTTTTTAAGESTR had no fucosylation detected, except in *sly1-10* and *sly1-10 sec-3*. The fucosylation level was higher in *sly1-10 sec-3*, but still only ~0.1%. GVIG had high levels of GlcNAcylation in *ga1-3*, *sly1-10*, and all the *sly1-10 spy* mutants, with over half the GVIG present having at least one GlcNAc. Hexose levels were also notable, with ~16% of GVIG having at least one hexose in each mutant besides *sly1-10 sec-3*. Along with the high levels of GlcNAcylation and hexosylation, PTM forms with multiple GlcNAcs and/or hexoses were found, giving multiple PTM forms in Table 2.2. Mono-GlcNAcylation levels were very similar (about 43%), though poly-GlcNAcylation levels varied across the samples. However, GVIG quantitation was more difficult due to the presence of a long tail trailing after the peak of

more abundant PTM variants (frequently stretching towards the end of the HPLC gradient), and the not-insignificant percentage of the $(M+2H)^{2+}$ charge state.

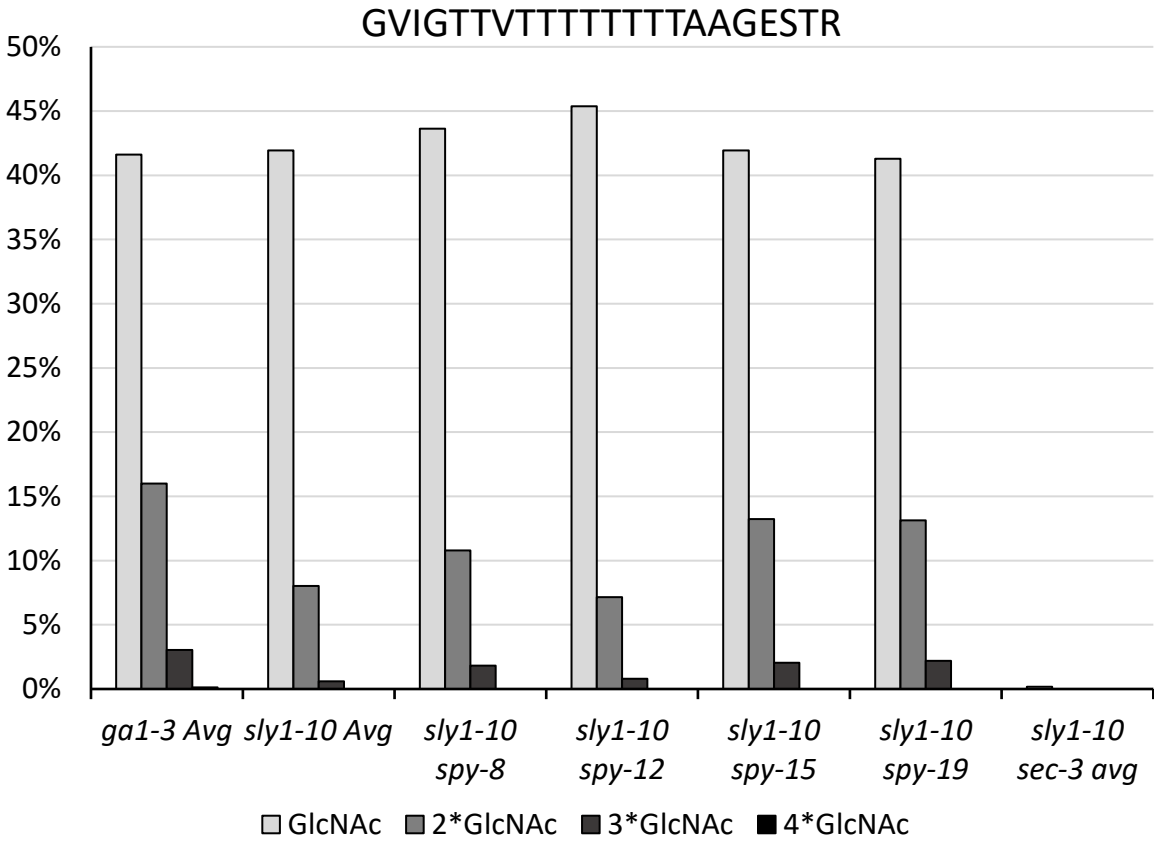


**Figure 2.7 GVIGTTVTTTTTTTTAAGESTR O-GlcNAcylation.** percent abundances are taken from Table 2.2. "Avg" denotes a value averaged over several samples (3 for *ga1-3* and *sly1-10*; 2 for *sly1-10 sec-3*).

The first *sly1-10* and *ga1-3* samples (5/17 *sly1-10* and 7/17 *ga1-3*) were analyzed with a 300-1200 m/z MS1 window. However, the subsequent samples (12/17 and 3/18 *sly1-10*, 8/18 *ga1-3*, *sly1-10 spy-8/-12/-15/-19*, and both *sly1-10 sec-3*) were analyzed with a 300-1500 m/z window. The m/z range was expanded because the $(M+2H)^{2+}$ charge state of some larger peptides (such as more heavily-modified ones) fell outside of that range. This affects accurate quantification of PTMs relative to the unmodified form; for example, unmodified and mono-GlcNAcylated GVIGTTVTTTTTTTTAAGESTR have +2 charge states below 1200 m/z, but the di- and tri-GlcNAcylated forms have +2 charge states above 1200 m/z (Table 2.4). An $(M+2H)^{2+}$

peptide at 1200 m/z has an $(M+H)^{1+}$ mass of 2399 m/z. The (LK)SCSSPDSMVTSTSTGTQIGK peptide also has some forms with +2 charge states above 1200 m/z. On the other hand, the LSNHGTSSSSSSISK(DK) and VIPGNAIYQFPAIDSSSSSNNQNK(R) peptides were not affected by the change, as the unmodified and modified species considered always fall below or above 2399 Da, respectively (GlcNAcylated phosphorylated LSN17: 2004 m/z. Unmodified VIP24: 2551 m/z). Table 2.4 lists the $(M+H)^{1+}$ charge state m/z values for SCSS and GVIG peptides and whether they fall above or below 2400 Da. The mass-to-charge value for a specific PTM form does not change (mono-phosphorylated SCSSPDSMVTSTSTGTQIGK always has an $(M+2H)^{2+}$ of 1067.4095 m/z), though different PTM forms have different charge state distributions (percent in the +2 vs. +3 vs. +4, etc. charge states).

| **Peptide** | **$(M+H)^+$ Mass** | **Range** |
|---|---|---|
| SCSSPDSMVTSTSTGTQIGK | 2030.9001 | below 2400 |
| SCSSPDSMVTSTSTGTQIGK + PO4 | 2110.8664 | |
| SCSSPDSMVTSTSTGTQIGK + Fucose | 2176.9580 | |
| SCSSPDSMVTSTSTGTQIGK + 2*PO4 | 2190.8327 | |
| SCSSPDSMVTSTSTGTQIGK + GlcNAc | 2233.9794 | |
| LKSCSSPDSMVTSTSTGTQIGK | 2272.0791 | |
| LKSCSSPDSMVTSTSTGTQIGK + PO4 | 2352.0454 | |
| LKSCSSPDSMVTSTSTGTQIGK + Fucose | 2418.1370 | above 2400 |
| LKSCSSPDSMVTSTSTGTQIGK + 2*PO4 | 2432.0117 | |
| LKSCSSPDSMVTSTSTGTQIGK + GlcNAc | 2475.1584 | |
| **Peptide** | **$(M+H)^+$ Mass** | **Range** |
| GVIGTTVTTTTTTTTAAGESTR | 2127.0776 | below 2400 |
| GVIGTTVTTTTTTTTAAGESTR + PO4 | 2207.0439 | |
| GVIGTTVTTTTTTTTAAGESTR + Fucose | 2273.1355 | |
| GVIGTTVTTTTTTTTAAGESTR + Hexose | 2289.1304 | |
| GVIGTTVTTTTTTTTAAGESTR + GlcNAc | 2330.1569 | |
| GVIGTTVTTTTTTTTAAGESTR + 2*Hexose | 2451.1832 | above 2400 |
| GVIGTTVTTTTTTTTAAGESTR + Fucose + GlcNAc | 2476.2148 | |
| GVIGTTVTTTTTTTTAAGESTR + Hexose + GlcNAc | 2492.2097 | |
| GVIGTTVTTTTTTTTAAGESTR + 2*GlcNAc | 2533.2362 | |
| GVIGTTVTTTTTTTTAAGESTR + 2*Hexose + GlcNAc | 2654.2625 | |
| GVIGTTVTTTTTTTTAAGESTR + 2*GlcNAc + Hexose | 2695.2890 | |
| GVIGTTVTTTTTTTTAAGESTR + 3*GlcNAc | 2736.3155 | |
| GVIGTTVTTTTTTTTAAGESTR + 4*GlcNAc | 2939.3948 | |

**Table 2.4 m/z of singly-charged modified SCSS and GVIG peptides.** In a given peptide, different PTM species are listed in order of increasing $(M+H)^+$ mass.

(LK)SCSSPDSMVTSTSTGTQIGK was not significantly affected by the change in MS1 window. The three forms with an $(M+H)^+$ above 2400 m/z (LKSCSSPDSMVTSTSTGTQIGK + fucose, GlcNAc, or 2*PO4) only had a small fraction of the total mass area (<3%) present at the +2 charge state. GVIGTTVTTTTTTTTAAGESTR was affected to a greater extent. Depending on the PTM(s) present and their charge state distribution in each sample, the percent in the 2+ charge state ranged from a few percent to over half of the ion current. For a given PTM form, the percent in the 2+ charge state also varied across different samples. If one were to average the percent of each PTM form that was not $(M+2H)^{2+}$ across the 300-1500 m/z *ga1-3* and *sly1-10* samples and adjust the 300-1200 m/z samples accordingly, the difference in the averages (adjusted vs. unadjusted) would be less than 8%, except for GVIG + fucose + GlcNAc in *ga1-3* (15%). The different percent abundance between *ga1-3* and *sly1-10* samples was not notably altered. However, these charge state considerations combined with the long chromatographic tail of higher abundance PTM forms means that the threshold for a significant change between two samples is higher. *Ga1-3* had five times the amount of tri-GlcNAcylated GVIG compared to *sly1-10* and was the only mutant with detectable quad-GlcNAcylation. *Sly1-10 spy-8/-15/-19* all had at least three-fold more tri-GlcNAc than *sly1-10*. *Ga1-3* and *sly1-10 spy-19* also had four or more times as much di-hexosylated mono-GlcNAcylated GVIG as well as mono-hexosylated di-GlcNAcylated GVIG relative to *sly1-10*. Fucosylated GlcNAcylated GVIG was only observed in *ga1-3* and *sly1-10*, at similar levels.

### *2.4.4 Discovery of Acetylation*

When analyzing the data for the 7/17 *ga1-3* sample, I discovered a triply-charged peptide with a mass of 14.0034 m/z more than triply-charged LSNHGTSSSSSSISKDK (m/z

588.6176 vs. m/z 574.6142) eluting 2.6 minutes after the unmodified peptide (Figure 2.8a-b). This difference corresponds to 42.0102 Da in the +1 charge state. 42.0102 Da is very close to the mass of an added acetyl group (+$COCH_2$ = +42.0106 Da), a PTM found on lysines. Trimethylation (+$C_3H_6$ = +42.0470 Da) is another lysine modification of similar mass to acetylation; the difference is merely 0.0364 Da (+0.0121 m/z in the +3 charge state). With the high mass accuracy of the Orbitrap Fusion, however, this difference (0.0121/588.6184 = 21 ppm) is well above the mass error of -1.4 ppm for acetylated LSN17 (m/z 588.6176 observed vs. m/z 588.6184 calculated, -0.0008 m/z difference). The average mass error of peptide PTM variants in this sample run was -1.3 ppm. The charge state distribution of the modified peptide also pointed to acetylation; the modified LSN17 was almost entirely in the +3 charge state, whereas the unmodified LSN17 had a +4 charge state with half the abundance of the +3 charge state. Acetylating a lysine removes the positive charge on its side-chain nitrogen and lowers the average charge of the peptide, whereas trimethylation locks in the positive charge on the nitrogen with four bonded carbons, which would retain or increase the average charge of the peptide.

The ETD and CAD of the $(M+3H)^{3+}$ precursor ion confirmed the identity of the modified peptide as acetylated LSN17. The acetylation was on K15 of the peptide (K26 of wild-type RGA). The ETD MS2 spectrum is shown in Figure 2.8c with the observed ions marked on the spectrum and peptide sequence (c ions in blue, z· ions in green). Table 2.5 shows the potential ions in the ETD and CAD of the $(M+3H)^{3+}$ peptide in tabular form, with the observed ions indicated in grey. There was no CAD from the $(M+2H)^{2+}$ species because there was very little of it in the MS1; consequently, the CAD b- and y-ion coverage was not as extensive.

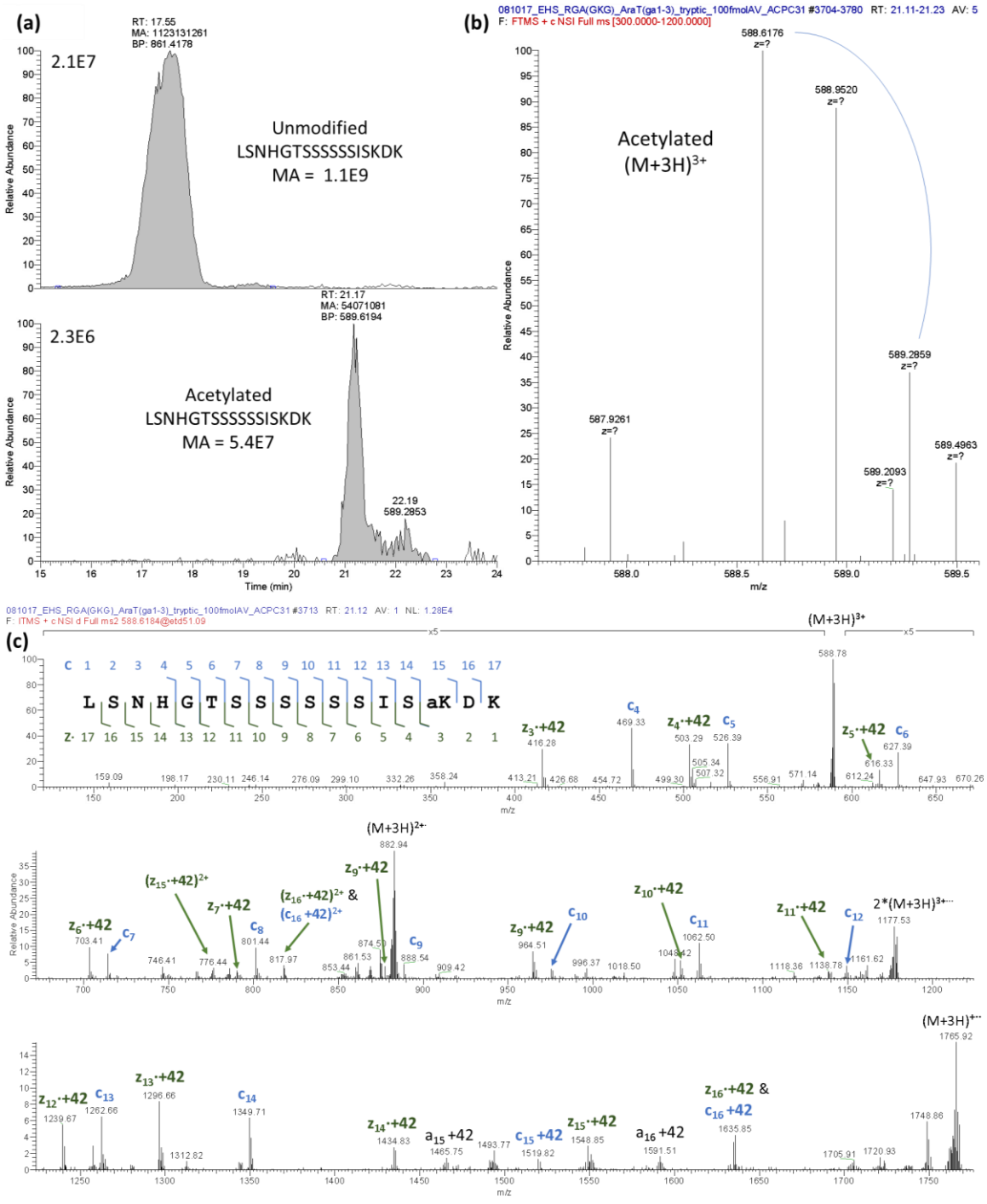


**Figure 2.8 Chromatograms and MS of acetylated LSNHGTSSSSSSISKDK from *ga1-3* (7/17).** (a) Chromatograms. The top and bottom chromatograms are LSNHGTSSSSSSISKDK unmodified and acetylated, respectively (the peaks include all major charge states and isotopes (10% cutoff)). RT – chromatographic retention time. MA – mass area of the peak (abundance of the peptide). The numbers to the left of each chromatogram are the intensity of the highest peak in each chromatogram. (b) The high-resolution MS1 of $(M+3H)^{3+}$, showing the isotopic distribution. (c) The ETD MS2 spectrum from $(M+3H)^{3+}$. c ions are shown in blue, and z· ions in green. Acetylated ions are indicated by "+42."

| 2+ b ions | 1+ b ions | 2+ c ions | 1+ c ions | | Sequence | | 1+ z· ions | 2+ z· ions | 1+ y ions | 2+ y ions |
|---|---|---|---|---|---|---|---|---|---|---|
| 57.59 | 114.09 | 66.10 | 131.12 | 1 | L | 17 | 1763.84 | 882.93 | 1763.84 | 882.93 |
| 101.13 | 201.12 | 109.64 | 218.15 | 2 | S | 16 | 1634.74 | 818.34 | 1650.76 | 826.35 |
| 158.18 | 315.17 | 166.69 | 332.19 | 3 | N | 15 | 1547.71 | 774.80 | 1563.72 | 782.81 |
| 226.75 | 452.23 | 235.26 | 469.25 | 4 | H | 14 | 1433.66 | 717.74 | 1449.68 | 725.76 |
| 255.28 | 509.25 | 263.79 | 526.27 | 5 | G | 13 | 1296.60 | 649.17 | 1312.62 | 657.19 |
| 305.83 | 610.29 | 314.34 | 627.32 | 6 | T | 12 | 1239.58 | 620.65 | 1255.60 | 628.66 |
| 349.37 | 697.33 | 357.88 | 714.35 | 7 | S | 11 | 1138.53 | 570.10 | 1154.55 | 578.11 |
| 392.91 | 784.36 | 401.42 | 801.39 | 8 | S | 10 | 1051.50 | 526.56 | 1067.52 | 534.57 |
| 436.45 | 871.39 | 444.96 | 888.42 | 9 | S | 9 | 964.47 | 483.02 | 980.49 | 491.03 |
| 479.98 | 958.42 | 488.50 | 975.45 | 10 | S | 8 | 877.44 | 439.48 | 893.46 | 447.49 |
| 523.52 | 1045.45 | 532.04 | 1062.48 | 11 | S | 7 | 790.41 | 395.94 | 806.43 | 403.95 |
| 567.06 | 1132.49 | 575.58 | 1149.51 | 12 | S | 6 | 703.37 | 352.40 | 719.39 | 360.41 |
| 623.64 | 1245.57 | 632.16 | 1262.60 | 13 | I | 5 | 616.34 | 308.86 | 632.36 | 316.87 |
| 667.18 | 1332.60 | 675.70 | 1349.63 | 14 | S | 4 | 503.26 | 252.28 | 519.28 | 260.29 |
| 752.29 | 1502.71 | 760.80 | 1519.73 | 15 | **K + 42** | 3 | 416.23 | 208.74 | 432.25 | 216.75 |
| 809.83 | 1617.74 | 818.35 | 1634.76 | 16 | D | 2 | 246.12 | 123.64 | 262.14 | 131.65 |
| 882.93 | 1763.84 | 882.93 | 1763.84 | 17 | K | 1 | 131.09 | 66.09 | 147.11 | 74.10 |

**Table 2.5 Sequence coverage of acetylated LSNHGTSSSSSSISKDK from *ga1-3* (7/17).** Grey cells indicate observed ions. The columns flanking the "Sequence" column give the number of residues in the b/c or y/z· ion (these numbers do not correspond to the residue number in the protein itself). c and z· ions come from the ETD of $(M+3H)^{3+}$; b and y ions come from the CAD of $(M+3H)^{3+}$, as there was no CAD of $(M+2H)^{2+}$ in this sample. The acetylated lysine is indicated as "K + 42."

CAD functions best at lower charge densities, as too many charges in a small peptide limits proton mobility. The problem was less marked for acetylated LSN17 than unmodified LSN17 (which had less complete b and y ion coverage), since the K15 side-chain nitrogen was neutralized as an amide that had lower basicity than an amine. CAD causes more random cleavages (and hence better sequence coverage) when there are more mobile protons than highly basic sites (lysine, arginine, and the N-terminus) (38). Furthermore, N-acetylation is a stable PTM that is not lost in CAD like phosphorylation or O-GlcNAcylation (39).

*Ga1-3* (7/17) also had acetylated VIPGNAIYQFPAIDSSSSSNNQNKR peptide, at K24 (K163 in wild-type RGA). No acetylated LKSCSSPDSMVTSTSTGTQIGK was observed in any sample. Trypsin does not cleave at acetylated lysines because the bulky, uncharged acetyl group on the nitrogen no longer interacts with the carboxyl group of the trypsin active site

aspartic acid (40–42). Because of this, no in vivo acetylated LSN15, VIP24, or SCSS20 was expected to be present.

### *2.4.5 Phosphorylation and Acetylation*

Figure 2.9 shows the levels of phosphorylation and acetylation across the seven mutants at LSNHGTSSSSSSISK(DK). Phosphorylation (mono-, di-, and tri-) of LSN was significantly increased in *ga1-3* relative to *sly1-10*. The percent of LSN bearing at least one phosphorylation increased eight-fold, and poly-phosphorylation increased fifteen-fold. Concordantly, acetylation increased 2.5-fold. In the *sly1-10* (*spy* or *sec*) double mutants, LSNHGTSSSSSSISK(DK) acetylation and phosphorylation changed in concert with one another relative to *sly1-10*. *Sly1-10 spy-8* had significantly increased di- and tri-phosphorylation and acetylation relative to *sly1-10*. *Sly1-10 spy-15* had the highest percentage of mono-phosphorylation of any mutant; its levels of di-phosphorylation and acetylation were elevated relative to *sly1-10* (but not quite significantly). *Sly1-10 spy-12* had significantly decreased di/tri-phosphorylation and acetylation relative to *sly1-10* (though a similar level of mono-phosphorylation). *Sly1-10 spy-19* had both di-phosphorylation and acetylation reduced three-fold relative to *sly1-10* (though the level of mono-phosphorylation was similar). *Sly1-10 sec-3* had increased mono-/di-/tri-phosphorylation and acetylation relative to *sly1-10*. *Ga1-3* and *Sly1-10 spy-15* had the highest percent of LSN phosphorylated at least once (36% and 22%, respectively), while *sly1-10 spy-12* and *sly1-10 spy-19* had the lowest (2.3% and 3.2%, respectively).

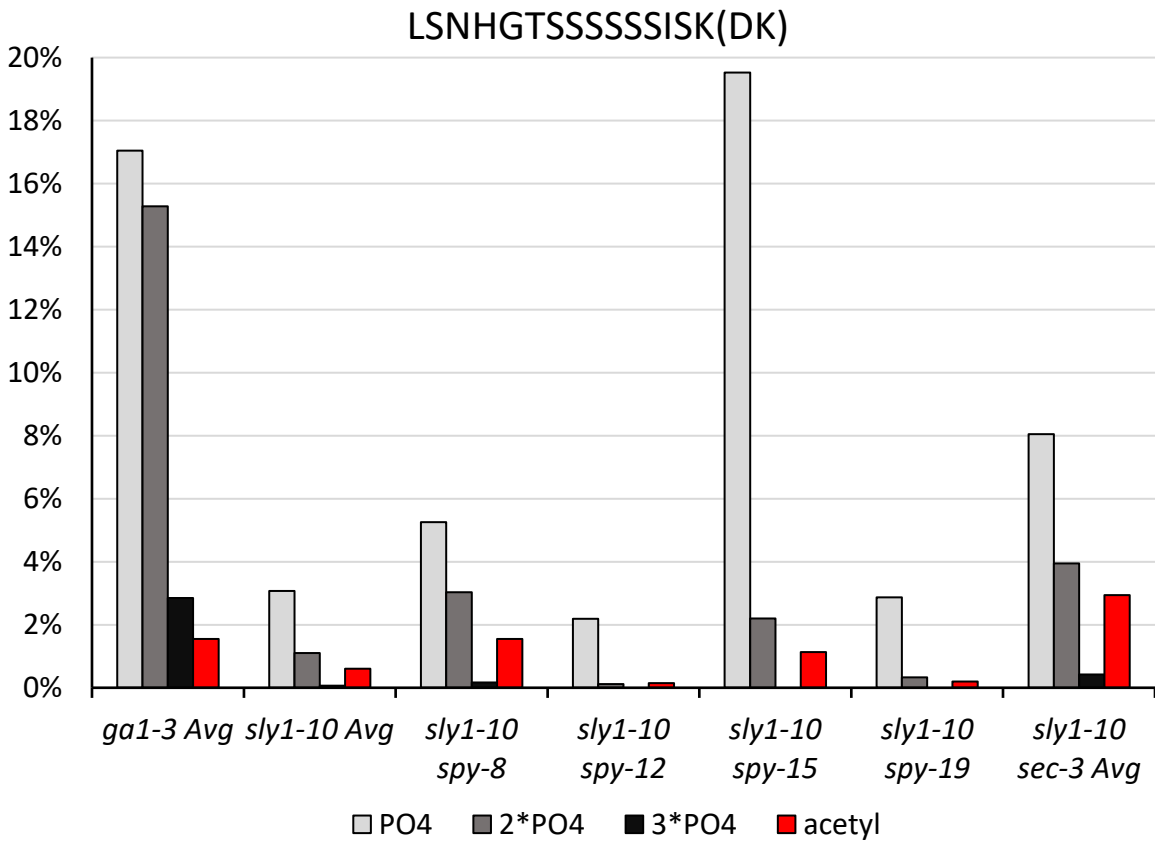


**Figure 2.9 Phosphorylation and acetylation of LSNHGTSSSSSSISK(DK) in RGA.** The percent abundances are taken from Table 2.2. "Avg" denotes a value averaged over several samples (3 for *ga1-3* and *sly1-10*; 2 for *sly1-10 sec-3*).

LSN was also found modified with simultaneous GlcNAcylation and phosphorylation. Relative to *sly1-10*, GlcNAcylated phosphorylated LSN was increased significantly in *ga1-3* and *sly1-10 spy-8/-15*, while it was of a similar level in *sly1-10 spy-12* and not detected in *sly1-10 spy-19* or *sly1-10 sec-3*. Interestingly, *sly1-10 spy-12/-19* had significantly increased GlcNAcylated LSN levels.

Acetylation on VIPGNAIYQFPAIDSSSSSNNQNK(R) was less than 0.1% of all VIP, except in *ga1-3* where it was 0.3%, so it is difficult to draw any conclusions on *sly1-10 (spy or sec)* double mutants relative to *sly1-10*. Phosphorylation was also most abundant in *ga1-3* with significantly more phosphorylation than the other samples, though *sly1-10 spy-12* had an

unexpectedly high proportion of di-phosphorylation. Several of the samples lacked MS2s for mono-phosphorylated and acetylated VIP (and all but *ga1-3* lacked them for di-phosphorylation).

Mono-phosphorylation levels of (LK)SCSSPDSMVTSTSTGTQIGK were high (on average 22%) but did not change significantly across samples. On the other hand, di-phosphorylation was significantly elevated in *ga1-3* relative to *sly1-10*, and absent in the double mutants. GVIGTTVTTTTTTTTAAGESTR phosphorylation was not detected in any sample.

### *2.4.6 Post-Translational Modification Sites*

Table 2.6 lists the locations of PTMs in the 5/17 *sly1-10* sample. Table 2.7 lists the locations of PTMs in the 7/17 *ga1-3* sample. In Tables 2.6 and 2.7, the first column gives the peptide with its modified amino acid(s) denoted in lower case with the modification as a superscript. For example, LSNHGTs$^{GlcNAc}$SSSSSISK was O-GlcNAcylated at the first S following the T (S18 in wild-type RGA). The second column gives the amino acid number in the wild-type RGA sequence, preceded by a single letter denoting the modification: f is O-fucosylation, g is O-GlcNAcylation, h is O-hexosylation, p is phosphorylation, and a is acetylation. Thus, LSNHGTs$^{GlcNAc}$SSSSSISK corresponds to gS18. In cases where the modified amino acid was uncertain, the possible amino acids are surrounded by brackets with the PTM as a superscript at the end (keeping in mind that O-fucosylation, O-GlcNAcylation, O-hexosylation, and phosphorylation only occur on S and T, and acetylation on K, in these peptides). Multiply modified peptides are also indicated; for example, GVIGTTV(TTT)$^{2*GlcNAc}$TTTTTAAGESTR has two GlcNAcs among T193, T194, and T195, indicated as gg(T193-T195) in the right column.

| ***sly1-10* (5/17) Modification Sites** | **Modification** |
|---|---|
| LSNHGt$^{Fucose}$SSSSSSISK | fT17 |
| LSNHGTSSs$^{Fucose}$SSSISK | fS20 |
| LSNHGT(SS)$^{Fucose}$SSSSISK (trace) | f(S18-S19) |
| LSNHGT(SSS)$^{Fucose}$SSSISKDK | f(S18-S20) |
| LSNHGt$^{GlcNAc}$SSSSSSISK | gT17 |
| LSNHGTs$^{GlcNAc}$SSSSSISK | gS18 |
| LSNHGTs$^{GlcNAc}$SSSSSISKDK | gS18 |
| LSNHGTSs$^{GlcNAc}$SSSSISK | gS19 |
| LSNHGTSs$^{GlcNAc}$SSSSISKDK | gS19 |
| LSNHGTs$^{Hexose}$SSSSSISK | hS18 |
| LSNHG(TSSS)$^{Hexose}$SSSISK | h(T17-S20) |
| LSNHGTSs$^{PO4}$SSSSISK | pS19 |
| L(SNHGTSSSSSS)$^{PO4}$ISK | p(S13-S23) |
| LSNHGTSSs$^{PO4}$SSSISKDK | pS20 |
| LSNHGT(SSSSSS)$^{PO4}$ISKDK | p(S18-S23) |
| LSNHGTSs$^{PO4}$SSSs$^{PO4}$ISKDK | pS19 + pS23 |
| LSNHGTSSs$^{PO4}$SSs$^{PO4}$ISKDK | pS20 + pS23 |
| LSNHGTSSSSSSISk$^{Acetyl}$DK | aK26 |
| LK(SCSSPDSMVTSTSTGT)$^{GlcNAc}$QIGK | g(S167-T182) |
| (SCSSPDSMVTSTSTGT)$^{GlcNAc}$QIGK | g(S167-T182) |
| LKSCSs$^{PO4}$PDSMVTSTSTGTQIGK | pS170 |
| SCSs$^{PO4}$PDSMVTSTSTGTQIGK | pS170 |
| GVIGTTVt$^{Hexose}$TTTTTTTAAGESTR | hT193 |
| GVIGTTVTt$^{Hexose}$TTTTTTAAGESTR | hT194 |
| GVIG(TTVTTTTTTTTAAGEST)$^{2*Hexose}$R | hh(T190-T206) |
| GVIGTTV(TTTTTT)$^{GlcNAc}$TTAAGESTR | g(T193-T198) |
| GVIGTTV(TTT)$^{2*GlcNAc}$TTTTTAAGESTR | gg(T193-T195) |
| GVIG(TTVTTT)$^{2*GlcNAc}$TTTTTAAGESTR | gg(T190-T195) |
| GVIG(TTVTTTTTTTTAAGEST)$^{3*GlcNAc}$R | ggg(T190-T206) |
| GVIG(TTVTTTTT)$^{Hexose+GlcNAc}$TTTAAGESTR | gh(T190-T197) |
| GVIG(TTVTTTTTTTTAAGEST)$^{Hexose+2*GlcNAc}$R | ggh(T190-T206) |
| GVIG(TTVTTTTTTTTAAGEST)$^{Fucose+GlcNAc}$R | fg(T190-T206) |

**Table 2.6 Modification sites in the 5/17 *sly1-10* sample.** Parentheses indicate that the modification(s) occur on a serine or threonine within the bracketed range. The second column gives the residue number(s) of wild-type RGA where the modification is found. Fucose, f: O-fucosylation. GlcNAc, g: O-GlcNAcylation. Hexose, h: O-hexosylation. PO4, p: phosphorylation. Acetyl, a: acetylation.

| ***ga1-3* (7/17) Modification Sites** | **Modification** |
|---|---|
| LSNHGTs$^{GlcNAc}$SSSSSISK | gS18 |
| LSNHGTS(SSSSS)$^{GlcNAc}$ISKDK | g(S19-S23) |
| L(SNHGTSSSSSS)$^{PO4}$ISKDK | p(S13-S23) |
| L(SNHGTSSSSSSIS)$^{2*PO4}$KDK | pp(S13-S25) |
| LSNHGT(SSSSSSIS)$^{3*PO4}$KDK | ppp(S18-S25) |
| LSNHGTSSSSSSISk$^{acetyl}$DK | aK26 |
| VIPGNAIYQFPAIDS(SSSS)$^{PO4}$NNQNKR | p(S155-S158) |
| VIPGNAIYQFPAID(SSSSS)$^{2*PO4}$NNQNKR | pp(S154-S158) |
| VIPGNAIYQFPAIDSSSSSNNQNk$^{Acetyl}$R | aK163 |
| LKSC(SSPDS)$^{GlcNAc}$MVTSTSTGTQIGK | g(S169-S173) |
| SCSSPDSMVt$^{GlcNAc}$STSTGTQIGK | gT176 |
| LKSCSs$^{PO4}$PDSMVTSTSTGTQIGK | pS170 |
| SCSs$^{PO4}$PDSMVTSTSTGTQIGK | pS170 |
| LK(SCSSPDSMVTSTSTGT)$^{2*PO4}$QIGK | pp(S167-T182) |
| GVIGTTV(TTTTTT)$^{GlcNAc}$TTAAGESTR | g(T193-T198) |
| GVIGTTV(TTTT)$^{2*GlcNAc}$TTTTAAGESTR | gg(T193-T196) |
| GVIG(TTVTTTT)$^{2*GlcNAc}$TTTTAAGESTR | gg(T190-T196) |
| GVIGT$^{GlcNAc}$TV(TTTTTTTTAAGEST)$^{2*GlcNAc}$R | gT190 + gg(T193-T206) |
| GVIGt$^{GlcNAc}$TV(TT)$^{GlcNAc}$(TTT)$^{GlcNAc}$t$^{GlcNAc}$TTAAGESTR | gT190 + g(T193-T194) + g(T195-T197) + gT198 |
| GVIGTTV(TT)$^{Hexose}$TTTTTTAAGESTR | h(T193-T194) |
| GVIG(TTVTTTTTTTTAAGEST)$^{2*Hexose}$R | hh(T190-T206) |
| GVIG(TTVTTTTTTTTAAGEST)$^{Hexose+GlcNAc}$R | gh(T190-T206) |
| GVIG(TTVTTTTTTTTAAGEST)$^{2*Hexose+GlcNAc}$R | ghh(T190-T206) |
| GVIG(TTVTTTTTTTT)$^{Hexose+2*GlcNAc}$AAGESTR | ggh(T190-T200) |
| GVIG(TTVTTTTTTTTAAGEST)$^{Fucose+GlcNAc}$R | fg(T190-T206) |

**Table 2.7 Modification sites in the 7/17 *ga1-3* sample.** Parentheses indicate that the modification(s) occur on a serine or threonine within the bracketed range. The second column gives the residue number(s) of wild-type RGA where the modification is found. Fucose, f: O-fucosylation. GlcNAc, g: O-GlcNAcylation. Hexose, h: O-hexosylation. PO4, p: phosphorylation. Acetyl, a: acetylation.

Figure 2.10 shows the PTM sites from Tables 2.6 and 2.7 on the sequence of RGA, in addition to some other sites from other samples. Residues where a PTM has been specifically site-localized are marked with the symbol of the PTM above the residue. Stretches where a

PTM exists but has not been site-localized are indicated by lines beneath the corresponding residues in the color of the modification. These underlines overlap site-localized residues. The results are drawn from multiple samples, and within a sample there may be multiple chromatographic peaks of a PTM form (such as seen in Figures 2.4 and 2.5 with O-GlcNAcylated LSNHGTSSSSSSISK). Also, three of the four peptides have two different forms due to a missed tryptic cleavage (such as LSNHGTSSSSSSISK and LSNHGTSSSSSSISKDK).

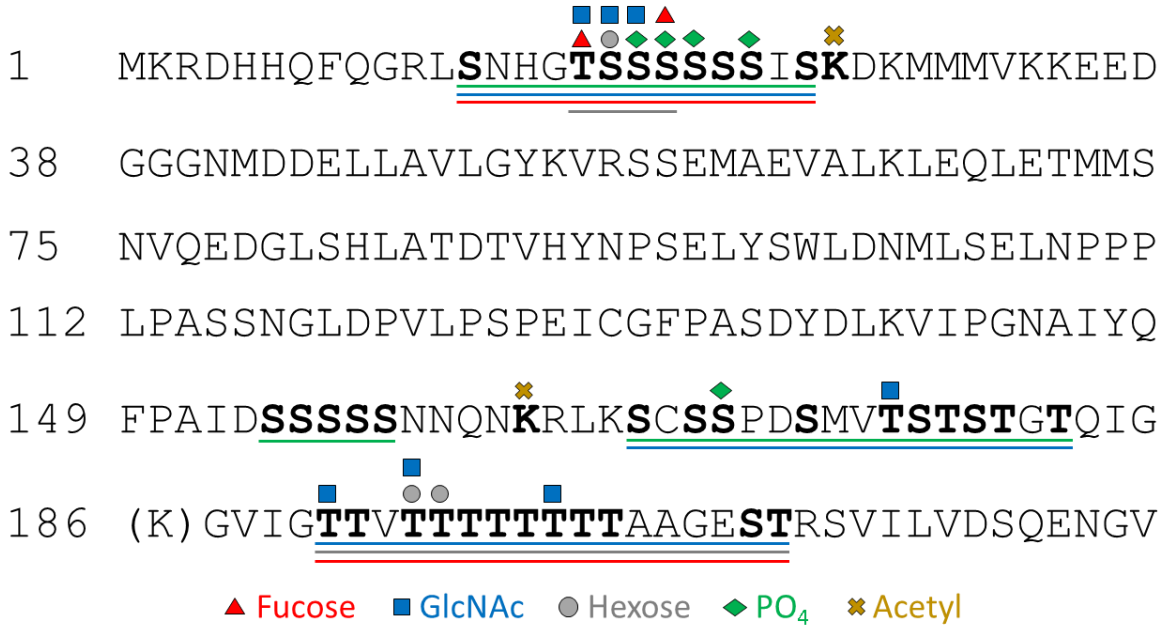


**Figure 2.10 Post-translational modification sites in RGA.** This figure shows the first 219 amino acids of RGA, where all the PTM sites in this work were found. Bolded letters indicate modified or potentially modified residues. For definitively localized PTMs, fucosylation sites are indicated with red triangles, GlcNAcylation sites with blue squares, hexosylation sites with grey circles, phosphorylation sites with green diamonds, and acetylation sites with gold crosses above the specific residue. Underlined residues indicate stretches where the modification (of the corresponding color) occurs, but the specific site was not able to be determined (these often overlap localized sites). The parenthesized K is the inserted lysine of $RGA^{GKG}$.

LSNHGTSSSSSSISK is the most diversely modified stretch, with all four modifications seen. In particular, LSNHG**TSSS**SSSISK (T17 – S20) has two modifications site-localized to each residue. Certain phosphorylation and acetylation sites occur very close together. They are as close as phosphorylated S23 and acetylated K26, though no LSNHGTSSSSSSISKDK peptide with

simultaneous phosphorylation and acetylation was detected in any sample. Acetylated K163 is also seven residues before phosphorylated S170 and five to nine residues after the phosphorylated poly-S stretch of S154-S158. Again, there was no VIPGNAIYQFPAIDSSSSSNNQNKR with simultaneous phosphorylation and acetylation detected.

## 2.5 Discussion

### *2.5.1 spy vs. sec Mutants*

GlcNAcylation and fucosylation were altered in the *sly1-10 spy* and *sly1-10 sec* mutants. *Sec-3* is a null mutant with no SEC expression (thus no OGT activity) (18), whereas *spy-8/-12/-15/-19* are mutants with little or no POFUT activity (3,19). Zentella et al. 2017 demonstrated that fucosylation and GlcNAcylation are competitive modifications on RGA in at least the LSN peptide, so reducing the activity of one enzyme was predicted to increase the presence of the other modification in *Arabidopsis* (3). However, this previous work could not characterize these modifications on the SCSS and GVIG peptides in the native organism of RGA, *Arabidopsis thaliana*, separated from the influence of any tobacco OGTs or POFUTs. In fact, while most of the PTM sites observed here were previously seen on RGA in tobacco (3,18,20), the only modifications seen in *Arabidopsis* were on the LSN peptide at T17-S20 (fucosylation, GlcNAcylation, and phosphorylation), and fucosylation on the VIP peptide (3).

As expected, the *sly1-10 sec-3* double mutant had no GlcNAcylation of LSNHGTSSSSSSISK(DK), (LK)SCSSPDSMVTSTSTGTQIGK, and only <0.2% mono-GlcNAcylated GVIGTTVTTTTTTTTAAGESTR. In contrast, *sly1-10* had several percent of LSN and SCSS in a

GlcNAcylated form, and over half of GVIG in a GlcNAcylated form. Concordantly, LSN fucosylation more than doubled in *sly1-10 sec-3* vs. *sly1-10*, and GVIG fucosylation increased.

In the *sly1-10 spy* mutants, the relation between fucosylation and GlcNAcylation levels was weaker than in the *sly1-10 sec-3* mutant. LSN fucosylation was absent or decreased relative to *sly1-10* as expected, except in *sly1-10 spy-8* where it was actually increased. *Spy-8* is a less severe phenotype with more POFUT activity than *spy-12,-15, or -19* (3,19), but this does not explain the increase. LSN GlcNAcylation levels were similar or higher in the *sly1-10 spy* mutants than in *sly1-10*. SCSS GlcNAcylation was not significantly changed, and neither was GVIG GlcNAcylation (except for some poly-GlcNAcylation in some *sly1-10 spy* samples). Fucosylated GVIG was not detected in any *sly1-10 spy* mutant, while it was present in *sly1-10* at a low level. It appears that the *sec* mutation has a more dramatic influence than the *spy* mutations, possibly because *sec-3* has no protein expressed at all, whereas the *spy* mutants still have the protein present, just in a less active mutant form. Still, this work has confirmed that fucosylation and GlcNAcylation are opposed modifications on RGA in *Arabidopsis*.

### *2.5.2 RGA Acetylation: DELLA Protein Alignment*

Acetylation sites were discovered on the peptide LSNHGTSSSSSSIS(aK)DK (residue K26 of wild-type RGA) and VIPGNAIYQFPAIDSSSSSNNQN(aK)R (residue K163). Acetylation has been observed in *Arabidopsis* before (43–45), but not in RGA. More generally, to the best of the author's knowledge, lysine acetylation has only been observed on a DELLA protein once before. Cao et al. 2019 found an acetylated DELLA protein in *Paulownia tomentosa* (princess tree), QHQPQA(aK)K (46). This aligns with SNNQN(aK)R from VIP of RGA (Figure 2.11). In fact, the motif of K-(K/R)-x-(K/R) (such as VIPGNAIYQFPAIDSSSSSNNQN**KRLK**SCSSPDS in RGA),

where x is usually hydrophobic, is highly conserved in DELLA proteins in Angiosperms (flowering plants). This motif is often followed by a stretch rich in serines and threonines (and sometimes preceded by a stretch with a few serines or threonines). For example, HvSLN1 from barley, OsSLR1 from rice, TaRHT1 from wheat, and ZmD8 from maize all have a long poly-S/T stretch paralleling parts of the RGA LKSCSSPDSMVTSTSTGT stretch, while grape VvGAI and *Paulownia tomentosa* PtDELLA also have a stretch with several serines and threonines. HvSLN1 and TaRHT1 also have a serine aligned with S170 of RGA, which is a phosphorylation site in RGA. The LSNHGTSSSSSSIS(aK) lysine is less conserved – of the seven DELLA proteins in Figure 2.11, the lysine is only present in RGA, PtDELLA, ZmD8, OsSLR1, and VvGAI. The serines and threonines of LSNHGTSSSSSSISK have scattered equivalents in the other proteins.

```
RGA      MKRDHHQFQGRLSNHGTSSSSSSISKDKMMMVKKEEDG-GGNMDDELLAVLGYKVRSSEM
PtDELLA MKRDHMNQEN-FSSSGSC---SSSSKAKMWSCETDDAG-----VDELFAVLGYNVKSSDM
HvSLN1   MKREYQDGGG--SGGGGD--EMGSSRDKMMVSSSEAG--EGEEVDELLAALGYKVRASDM
TaRHT1   MKREYQDAGG--SGGGGG--GMGSSEDKMMVSAA-AG--EGEEVDELLAALGYKVRASDM
ZmD8     MKREYQDAGG--SGG-----DMGSSKDKMMAAAAGAGEQEEEDVDELLAALGYKVRSSDM
OsSLR1   MKREYQEAGG--SSGGGSSADMGSCKDKVMAGAAG----EEEDVDELLAALGYKVRSSDM
VvGAI    MKREYHH------PHHPTCSTSPTGKGKMWDADPQQDA-GM---DELLAVLGYNVKASDM

RGA      VIPGNAIY--------QFPAIDSSSSSNNQNKRLKSCSSPDSMVTSTSTGTQIGGVIGTT
PtDELLA AIPGKAVYPRTQPPPPPPPPKQ----HQPQAKKLKTSPY-------------VGNSVGGL
HvSLN1   PIISPP--V----APADLSA-D----SVRDPKRMRTGGS-----STSSSSSSSSSLGGGA
TaRHT1   PIPSPAGAT----APADLSA-D----SVRDPKRMRTGGS-----STSSSSSSSSSLGGGA
ZmD8     PIPSPV-AA----PSADPST-D----SAREPKRMRTGGG-----STSSSSSSSSSMDGGR
OsSLR1   PISLPV-VA-----TADPSAAD----SARDTKRMRTGGG-----STSSSSSSSSSLGGGA
VvGAI    AIPGKALYSHIEQPPQQPPA---PPLYQRDNKRLKP--------TTSATANSVSSVIGGW
```

**Figure 2.11: Alignment of DELLA proteins.** Proteins from *Arabidopsis thaliana* (RGA), *Paulownia tomentosa* (PtDELLA), barley (HvSLN1), wheat (TaRHT1), maize (ZmD8), rice (OsSLR1), and grape (VvGAI). The two stretches are not contiguous (RGA 1-59, and 140-191). The K-(K/R)-x-(K/R) motif is colored purple. The RGA acetylated lysines (K26 and K170), and the aligned K in the other proteins, are highlighted in yellow. The S and T residues aligning to those in RGA are indicated in green. Alignment performed using COBALT from https://www.ncbi.nlm.nih.gov/tools/cobalt/ (33). PtDELLA sequence assembled from (47) using BLAST (33) and ExPASy (48).

Besides the acetylation found in Cao 2019, a search of the literature found only one other proposed DELLA acetylation site, in AtRGL3 in *Arabidopsis* (in Uhrig et al. 2019). However, the one AtRGL3 peptide in their dataset had the acetylated lysine at the C-terminus of a tryptic peptide (the only lysine in the peptide) (49), when trypsin cannot cleave at acetylated lysines, since the bulky, uncharged head group is not stabilized by the aspartic acid in the trypsin active site (40–42). C-terminally acetylated peptides could be a search program misassignment or an experimental artifact, as demonstrated by Choudhary et al. with deuterated acetic acid (42).

### *2.5.3 Phosphorylation and Acetylation Stabilize RGA*

Phosphorylation and acetylation levels were increased in the gibberellin-deficient *ga1-3* mutant relative to the *sly1-10* mutant, especially poly-phosphorylation. K163 acetylation was increased 20-fold in *ga1-3* over *sly1-10*. SCSS mono-phosphorylation was not increased, but VIP mono-/di-phosphorylation were increased, and SCSS di-phosphorylation was increased over 100-fold. This increase, and the linkage of phosphorylation and acetylation levels across the different mutants, sheds new light on the mechanism of gibberellin-mediated DELLA degradation. Figure 2.12a shows an updated model. It is known that the *sly1-10* mutant does not ubiquitylate RGA, since the F-box protein *sly1* is unable to bind to RGA and recruit the SCF E3 ubiquitin ligase complex to it. On the other hand, the *ga1-3* mutant does not produce gibberellin to cause GID1 to bind to RGA and recruit SLY1. Both mutants lack proteasomal degradation of RGA; the main difference is that in *sly1-10* GID1 can bind to RGA, whereas in *ga1-3* it cannot (16).

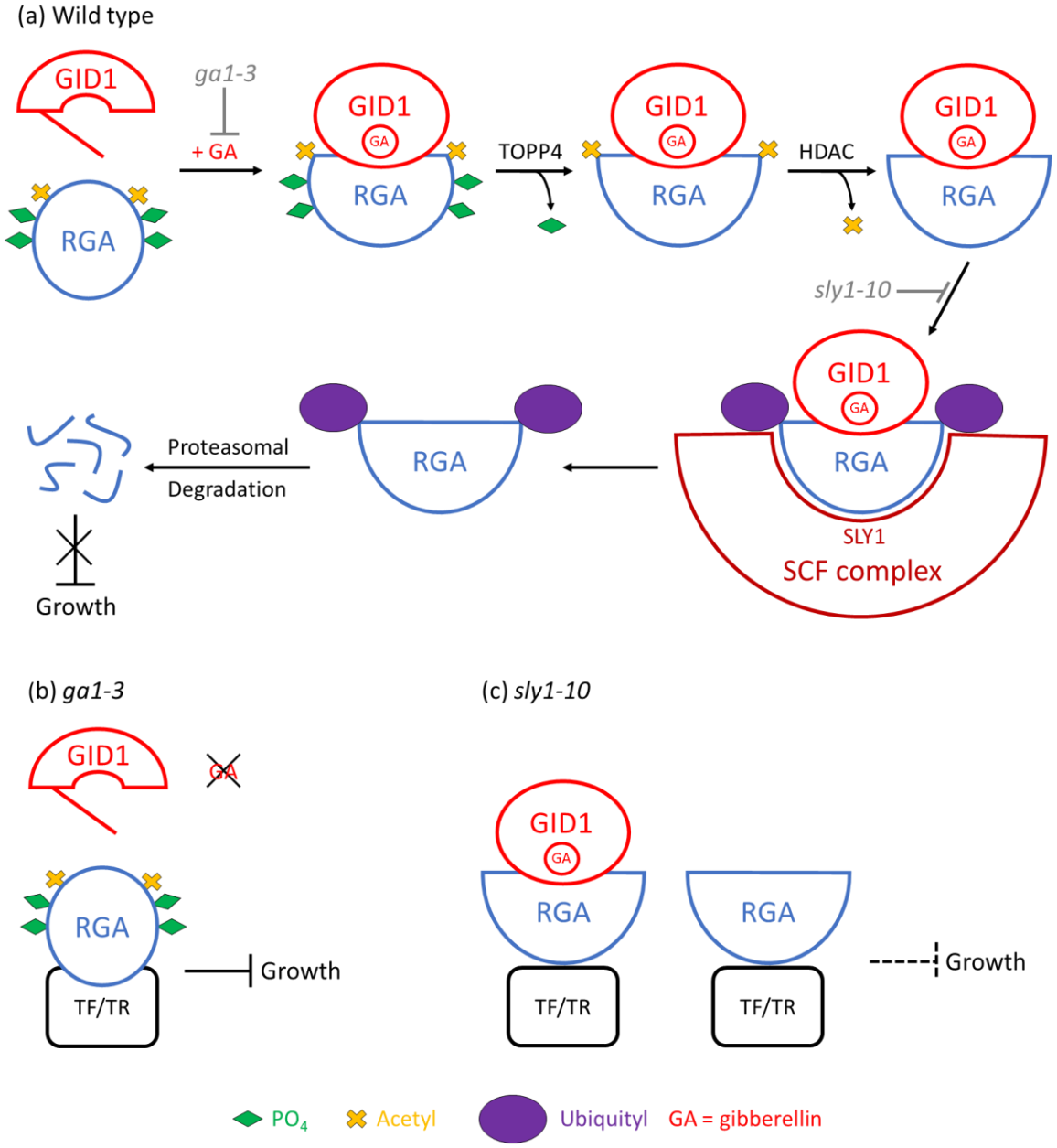


**Figure 2.12 Model of phosphorylation and acetylation regulation of RGA.** a) Wild type *Arabidopsis*. Gibberellin (GA) binds to GID1, then GA-GID1 binds to RGA. This changes the conformation of RGA to allow TOPP4 and a deacetylase (HDAC) to remove phosphates and acetyls, respectively, from RGA. SLY1 targets the SCF E3 ubiquitin ligase to RGA to ubiquitylate it at the deacetylated lysines, causing proteasomal degradation of RGA. The steps where *ga1-3* and *sly1-10* would interfere are shown. A blunt-end arrow indicates inhibition or repression. (b) *ga1-3* has no gibberellin, no GID1 binding to RGA, higher levels of RGA phosphorylation and acetylation, stronger binding to target transcription factors and regulators (TF/TR), and thus more repression of growth. (c) *sly1-10* has decreased phosphorylation and acetylation due to the presence of gibberellin and GID1 binding, which causes TOPP4 dephosphorylation and lysine deacetylation. RGA is not ubiquitylated by the SCF E3 ubiquitin ligase complex, but it binds to its target TFs/TRs less and thus has less repression of growth than in *ga1-3*. Green diamond: phosphorylation. Gold cross: acetylation. Purple oval: ubiquitylation.

Studies have shown that the phosphorylation of DELLA proteins enhances their stability (6,50). Early papers proposed that gibberellin-induced phosphorylation of DELLAs enhanced their degradation (51,52), since F-box proteins usually recognize and bind to the phosphorylated form of their substrates preferentially (53). However, Itoh et al. (54) demonstrated that phosphorylation of rice OsSLR1 does not affect the binding of the rice SLY1 equivalent, require gibberellin, or impair OsSLR1 stability. Dai and Xue (55) showed that the kinase EARLY-FLOWERING1 (EL1) phosphorylates OsSLR1 and actually enhances its stability. Hussain et al. (56) and Wang et al. (57) demonstrated that phosphatase inhibitors increase the stability of *Arabidopsis* AtRGL2 and RGA, respectively. In accordance with this, Qin et al. (58) showed that dephosphorylation of RGA and *Arabidopsis* AtGAI by the phosphatase TOPP4 downregulates their repression of gibberellin signaling by enhancing their gibberellin-mediated degradation.

However, none of these studies used mass spectrometry to characterize the phosphorylation of DELLAs across mutants of different proteins in the degradation pathway. They looked for the presence or absence of phosphorylation on a DELLA protein with gel electrophoresis (all studies), Western blots with a phospho-specific antibody (55), phosphate radiolabeling (51,52,54), etc. Such techniques can give rough relative levels of phosphorylated vs. unphosphorylated DELLA protein. However, they were unable to confirm sites of phosphorylation, merely identify possible ones with techniques like deleting portions of the protein (54,55,57) or testing phospho-mimic amino acid substitution variants (55,56). They were also unable to differentiate between mono- and poly-phosphorylation, or detect new PTMs. Mass spectrometry overcomes these limitations, and in this work revealed the

difference in phosphorylation levels between *ga1-3* and *sly1-10*, as well as the existence of DELLA acetylation.

How does TOPP4 dephosphorylation of RGA cause its proteasomal degradation, and how does it relate to GID1 binding? Considering this data, GID1 binding likely precedes and enables TOPP4 dephosphorylation of RGA (see Figure 2.12a). *Topp4-1* is a mutant of TOPP4 with a single amino acid substitution that does not impair *topp4-1* binding to RGA but does prevent its dephosphorylation and degradation. *Topp4-1* plants are severely dwarfed, like *ga1-3* plants, but they are unresponsive to gibberellin treatment, unlike *ga1-3*. It is the increased phosphorylation that causes the pronounced RGA activity (58). However, the Qin 2014 data did not establish the order of GID1 binding and dephosphorylation of RGA. On the one hand, if GID1 binding were to follow the dephosphorylation, *topp4-1* would prevent GID1 binding. On the other hand, if GID1 binding were to precede the dephosphorylation, the lack of *topp4-1* activity would mean the level of GID1 binding was irrelevant. However, the data presented here allows one to distinguish the two possibilities, and supports the second model: *ga1-3* (no GID1 binding) has increased phosphorylation over *sly1-10*, even in the presence of functional TOPP4. If TOPP4 dephosphorylation occurred first, one would expect to see no difference in phosphorylation levels between *ga1-3* and *sly1-10*.

While the N-terminal portion of RGA is intrinsically disordered, it is known that GID1 binding causes the AtGAI equivalent of RGA residues 44-107 (with the DELLA, LExLE, and TVHYNP motifs) to fold into a bundle of alpha helices. This bundle lies between the N-terminus of RGA (with LSN peptide) and the poly-S/T region (with the other three peptides). The change in RGA conformation upon GID1 binding would affect the conformation of those flanking stretches (5,59), likely exposing the phosphorylated residues for TOPP4 access.

Phosphorylation and acetylation are able to affect each other. The N-terminal tail of histones is unstructured and the site of many PTMs, similar to RGA (59,60). Cheung et al. found that histone H3 can be phosphorylated on S10. This phosphorylation recruits the histone acetyltransferase (HAT) GCN5 to H3 to acetylate K14, which then increases expression of the genes in the DNA bound to the modified histones (61). Tran et al. found that the *Arabidopsis* deacetylase (HDAC) HDA14 can bind to a phosphatase PP2A-A2 subunit, which helps HDA14 bind to α-tubulin and deacetylate it (62). Since phosphorylation and acetylation levels corresponded to each other in this work, increasing or decreasing in concert across different mutants, TOPP4 dephosphorylation likely causes deacetylation of RGA. The removal of the phosphates would change the interaction surface electrostatics and shape, increasing the affinity of RGA for the histone deacetylase. TOPP4 and the HDAC may even form a complex to sequentially remove their respective PTMs.

The discovery of lysine acetylation on DELLA proteins is notable, because ubiquitylation also occurs on lysines. Plants display crosstalk between acetylation, ubiquitylation, and phosphorylation (63). Lin et al. demonstrated that three lysines of human ATP-citrate lyase are sites of both acetylation and ubiquitylation, and that acetylation blocks ubiquitylation on them to stabilize the protein (64). The lysines of RGA that are ubiquitylated have not been identified (6), but it is likely that they comprise or include K26 and K163. This would explain why phosphorylation of RGA prevents its degradation. Even though SLY1 can bind to phosphorylated RGA (as shown with the rice equivalents) (54), the SCF E3 ubiquitin ligase complex cannot ubiquitylate an acetylated lysine. This completes the model in Figure 2.12a, where gibberellin causes RGA degradation by binding to GID1, which then binds to RGA, changing its conformation. This causes TOPP4 to dephosphorylate RGA, which allows a

HDAC to deacetylate RGA, exposing the lysines for ubiquitylation by the SCF complex. The ubiquitylated RGA is then degraded by the proteasome, freeing the transcription factors and regulators that it had sequestered in order to promote the expression of growth-activating genes.

Furthermore, the dephosphorylation and deacetylation of RGA caused by GID1 binding likely decreases the activity of RGA as well (Figure 2.12b-c). *ga1-3*, *gid1* (triple mutant of all 3 homologs), and *topp4-1* are severely dwarfed, while *sly1-10* has a less severe phenotype (16,58,65). The change in phenotype occurs at the dephosphorylation step, which follows GID1 binding in this model. It is known that GID1 binding decreases RGA activity even without RGA degradation (16), and that phosphorylation of OsSLR1 increases its activity (55). The relatively open and flexible structure of intrinsically disordered regions (IDRs), like the N-terminal regions of histones and DELLA proteins, allows for binding to multiple protein partners with high specificity and low affinity. The low affinity allows for the partners to bind and dissociate easily. The flexibility also reduces steric hindrance that might interfere with binding. PTMs like phosphorylation and acetylation change the shape and electrostatics of the surface of the protein, and thus influence the interactions of RGA with other proteins (59). However, RGA binds most of its targets via the LHRI subdomain of the C-terminal GRAS domain (6,9), so perhaps the N-terminal region affects initial target recognition while the LHRI subdomain binds the target more strongly to sequester it. Thus, unmodified RGA (in *sly1-10*) would have lower initial binding to its targets than would the phosphorylated acetylated form (in *ga1-3*), explaining why *ga1-3* has more repression of growth and a worse phenotype than *sly1-10*.

Finally, how do GlcNAcylation and fucosylation interact with phosphorylation and acetylation? SEC and SPY do not affect RGA nuclear localization or protein level (18,19). Serines and threonines are sites of the first three PTMs, which can compete for occupancy (3,66). This is especially likely in the LSN peptide, which has all three. While the effect of fucosylation is less clear, the *sly1-10 sec-3* mutant had no LSN GlcNAcylation but increased LSN phosphorylation and acetylation relative to *sly1-10*. This agrees with the Zentella papers where GlcNAcylation reduces RGA binding to target TFs and TRs (3,18), and suggests that the effect of GlcNAcylation could be due to it preventing phosphorylation on certain serines and threonines in LSN, which also affects acetylation of K26.

**2.6 Conclusions and Future Directions**

With this analysis of post-translational modification sites on RGA, questions remain. The model proposed above should be evaluated with mass spectrometry characterization of *topp4* and *gid1* mutants. If the model is correct, then *gid1* and *topp4* mutants would have levels of phosphorylation and acetylation similar to *ga1-3*, and higher than *sly1-10*. Since Qin 2014 did not determine the residues dephosphorylated by TOPP4 (58), mass spectrometry should be used to confirm if they are the same phosphosites seen here.

Furthermore, the kinase that modifies RGA has not been identified. It is probably a casein kinase, like EL1 which phosphorylates OsSLR1 (55). Since this is the first known instance of acetylation of RGA, the HAT and HDAC involved are also unknown. Several methods for determining the identity of these specific enzymes exist. Similar to the *spy* and *sec* mutants examined in this study, *Arabidopsis* plants with an inactivating mutation in the enzyme of interest could be analyzed. If the enzyme was the one that altered the PTM status of RGA at

that site, the mutant would show altered levels of that PTM. Kinase and HAT mutants would have reduced or abolished phosphorylation and acetylation at specific residues, respectively, while phosphatase and HDAC mutants would have increased phosphorylation and acetylation. Biological assays can determine a general increase or decrease in the level of a PTM on a protein. However, mass spectrometry is uniquely suited to identify changes at specific sites, as demonstrated here. Protein-protein interaction assays like yeast-two-hybrid techniques and co-immunoprecipitation can demonstrate that an enzyme and RGA physically interact. In vitro assays can verify the ability of an enzyme to modify RGA (3,18,58).

The RGA analyzed here possessed tags for purification on the N-terminus and a lysine inserted between G185 and G186, giving FLAG-RGA$^{GKG}$. In previous studies, FLAG-RGA (non-GKG) expressed in *Arabidopsis* was able to be pulled down by target transcription factors (3), so the N-terminal tags did not interfere. Also, FLAG-RGA$^{GKG}$ had similar mobility to FLAG-RGA in an anti-FLAG Western blot in tobacco with and without SEC overexpression (18). Still, it would be wise to characterize acetylation on non-GKG RGA. Acetylated RGA can be detected by Western blot with an antibody specific for acetylated lysine (42). More importantly, acetylated K26 and K163 should be observable by mass spectrometry in non-GKG RGA, since the LSN and VIP peptides are still present.

Determining what PTMs exist simultaneously on RGA is important for understanding how they affect one another. Trypsin tends to generate short peptides. A digestion method that generates longer peptides which encompass more PTM sites would be preferable. A new method of digesting RGA would be to use the Aspergillopepsin enzyme reactor developed in the Hunt lab by Lichao Zhang et al. (67) and improved upon by Joshua Hinkle et al. (68). Unlike most commonly-used proteases, Aspergillopepsin is a non-specific protease that cleaves at

several different sites, usually N and C-terminal to hydrophobic residues or C-terminal to Lys. Aspergillopepsin I (from *Aspergillus saitoi*) is notable for being active in 8 M urea (pH 3-4) for at least 1 hour. These are highly denaturing and chaotropic conditions that will cause most other proteins to unfold, making them more accessible for proteolysis along their entire length. Zhang et al. demonstrated that by attaching aspergillopepsin to beads and loading them in a fused silica column, a sample may be digested to different extents depending on the flow rate of the sample through the column. A faster flow rate means the sample proteins are exposed to the aspergillopepsin for a shorter time period, so they are cleaved to a lesser extent, resulting in larger peptides (67). Due to the non-specific nature of Aspergillopepsin cleavage, it produces series of overlapping peptides. This, combined with the larger peptides generated, and the parallelization and sophisticated method design offered by the Orbitrap Fusion (68), would allow for better analysis of simultaneous PTMs of RGA on distant sites (such as on the SCSS and GVIG stretches). Joshua Hinkle and I performed some initial work on RGA digestion with the aspergillopepsin enzyme reactor, but the digestion unexpectedly produced shorter than expected peptides at higher abundances, likely due to the low concentration of RGA in the solution. Still, some GlcNAcylated RGA peptides were observed, indicating that this would be a promising avenue to follow.

Acetylation likely exists on other DELLA proteins than RGA (such as in *Paulownia tomentosa* (46)). There have been high-throughput analyses of plant acetylation (43–46,49) and studies of histone acetylation (60), but no specific studies of DELLA protein acetylation before now. OsSLR1 is a good candidate for the first studies outside of *Arabidopsis*. The gibberellin degradation pathway is well characterized in rice, and the involvement of OsSLR1 phosphorylation has been studied (6,51,52,54,55). Unlike in *Arabidopsis*, the identity of the

kinase (55) but not the phosphatase is known. If acetylation exists on OsSLR1 and has an effect similar to RGA acetylation, that would support the mechanism described here. It would also suggest that this occurs across different plant species, and could be a new molecular mechanism to modulate to improve crops.

The K-(K/R)-x-(K/R) motif is conserved in DELLA proteins in Angiosperms, which include virtually all food plants (69). New dwarfing alleles would be useful for at least some of these. For example, the commercially used rice dwarfing alleles are all in OsSD1, a gibberellin biosynthesis gene (further in the pathway than GA1). These mutants are recessive, requiring both copies in a plant to have the allele (homozygosity), which restricts breeding options, overall genetic diversity, and thus increases the vulnerability of the crop to diseases and pests. The existing *Osslr1* mutants have phenotypes too severe to be useful (extreme dwarfing and reduced fertility). However, in 2018, Wu et al. (70) found a new semi-dominant allele with a mutation in the serine following the TVHYNP motif with milder dwarfing and good yield. The mutant was able to bind to GID1, though less strongly than wild type.

The development of Green Revolution mutants in the 1950s and 1960s relied on breeding plants based on beneficial phenotypes; the genes responsible were only identified several decades later (13). Exposing plants to mutagens like ethyl methanesulfonate has been used to generate mutants for phenotypic screening, but the changes occur randomly throughout the genome (71,72). In contrast, CRISPR/Cas9 can be used to produce mutations at a specific point in a gene. in 2019, Tomlinson et al. (73) used this technique in the tomato DELLA protein PROCERA and produced a dominant dwarfing mutant, deleting the A in the DELLA motif. Shorter tomato plants require less labor-intensive management to cultivate. If the model presented here is correct, deleting or mutating the K163 equivalent would not

impair GID1 binding or dephosphorylation, but would prevent DELLA ubiquitylation and degradation (and also acetylation). This would probably have a milder phenotype than the prevalent DELLA/TVHYNP motif mutants, more similar to *sly1-10*. Alternatively, modulating the enzymes responsible for phosphorylation and acetylation could be useful. Mutating the phosphatase is unlikely to be beneficial, as mutant *topp4* in *Arabidopsis* has a very severe phenotype (58). Mutating the HDAC or HAT, once identified, would again be likely to have a milder phenotype. However, CRISPR/Cas9 alteration of the DELLA protein would be the best place to start, since the DELLA protein has been identified in most of these species, while much less is currently known about PTMs of DELLA proteins outside *Arabidopsis* and rice.

Using the formidable analytical power of mass spectrometry, this work examined post-translational modifications of the DELLA protein RGA, and identified the first acetylation on RGA. This acetylation was some of the first seen on any DELLA protein. The acetylated lysine K163 aligns with the only other known acetylation site on a DELLA protein, which is part of a K-(K/R)-x-(K/R) motif conserved across many DELLA proteins. This conservation suggests that acetylation of this lysine may play an important role in DELLA activity and stability, as does the different levels of RGA acetylation in mutants of the gibberellin signaling pathway such as *ga1-3* and *sly1-10*. Phosphorylation and acetylation levels in these mutants were correlated, varying in concert across different mutants. This is likely because GID1 binding to RGA causes TOPP4 to dephosphorylate RGA, which recruits a HDAC to deacetylate K163 and K26, which allows the SCF complex to ubiquitylate those lysines and thus cause RGA proteasomal degradation. These findings enhance our knowledge of DELLA post-translational modification in the model plant *Arabidopsis thaliana*, and thus our understanding of this class of master growth regulator proteins in plants as a whole. Malthus may ever be dogging our

steps, but the application of this knowledge continues the legacy of the Green Revolution and may stave off famine for another day.

## 2.7 References


1. Davière J-M, Achard P, Gibberellin Signaling in Plants, Development, **2013**;140(6):1147–1151.
2. Cenci A, Rouard M, Evolutionary Analyses of GRAS Transcription Factors in Angiosperms, Front Plant Sci, **2017**;8(273):1–15.
3. Zentella R, Sui N, Barnhill B, Hsieh W-P, Hu J, Shabanowitz J, Boyce M, Olszewski NE, Zhou P, Hunt DF, Sun T, The *Arabidopsis* O-Fucosyltransferase SPINDLY Activates Nuclear Growth Repressor DELLA, Nat Chem Biol, **2017**;13(5):479–485.
4. Hirano Y, Nakagawa M, Suyama T, Murase K, Shirakawa M, Takayama S, Sun T, Hakoshima T, Structure of the SHR-SCR Heterodimer Bound to the BIRD/IDD Transcriptional Factor JKD, Nat Plants, **2017**;3(17010).
5. Murase K, Hirano Y, Sun T, Hakoshima T, Gibberellin-Induced DELLA Recognition by the Gibberellin Receptor GID1, Nature, **2008**;456(7221):459–463.
6. Nelson SK, Steber CM, Gibberellin Hormone Signal Perception: Down-Regulating DELLA Repressors of Plant Growth and Development, Annu Plant Rev, **2016**;49:153–188.
7. Vaschetto LM, Understanding the Role of Protein Interaction Motifs in Transcriptional Regulators: Implications for Crop Improvement, Brief Funct Genomics, **2017**;16(3):152–155.
8. Davière J-M, Achard P, A Pivotal Role of DELLAs in Regulating Multiple Hormone Signals, Mol Plant, **2016**;9(1):10–20.
9. Van De Velde K, Ruelens P, Geuten K, Rohde A, Van Der Straeten D, Exploiting DELLA Signaling in Cereals, Trends Plant Sci, **2017**;22(10):880–893.
10. Stark C, Breitkreutz B-J, Reguly T, Boucher L, Breitkreutz A, Tyers M, BioGRID: a General Repository for Interaction Datasets, Nucleic Acids Res, **2006**;34(Database):D535–D539.
11. Hill K, Post-Translational Modifications of Hormone-Responsive Transcription Factors: The Next Level of Regulation, J Exp Bot, **2015**;66(16):4933–4945.
12. Bai M-Y, Shang J-X, Oh E, Fan M, Bai Y, Zentella R, Sun T, Wang Z-Y, Brassinosteroid, Gibberellin, and Phytochrome Impinge on a Common Transcription Module in *Arabidopsis*, Nat Cell Biol, **2012**;14(8):810–817.
13. Hedden P, The Genes of the Green Revolution, Trends Genet, **2003**;19(1):5–9.
14. McGinnis KM, Thomas SG, Soule JD, Strader LC, Zale JM, Sun T, Steber CM, The *Arabidopsis* SLEEPY1 Gene Encodes a Putative F-Box Subunit of an SCF E3 Ubiquitin

Ligase, Plant Cell, **2003**;15(5):1120–1130.

15. Dill A, Thomas SG, Hu J, Steber CM, Sun T, The *Arabidopsis* F-box protein SLEEPY1 Targets Gibberellin Signaling Repressors for Gibberellin-Induced Degradation, Plant Cell, **2004**;16(6):1392–1405.

16. Ariizumi T, Murase K, Sun T, Steber CM, Proteolysis-Independent Downregulation of DELLA Repression in *Arabidopsis* by the Gibberellin Receptor GIBBERELLIN INSENSITIVE DWARF1, Plant Cell, **2008**;20(9):2447–2459.

17. Sun T, Kamiya Y, The *Arabidopsis* GA1 Locus Encodes the Cyclase ent-Kaurene Synthetase A of Gibberellin Biosynthesis, Plant Cell, **1994**;6(10):1509–1518.

18. Zentella R, Hu J, Hsieh W-P, Matsumoto PA, Dawdy AW, Barnhill B, Oldenhof H, Hartweck LM, Maitra S, Thomas SG, Cockrell S, Boyce M, Shabanowitz J, Hunt DF, Olszewski NE, Sun T, O-GlcNAcylation of Master Growth Repressor DELLA by SECRET AGENT Modulates Multiple Signaling Pathways in *Arabidopsis*, Genes Dev, **2016**;30(2):164–176.

19. Silverstone AL, Tseng T-S, Swain SM, Dill A, Jeong SY, Olszewski NE, Sun T, Functional Analysis of SPINDLY in Gibberellin Signaling in *Arabidopsis*, Plant Physiol, **2007**;143(2):987–1000.

20. Dawdy AW, Characterization of O-Glycosylation and Phosphorylation on Nuclear Protein by Mass Spectrometry, **2013**.

21. Derevnina L, Kamoun S, Wu C, Dude, Where is My Mutant? *Nicotiana benthamiana* Meets Forward Genetics, New Phytol, **2019**;221(2):607–610.

22. Zentella R, Zhang Z-L, Park M, Thomas SG, Endo A, Murase K, Fleet CM, Yusuke J, Nambara E, Kamiya Y, Sun T, Global Analysis of DELLA Direct Targets in Early Gibberellin Signaling in *Arabidopsis*, Plant Cell, **2007**;19(10):3037–3057.

23. Hochuli E, Döbeli H, Schacher A, New Metal Chelate Adsorbent Selective for Proteins and Peptides Containing Neighbouring Histidine Residues, J Chromatogr A, **1987**;411:177–184.

24. Einhauer A, Jungbauer A, The FLAG™ Peptide, a Versatile Fusion Tag for the Purification of Recombinant Proteins, J Biochem Biophys Methods, **2001**;49:455–465.

25. Bowler C, Benvenuto G, Laflamme P, Molino D, Probst A V, Tariq M, Paszkowski J, Chromatin Techniques for Plant Cells, Plant J, **2004**;39(5):776–789.

26. Bern M, Cai Y, Goldberg D, Lookup Peaks: A Hybrid of de Novo Sequencing and Database Search for Protein Identification by Tandem Mass Spectrometry, Anal Chem, **2007**;79(4):1393–1400.

27. Bern M, Kil YJ, Becker C, Byonic: Advanced Peptide and Protein Identification Software, Curr Protoc Bioinforma, **2012**;40(1):13.20.1-14.

28. Sleno L, The Use of Mass Defect in Modern Mass Spectrometry, J Mass Spectrom, **2012**;47(2):226–236.

29. Steen H, Mann M, The ABC's (and XYZ's) of Peptide Sequencing, Nat Rev Mol Cell Biol,

**2004**;5(9):699–711.

30. Monroe ME, Molecular Weight Calculator, **2014**; Available from: https://omics.pnl.gov/software.

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

32. Eliuk SM, Makarov A, Evolution of Orbitrap Mass Spectrometry Instrumentation, Annu Rev Anal Chem, **2015**;8(1):61–80.

33. Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V, Church DM, DiCuccio M, Edgar R, Federhen S, Geer LY, Kapustin Y, Khovayko O, Landsman D, Lipman DJ, Madden TL, Maglott DR, Ostell J, Miller V, et al., Database Resources of the National Center for Biotechnology Information, Nucleic Acids Res, **2007**;35(Database):D5–D12.

34. Millar AH, Heazlewood JL, Giglione C, Holdsworth MJ, Bachmair A, Schulze WX, The Scope, Functions, and Dynamics of Posttranslational Protein Modifications, Annu Rev Plant Biol, **2019**;70(1):119–151.

35. Darula Z, Medzihradszky KF, Carbamidomethylation Side Reactions May Lead to Glycan Misassignments in Glycopeptide Analysis, Anal Chem, **2015**;87(12):6297–6302.

36. Rathore D, Aboufazeli F, Huang Y, Kolli V, Fernando GS, Dodds ED, Ion Dissociation Methods in Proteomics, Encycl Anal Chem, **2015**;1–26.

37. Savaryn JP, Toby TK, Kelleher NL, A Researcher's Guide to Mass Spectrometry-Based Proteomics, Proteomics, **2016**;16(18):2435–2443.

38. Brodbelt JS, Ion Activation Methods for Peptides and Proteins, Anal Chem, **2016**;88(1):30–51.

39. Mann M, Jensen ON, Proteomic Analysis of Post-Translational Modifications, Nat Biotechnol, **2003**;21(3):255–261.

40. Ansari AA, Kidwai SA, Salahuddin A, Acetylation of Amino Groups and Its Effect on the Conformation and Immunological Activity of Ovalbumin, J Biol Chem, **1975**;250(5):1625–1632.

41. Zee BM, Garcia BA, Discovery of Lysine Post-Translational Modifications through Mass Spectrometric Detection, Essays Biochem, **2012**;52:147–163.

42. Choudhary C, Kumar C, Gnad F, Nielsen ML, Rehman M, Walther TC, Olsen J V, Mann M, Lysine Acetylation Targets Protein Complexes and Co-Regulates Major Cellular Functions, Science, **2009**;325(5942):834–840.

43. Wu X, Oh M-H, Schwarz EM, Larue CT, Sivaguru M, Imai BS, Yau PM, Ort DR, Huber SC, Lysine Acetylation is a Widespread Protein Modification for Diverse Proteins in *Arabidopsis*, Plant Physiol, **2011**;155(4):1769–1778.

44. Hartl M, Füßl M, Boersema PJ, Jost J-O, Kramer K, Bakirbas A, Sindlinger J, Plöchinger M, Leister D, Uhrig RG, Moorhead GBG, Cox J, Salvucci ME, Schwarzer D, Mann M, Finkemeier I, Lysine Acetylome Profiling Uncovers Novel Histone Deacetylase Substrate Proteins in *Arabidopsis*, Mol Syst Biol, **2017**;13(10):949.

45. Finkemeier I, Laxa M, Miguet L, Howden AJM, Sweetlove LJ, Proteins of Diverse Function and Subcellular Location Are Lysine Acetylated in *Arabidopsis*, Plant Physiol, **2011**;155(4):1779–1790.

46. Cao Y, Fan G, Wang Z, Gu Z, Phytoplasma-Induced Changes in the Acetylome and Succinylome of *Paulownia tomentosa* Provide Evidence for Involvement of Acetylated Proteins in Witches' Broom Disease, Mol Cell Proteomics, **2019**;18(6):1210–1226.

47. Fan G, Cao X, Niu S, Deng M, Zhao Z, Dong Y, Transcriptome, MicroRNA, and Degradome Analyses of the Gene Expression of *Paulownia* with Phytoplamsa, BMC Genomics, **2015**;16(896):1–15.

48. Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, de Castro E, Duvaud S, Flegel V, Fortier A, Gasteiger E, Grosdidier A, Hernandez C, Ioannidis V, Kuznetsov D, Liechti R, Moretti S, Mostaguir K, Redaschi N, Rossier G, et al., ExPASy: SIB Bioinformatics Resource Portal, Nucleic Acids Res, **2012**;40(Web Server):W597–W603.

49. Uhrig RG, Schläpfer P, Roschitzki B, Hirsch-Hoffmann M, Gruissem W, Diurnal Changes in Concerted Plant Protein Phosphorylation and Acetylation in *Arabidopsis* Organs and Seedlings, Plant J, **2019**;99(3):176–194.

50. Zheng H, Zhang F, Wang S, Su Y, Ji X, Jiang P, Chen R, Hou S, Ding Y, MLK1 and MLK2 Coordinate RGA and CCA1 Activity to Regulate Hypocotyl Elongation in *Arabidopsis* thaliana, Plant Cell, **2018**;30(1):67–82.

51. Sasaki A, Itoh H, Gomi K, Ueguchi-Tanaka M, Ishiyama K, Kobayashi M, Jeong D-H, An G, Kitano H, Ashikari M, Matsuoka M, Accumulation of Phosphorylated Repressor for Gibberellin Signaling in an F-Box Mutant, Science, **2003**;299(5614):1896–1898.

52. Gomi K, Sasaki A, Itoh H, Ueguchi-Tanaka M, Ashikari M, Kitano H, Matsuoka M, GID2, an F-box Subunit of the SCF E3 Complex, Specifically Interacts with Phosphorylated SLR1 Protein and Regulates the Gibberellin-Dependent Degradation of SLR1 in Rice, Plant J, **2004**;37(4):626–634.

53. Deshaies R, SCF and Cullin/RING H2-Based Ubiquitin Ligases, Annu Rev Cell Dev Biol, **1999**;15(1):435–467.

54. Itoh H, Sasaki A, Ueguchi-Tanaka M, Ishiyama K, Kobayashi M, Hasegawa Y, Minami E, Ashikari M, Matsuoka M, Dissection of the Phosphorylation of Rice DELLA Protein, SLENDER RICE1, Plant Cell Physiol, **2005**;46(8):1392–1399.

55. Dai C, Xue H-W, Rice early flowering1, a CKI, Phosphorylates DELLA Protein SLR1 to Negatively Regulate Gibberellin Signalling, EMBO J, **2010**;29(11):1916–1927.

56. Hussain A, Cao D, Cheng H, Wen Z, Peng J, Identification of the Conserved Serine/Threonine Residues Important for Gibberellin-Sensitivity of *Arabidopsis* RGL2 Protein, Plant J, **2005**;44(1):88–99.

57. Wang F, Zhu D, Huang X, Li S, Gong Y, Yao Q, Fu X, Fan L-M, Deng XW, Biochemical Insights on Degradation of *Arabidopsis* DELLA Proteins Gained from a Cell-Free Assay System, Plant Cell, **2009**;21(8):2378–2390.

58. Qin Q, Wang W, Guo X, Yue J, Huang Y, Xu X, Li J, Hou S, *Arabidopsis* DELLA Protein Degradation Is Controlled by a Type-One Protein Phosphatase, TOPP4, PLoS Genet, **2014**;10(7):e1004464.

59. Sun X, Jones WT, Rikkerink EHA, GRAS Proteins: The Versatile Roles of Intrinsically Disordered Proteins in Plant Signalling, Biochem J, **2012**;442(1):1–12.

60. Johnson L, Mollah S, Garcia BA, Muratore TL, Shabanowitz J, Hunt DF, Jacobsen SE, Mass Spectrometry Analysis of *Arabidopsis* Histone H3 Reveals Distinct Combinations of Post-Translational Modifications, Nucleic Acids Res, **2004**;32(22):6511–6518.

61. Cheung P, Tanner KG, Cheung WL, Sassone-Corsi P, Denu JM, Allis CD, Synergistic Coupling of Histone H3 Phosphorylation and Acetylation in Response to Epidermal Growth Factor Stimulation, Mol Cell, **2000**;5(6):905–915.

62. Tran HT, Nimick M, Uhrig RG, Templeton G, Morrice N, Gourlay R, DeLong A, Moorhead GB, *Arabidopsis thaliana* Histone Deacetylase 14 (HDA14) is an α-Tubulin Deacetylase that Associates with PP2A and Enriches in the Microtubule Fraction with the Putative Histone Acetyltransferase ELP3, Plant J, **2012**;71(2):263–272.

63. Friso G, van Wijk KJ, Posttranslational Protein Modifications in Plant Metabolism, Plant Physiol, **2015**;169(3):1469–1487.

64. Lin R, Tao R, Gao X, Li T, Zhou X, Guan K-L, Xiong Y, Lei Q-Y, Acetylation Stabilizes ATP-Citrate Lyase to Promote Lipid Biosynthesis and Tumor Growth, Mol Cell, **2013**;51(4):506–518.

65. Griffiths J, Murase K, Rieu I, Zentella R, Zhang Z-L, Powers SJ, Gong F, Phillips AL, Hedden P, Sun T, Thomas SG, Genetic Characterization and Functional Analysis of the GID1 Gibberellin Receptors in *Arabidopsis*, Plant Cell, **2006**;18(12):3399–3414.

66. Martínez-Turiño S, Pérez JDJ, Hervás M, Navajas R, Ciordia S, Udeshi ND, Shabanowitz J, Hunt DF, García JA, Phosphorylation Coexists with O-GlcNAcylation in a Plant Virus Protein and Influences Viral Infection, Mol Plant Pathol, **2018**;19(6):1427–1443.

67. Zhang L, English AM, Bai DL, Ugrin SA, Shabanowitz J, Ross MM, Hunt DF, Wang W-H, Analysis of Monoclonal Antibody Sequence and Post-Translational Modifications by Time-Controlled Proteolysis and Tandem Mass Spectrometry, Mol Cell Proteomics, **2016**;15(4):1479–1488.

68. Hinkle JD, D'Ippolito RA, Panepinto MC, Wang W-H, Bai DL, Shabanowitz J, Hunt DF, Unambiguous Sequence Characterization of a Monoclonal Antibody in a Single Analysis Using a Nonspecific Immobilized Enzyme Reactor, Anal Chem, **2019**;91(21):13547–13554.

69. Taylor TN, Taylor EL, Krings M, Paleobotany: The Biology and Evolution of Fossil Plants, 2nd ed. Academic Press; 2009.

70. Wu Z, Tang D, Liu K, Miao C, Zhuo X, Li Y, Tan X, Sun M, Luo Q, Cheng Z, Characterization

of a New Semi-Dominant Dwarf Allele of SLR1 and its Potential Application in Hybrid Rice Breeding, J Exp Bot, **2018**;69(20):4703–4713.

71. Silverstone AL, Mak PYA, Martínez EC, Sun T, The New RGA Locus Encodes a Negative Regulator of Gibberellin Response in *Arabidopsis thaliana*, Genetics, **1997**;146(3):1087–1099.

72. Shinozaki Y, Ezura K, Hu J, Okabe Y, Bénard C, Prodhomme D, Gibon Y, Sun T, Ezura H, Ariizumi T, Identification and Functional Study of a Mild Allele of SlDELLA Gene Conferring the Potential for Improved Yield in Tomato, Sci Rep, **2018**;8(1):1–15.

73. Tomlinson L, Yang Y, Emenecker R, Smoker M, Taylor J, Perkins S, Smith J, MacLean D, Olszewski NE, Jones JDG, Using CRISPR/Cas9 Genome Editing in Tomato to Create a Gibberellin-Responsive Dominant Dwarf DELLA Allele, Plant Biotechnol J, **2019**;17(1):132–140.

## Appendix: Relative levels of post-translational modifications of RGA in *Arabidopsis* mutants

| LSNHGTSSSSSSISK(DK) | 7/17 *ga1-3* | 8/18 EHS *ga1-3* | 8/18 EBZ *ga1-3* | 5/17 *sly1-10* | 12/17 *sly1-10* | 3/18 *sly1-10* | *sly1-10 spy-8* | *sly1-10 spy-12* | *sly1-10 spy-15* | *sly1-10 spy-19* | 4/18 *sly1-10 sec-3* | 8/18 *sly1-10 sec-3* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Unmodified | 67% | 42% | 55% | 96% | 93% | 90% | 85% | 89% | 74% | 86% | 80% | 87% |
| + Fucose | 0.2% | 3.8% | 0.5% | 0.4% | 0.4% | 0.3% | 1.7% | ND | <0.1% | ND | 0.5% | 1.4% |
| + Hexose | ND | ND | ND | <0.1% | ND | ND | ND | <0.1% | ND | <0.1% | ND | ND |
| + GlcNAc | 2.8% | 15.3% | 1.2% | 2.2% | 0.9% | 2.3% | 2.5% | 8.1% | 2.6% | 10.1% | ND | ND |
| + PO4 | 4.3% | 28.7% | 18.1% | 1.3% | 5.2% | 2.7% | 5.3% | 2.2% | 19.5% | 2.9% | 7.1% | 9.0% |
| + 2*PO4 | 17.0% | 8.2% | 20.7% | <0.1% | 0.2% | 3.0% | 3.0% | 0.1% | 2.2% | 0.3% | 6.3% | 1.6% |
| + 3*PO4 | 5.1% | 0.9% | 2.6% | ND | ND | 0.2% | 0.2% | ND | ND | ND | 0.8% | ND |
| + GlcNAc + PO4 | 0.1% | 0.4% | 0.8% | <0.1% | <0.1% | ND | 0.5% | <0.1% | 0.3% | ND | ND | ND |
| + Acetyl | 3.0% | 0.6% | 1.0% | <0.1% | 0.5% | 1.2% | 1.5% | 0.2% | 1.1% | 0.2% | 4.9% | 1.0% |

| VIPGNAIYQFPAIDSSSSSNNQNK(R) | 7/17 *ga1-3* | 8/18 EHS *ga1-3* | 8/18 EBZ *ga1-3* | 5/17 *sly1-10* | 12/17 *sly1-10* | 3/18 *sly1-10* | *sly1-10 spy-8* | *sly1-10 spy-12* | *sly1-10 spy-15* | *sly1-10 spy-19* | 4/18 *sly1-10 sec-3* | 8/18 *sly1-10 sec-3* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Unmodified | 97% | 98% | 98% | 100% | 100% | 100% | 100% | 99% | 100% | 100% | 99% | 100% |
| + GlcNAc | ND | ND | ND | ND | ND | ND | ND | ND | ND | ND | ND | ND |
| + PO4 | 1.9% | 2.0% | 2.0% | 0.3% | 0.2% | 0.2% | 0.2% | <0.1% | 0.2% | 0.1% | 0.5% | ND |
| + 2*PO4 | 0.9% | ND | 0.2% | <0.1% | 0.2% | 0.2% | ND | 1.0% | 0.1% | ND | ND | ND |
| + Acetyl | 0.6% | 0.2% | 0.2% | <0.1% | ND | ND | <0.1% | ND | <0.1% | ND | 0.1% | ND |

| (LK)SCSSPDSMVTSTSTGTQIGK | 7/17 *ga1-3* | 8/18 EHS *ga1-3* | 8/18 EBZ *ga1-3* | 5/17 *sly1-10* | 12/17 *sly1-10* | 3/18 *sly1-10* | *sly1-10 spy-8* | *sly1-10 spy-12* | *sly1-10 spy-15* | *sly1-10 spy-19* | 4/18 *sly1-10 sec-3* | 8/18 *sly1-10 sec-3* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Unmodified | 58% | 68% | 68% | 72% | 72% | 71% | 75% | 80% | 75% | 74% | 75% | 76% |
| + Fucose | ND | ND | ND | ND | ND | ND | ND | ND | ND | ND | ND | ND |
| + GlcNAc | 7.4% | 3.2% | 2.4% | 4.7% | 2.6% | 3.7% | 5.3% | 4.6% | 5.1% | 4.4% | ND | ND |
| + PO4 | 33.1% | 26.6% | 28.5% | 23.6% | 25.0% | 25.3% | 19.9% | 15.2% | 19.6% | 22.0% | 25.0% | 24.2% |
| + 2*PO4 | 2.0% | 1.8% | 1.4% | <0.1% | ND | ND | ND | ND | ND | ND | ND | ND |

| GVIGTTVTTTTTTTTAAGESTR | 7/17 *ga1-3* | 8/18 EHS *ga1-3* | 8/18 EBZ *ga1-3* | 5/17 *sly1-10* | 12/17 *sly1-10* | 3/18 *sly1-10* | *sly1-10 spy-8* | *sly1-10 spy-12* | *sly1-10 spy-15* | *sly1-10 spy-19* | 4/18 *sly1-10 sec-3* | 8/18 *sly1-10 sec-3* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Unmodified | 11% | 36% | 23% | 31% | 38% | 30% | 29% | 30% | 27% | 30% | 99% | 100% |
| + Fucose | ND | ND | ND | 0.1% | ND | ND | ND | ND | ND | ND | 0.2% | ND |
| + Hex | 12.9% | 1.8% | 4.0% | 11.9% | 11.5% | 12.0% | 9.2% | 12.0% | 9.3% | 4.1% | <0.1% | ND |
| + 2*Hex | 0.7% | 0.2% | ND | 0.2% | 0.3% | 0.4% | 0.3% | 0.5% | 0.4% | 0.6% | ND | ND |
| + GlcNAc | 33.0% | 44.7% | 47.2% | 42.0% | 41.3% | 42.5% | 43.6% | 45.4% | 41.9% | 41.3% | 0.3% | ND |
| + 2*GlcNAc | 21.0% | 10.8% | 16.3% | 9.3% | 5.3% | 9.5% | 10.8% | 7.2% | 13.2% | 13.1% | ND | ND |
| + 3*GlcNAc | 4.8% | 1.9% | 2.4% | 0.6% | 0.3% | 0.8% | 1.8% | 0.8% | 2.0% | 2.2% | ND | ND |
| + 4*GlcNAc | 0.4% | ND | ND | ND | ND | ND | ND | ND | ND | ND | ND | ND |
| + Hex + GlcNAc | 11.4% | 4.1% | 5.9% | 4.2% | 2.8% | 4.4% | 4.4% | 4.0% | 5.4% | 6.9% | ND | ND |
| + 2*Hex + GlcNAc | 0.5% | ND | ND | <0.1% | ND | ND | <0.1% | ND | ND | 0.1% | ND | ND |
| + Hex + 2*GlcNAc | 3.4% | 0.5% | 1.0% | 0.3% | 0.3% | 0.5% | 1.0% | 0.5% | 1.1% | 1.8% | ND | ND |
| + Fucose + GlcNAc | 0.6% | ND | ND | 0.2% | 0.1% | 0.3% | ND | ND | ND | ND | ND | ND |
| + PO4 | ND | ND | ND | ND | ND | ND | ND | ND | ND | ND | ND | ND |

**Appendix: Relative levels of post-translational modifications of RGA in *Arabidopsis* mutants.** ND: not detected (threshold ~1e6). Grey cells: no MS2s from the peptide with this modification were acquired in this sample, but it had a comparable retention time to the other sample and low ppm mass error. All columns constitute a different biological sample except for the 8/18 *ga1-3* sample, which was analyzed in two separate instrument runs. The 7/17 *ga1-3* and 5/17 *sly1-10* samples were analyzed with a 300-1200 m/z MS1 window; the rest of the samples were analyzed with a 300-1500 m/z window. The 12/17 *sly1-10* and *sly1-10 spy-12* samples were analyzed with the QMF operated in mass filter mode, while the rest of the samples were analyzed with the QMF in ion transmission mode. The 8/18 EBZ *ga1-3*, 3/18 *sly1-10*, and 8/18 *sly1-10 sec-3* samples were analyzed on Fusion by Emily Zahn. Those runs and the EHS 8/18 *ga1-3* and *sly1-10 spy-19* runs were numerically analyzed by Emily Zahn and/or Mark Ross. The other instrument runs and numerical analyses were performed by Ellen Speers, as was most of the site-localization of PTMs (working with Jeffrey Shabanowitz). Where there were multiple analytical runs of a given mutant (*ga1-3*, *sly1-10*, and *sly1-10 sec-3*), the percent relative abundances in each sample of a given PTM form were averaged to give the numbers in Table 2.2 (with samples where the form was not detected counted as 0%).